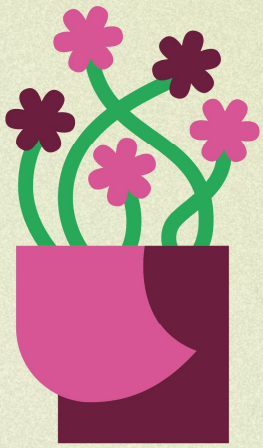
37º CICT
Ciência Delas

Anais

37º Congresso de Iniciação Científica e Tecnológica da UFRN





37º CICT

Ciência Delas

UFRN PROPESQ CNPq

EXPEDIENTE

APRESENTAÇÃO

O Congresso de Iniciação Científica e Tecnológica (CICT) é um evento aberto à comunidade no qual todos os aproximadamente 1.800 alunos de Iniciação Científica e Tecnológica da UFRN (bolsistas e voluntários) apresentam os resultados de suas pesquisas desenvolvidas ao longo de um ano, como cumprimento de um plano de trabalho elaborado e orientado por um professor/pesquisador do quadro permanente da Instituição.

O CICT está entre as ações inseridas no âmbito do Programa Institucional de Bolsas de Iniciação Científica e Tecnológica da UFRN, que possui entre seus objetivos: a) Contribuir para a formação e engajamento de recursos humanos em atividades de pesquisa, desenvolvimento tecnológico e inovação; b) Contribuir para a formação científica de recursos humanos que se dedicarão a qualquer atividade profissional; e c) Contribuir para a formação de recursos humanos que se dedicarão ao fortalecimento da capacidade inovadora das empresas no País.

A 37ª edição do Congresso teve como tema “Ciência Delas", estando alinhada aos 17 Objetivos do Desenvolvimento Sustentável da Agenda 2030 da ONU. Ao submeter os trabalhos ao Congresso, os alunos indicaram os objetivos da Agenda 2030 aos quais suas pesquisas estavam relacionadas. Essa iniciativa visa estimular a pesquisa científica e tecnológica voltada para a busca por soluções para os desafios sociais contemporâneos. Em uma etapa inicial do CICT, denominada Mostra de Trabalhos e Vídeos de Ciência, Tecnologia e Inovação, foram realizadas as apresentações dos trabalhos submetidos pelos participantes do evento. As apresentações ocorreram de forma assíncrona, por meio de arquivos digitais (trabalho completo e vídeo), sendo disponibilizadas não só aos avaliadores do congresso pelos Sistemas da UFRN, mas também ao público em geral por meio do site do evento (www.cic.propesq.ufrn.br). Após essa etapa inicial, que ocorreu sob o formato virtual, os 48 melhores trabalhos do evento foram selecionados para apresentação oral em uma etapa complementar do congresso que ocorreu de forma presencial.

Os alunos concorreram ao 10º Prêmio Destaque na Iniciação Científica e Tecnológica da UFRN, que foi instituído pela Pró-Reitoria de Pesquisa em 2017 para reforçar as ações dos programas institucionais de iniciação científica e tecnológica. O prêmio contemplou as categorias Trabalho Destaque de Iniciação Científica e Tecnológica e Vídeo Destaque de Iniciação Científica e Tecnológica e na subcategoria de Inovação. Os premiados foram agraciados com bolsas mensais de até R$ 1100,00 (mil e cem reais) durante o período de um ano. Os orientadores dos alunos premiados adquiriram precedência em relação aos demais para efeitos de concorrência no Edital de Bolsas de Pesquisa da UFRN em 2027.

Ainda em 2026, a Pró-Reitoria de Pesquisa realizou a 8ª Edição do Prêmio Pesquisador Destaque da UFRN, certame instituído em 2019 e que é concedido anualmente aos pesquisadores da instituição que tenham apresentado relevantes contribuições para o desenvolvimento da ciência em cada uma das três grandes áreas do conhecimento: Ciências da Vida; Ciências Exatas, da Terra e Engenharias; e Ciências Humanas e Sociais, Letras e Artes. Além de troféu e certificado, todos os premiados foram contemplados com auxílio no valor de R$ 10.000,00 (dez mil reais) para custear despesas específicas, discriminadas em edital, e tiveram a oportunidade de compartilhar suas experiências com os participantes do CICT 2026 em uma mesa redonda.

Além de valorizar a contribuição das mulheres para a produção do conhecimento científico, promovendo a visibilidade de suas trajetórias e reconhecendo seu papel no desenvolvimento da ciência, o congresso cumpriu três funções valiosas para a instituição: a inserção dos discentes em um ambiente acadêmico de publicação e apresentação dos resultados de pesquisa; a divulgação científica, por meio de vídeos que utilizam uma linguagem científica, porém acessível ao público em geral; e, por fim, a popularização da ciência.

PROGRAMAÇÃO

DATA	ATIVIDADE
09/10	Divulgação da lista dos pré-selecionados para concorrer ao 9º Prêmio Destaque na Iniciação Científica e Tecnológica da UFRN
19/10	Abertura oficial do evento
19/10 a 23/10	CICT 2026 - Etapa presencial
20/10 a 22/10	Sessões de apresentações orais
23/10	Cerimônia de premiação e encerramento do evento

INSTITUTO METROPOLE DIGITAL

CÓDIGO: ET0396

AUTOR: MIGUEL CARPEGIANE MARQUES FARIAS

ORIENTADOR: HEITOR MEDEIROS FLORENCIO

TÍTULO: Automação com CLP e Monitoramento via Internet com dispositivo IoT

Resumo

Este trabalho apresenta a modernização e a integração ciber-física de uma bancada de automação industrial, elevando o monitoramento de um ambiente local para uma arquitetura baseada em nuvem (Cloud) e computação de borda (Edge Computing). O projeto utilizou o CLP Altus XP340 para realizar o controle dinâmico PID do nível de uma planta piloto hidráulica. Para efetivar a convergência TI/TO, empregou-se um gateway industrial Siemens SIMATIC IOT2050 responsável por rotear os dados do processo via protocolo MQTT até a plataforma Node-RED, hospedada em um servidor da Oracle Cloud. O estudo focou na engenharia de rede, avaliando o desempenho da comunicação através de métricas de latência (Round-Trip Time - RTT) e níveis de Qualidade de Serviço (QoS), estabelecendo políticas distintas para o tráfego de telemetria contínua e para o envio de comandos críticos. Os ensaios dinâmicos em malha fechada demonstraram que o controle físico local permaneceu estável e responsivo às mudanças de Setpoint, enquanto a supervisão remota operou com alta confiabilidade e baixos índices de atraso. Adicionalmente, estruturou-se a topologia para testes com controladores virtualizados (vm-xp340). Os resultados validam a eficiência da arquitetura proposta, consolidando a bancada como um ambiente de pesquisa avançada e alinhado aos rigorosos requisitos de comunicação da Indústria 4.0.

Palavras-chave: IIoT; Automação; Edge Computing; MQTT; Node-RED; CLP

TITLE: PLC Automation and Internet Monitoring with IoT Device

Abstract

This work presents the modernization and cyber-physical integration of an industrial automation test bench, elevating the monitoring from a local environment to a cloud-based and edge computing architecture. The project used the Altus XP340 PLC to perform dynamic PID control of a hydraulic pilot plant's level. To effectuate IT/OT convergence, a Siemens SIMATIC IOT2050 industrial gateway was used to route process data via the MQTT protocol to the Node-RED platform, hosted on an Oracle Cloud server. The study focused on network engineering, evaluating communication performance through latency metrics (Round-Trip Time - RTT) and Quality of Service (QoS) levels, establishing distinct policies for continuous telemetry traffic and critical

command sending. Dynamic closed-loop tests demonstrated that local physical control remained stable and responsive to Setpoint changes, while remote supervision operated with high reliability and low delay rates. Additionally, the topology was structured for tests with virtualized controllers (vm-xp340). The results validate the efficiency of the proposed architecture, consolidating the bench as an advanced research environment aligned with the rigorous communication requirements of Industry 4.0.

Keywords: IIoT; Automation; Edge Computing; MQTT; Node-RED; PLC

Introdução

Os sistemas de automação industrial são essenciais para a otimização de processos e o aumento da produtividade. Tradicionalmente baseados em Controladores Lógicos Programáveis (CLPs) e redes locais de campo, esses sistemas operavam de forma isolada. Contudo, a evolução para a Indústria 4.0 exige que essas arquiteturas se tornem sistemas ciber-físicos, permitindo o acesso e a análise de dados em tempo real para garantir a rastreabilidade e a qualidade do fluxo de produção [3].
A Internet das Coisas Industrial (IIoT) surge como a principal tecnologia habilitadora dessa transição, promovendo a convergência entre a Tecnologia da Informação (TI) e a Tecnologia da Operação (TO) [6]. Dando continuidade às pesquisas desenvolvidas no laboratório LIND (NPITI/IMD) — que anteriormente estabeleceram a infraestrutura elétrica e a comunicação básica em rede local —, este plano de trabalho justifica-se pela necessidade de elevar a planta piloto a um padrão real de nuvem e computação de borda (Edge Computing).
A pesquisa concentrou-se na automação e no acompanhamento de uma planta piloto hidráulica, migrando o monitoramento de um computador local para um servidor na nuvem (Oracle Cloud), interligado através de um gateway industrial (Siemens SIMATIC IOT2050) [5][8]. Adicionalmente, o estudo explora o comportamento do controle local frente à latência da internet, analisa diferentes métricas de rede (QoS) e introduz testes com controladores virtuais (CLPs virtualizados) operando via software [2][3].

Metodologia

O projeto foi conduzido em etapas práticas para garantir a transição da bancada de um modelo local para uma arquitetura em nuvem:

1. Revisão Bibliográfica: Levantamento de literatura técnica sobre automação de processos, integração TI/TO, tecnologias de borda (Edge Computing) e virtualização de CLPs [2][6].
2. Implementação da Arquitetura Edge-to-Cloud: O monitoramento via computador local foi substituído por uma arquitetura IIoT robusta. Um gateway industrial (Siemens IOT2050) foi configurado para atuar como ponte entre a rede de operação local da bancada e a rede externa, comunicando-se com um servidor virtual na Oracle Cloud onde a plataforma Node-RED foi hospedada [4][5].
3. Controle Local e Virtualização: Configuração da malha de controle PID no CLP físico (Altus XP340) para o controle dinâmico da bomba e do nível da água. Em

paralelo, foi estruturado um ambiente de simulação com um controlador virtualizado (vm-xp340) operando no switch da topologia de rede para testes de roteamento e intercâmbio de dados.
4. Engenharia de Redes e Configuração de QoS: O fluxo de dados MQTT foi otimizado aplicando lógicas de separação de Quality of Service (QoS) [7]. Definiu-se o nível QoS 0 para o tráfego contínuo de telemetria (visando menor latência) e o QoS 1 para a garantia de entrega de comandos críticos de Setpoint e estado do motor. Adicionou-se uma lógica de carimbo de tempo (timestamping) para calcular o tempo total de ida e volta (Round-Trip Time - RTT) dos pacotes [1].
5. Ensaios Práticos Dinâmicos: Execução de testes contínuos em malha fechada. A planta foi submetida a perfis de degrau com múltiplas alterações de nível programado, gerando arquivos de dados estruturados em .CSV para análise conjunta do comportamento físico do fluido e dos atrasos impostos pela rede externa.

Resultados e Discussões

O sistema de controle local e o monitoramento via IoT funcionaram conforme o esperado durante os testes de bancada.
Para analisar a resposta da malha PID, foi realizado um ensaio contínuo de 30 minutos, configurado com uma frequência de amostragem de 1 Hz (totalizando 1.800 amostras). O teste consistiu em submeter a planta a três degraus de Setpoint (30, 20 e 35), com duração de 10 minutos cada, para avaliar a resposta do sistema nos transientes de subida e descida:

• Estágio 1 (Setpoint = 30): O sistema partiu do repouso. O controlador acionou a bomba (variável de comando) para elevar o nível, estabilizando a variável de processo (PV) na referência determinada com baixo sobressinal.
• Estágio 2 (Setpoint = 20): Foi aplicado um degrau negativo. O CLP reduziu o sinal enviado ao motor, permitindo o escoamento natural da água até o sistema acomodar no novo patamar.
• Estágio 3 (Setpoint = 35): Com a aplicação de um degrau positivo, a bomba retomou a aceleração para vencer a inércia e elevar o volume de fluido, estabilizando novamente o nível em regime permanente.

Durante todo o processo, o painel de monitoramento no Node-RED registrou a dinâmica da planta sem apresentar falhas de conexão. A análise do arquivo .CSV gerado confirmou a integridade dos dados, validando a confiabilidade da coleta de telemetria via IoT. Adicionalmente, as análises de Round-Trip Time (RTT) demonstraram que a configuração de telemetria em QoS 0 garantiu uma latência média estável e inferior a testes com QoS maiores, evitando gargalos no processamento do gateway [1][6].

Conclusão

As atividades desenvolvidas durante o período da bolsa PIBIC-EM permitiram validar o funcionamento prático de uma planta ciber-física. O CLP manteve o controle eficiente do processo físico, enquanto a arquitetura IoT com o gateway de borda Siemens

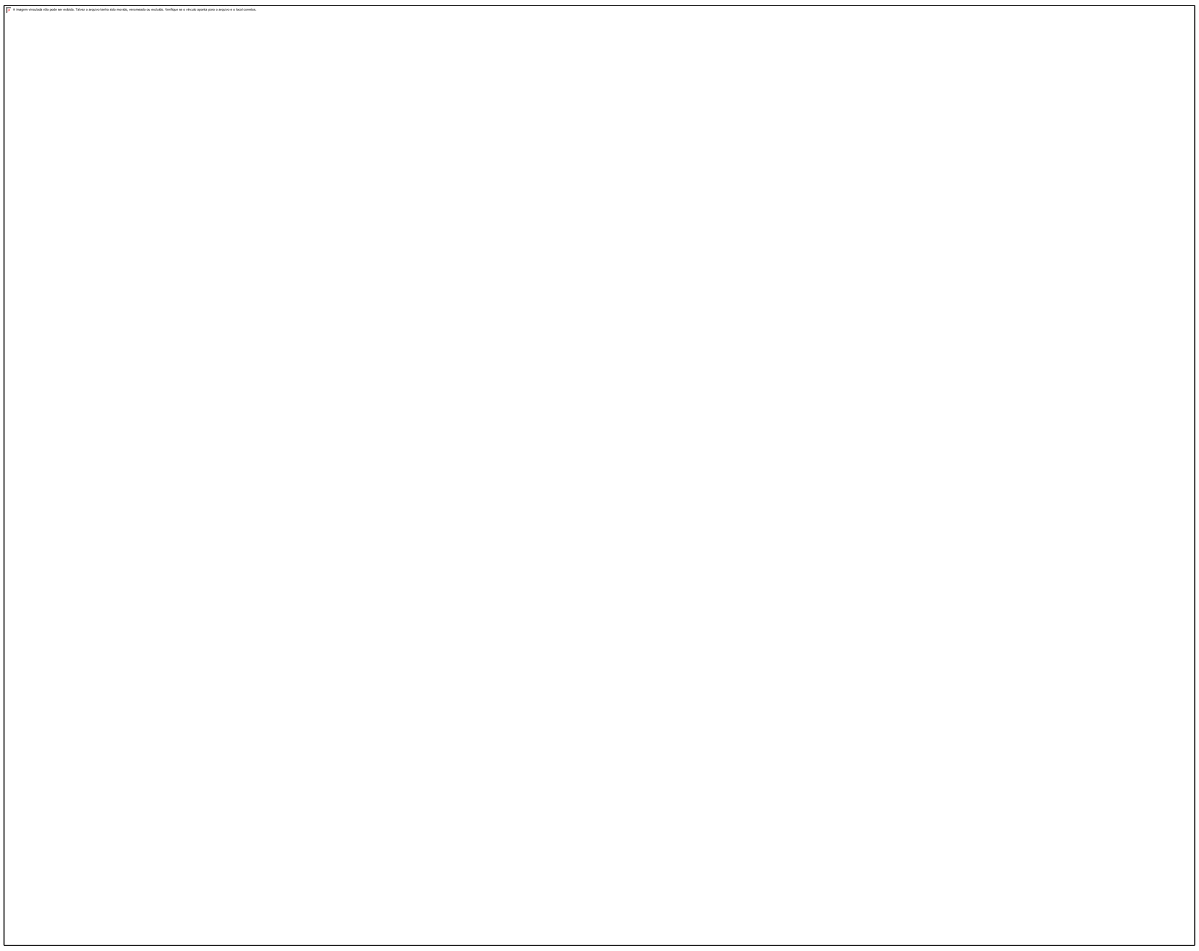
IOT2050 garantiu a supervisão remota e a transição dos dados para a Oracle Cloud sem comprometer o tempo de ciclo do controlador local.
A bancada encontra-se totalmente operacional e modernizada. O desenvolvimento da interface gráfica, a estruturação dos logs de dados e o estudo das métricas de rede e QoS estabelecem uma base sólida para futuros ensaios de controle avançado, testes com CLPs virtuais e implementação de novos protocolos da Indústria 4.0 no laboratório.
HABILIDADES ADQUIRIDAS:

• Operação e Controle de Processos: Domínio prático sobre malhas de controle em malha fechada (PID), incluindo a análise da resposta dinâmica de variáveis de processo e o ajuste de atuadores em plantas hidráulicas.
• Comunicação e Redes Industriais: Aprendizado prático na integração e configuração de protocolos da Tecnologia da Operação (TO) e da Informação (TI), como Modbus TCP e MQTT.
• Análise de Desempenho de Redes: Capacidade de avaliar métricas de telemetria ciber-física, como RTT e percentis de tempo de resposta em diferentes níveis de Qualidade de Serviço (QoS).
• Desenvolvimento de Interfaces IoT: Construção de dashboards gráficos de supervisão em tempo real utilizando a plataforma Node-RED em nuvem.
• Diagnóstico de Hardware/Software: Identificação e resolução de problemas em bancadas industriais integrando CLPs físicos (Altus XP340) e gateways de borda (IOT2050).

Referências

[1] BHATTACHARJEE, S. et al. Latency-Aware Function Placement, Routing, and Scheduling in TSN-based Industrial Networks. In: ICC 2022-IEEE International Conference on Communications. IEEE, 2022. p. 4248-4254.
[2] CRUZ, T.; SIMÕES, P.; MONTEIRO, E. Virtualizing Programmable Logic Controllers: Toward a Convergent Approach. IEEE Embedded Systems Letters, v. 8, n. 4, p. 69-72, 2016.
[3] LOU, D.; GRAF, U.; TSENG, M. Virtualized Programmable Logic Controllers: A Paradigm Shift Toward Industrial Edge and Cloud Computing. Industrial Internet Consortium, 2021. Disponível em: https://www.iiconsortium.org/pdf/IIC-Edge-vPLC-Tech-Brief-20210907.pdf. Acesso em: 13 ago. 2026.
[4] MOŻARYN, J.; BOGUSZ, K.; JUSZCZYŃSKI, S. Development of PLC Based Fault Isolation and Remote IIoT Monitoring of Three Tank System. IFAC-PapersOnLine, v. 55, n. 6, p. 175-180, 2022.
[5] GAVLAS, A.; ZWIERZYNA, J.; KOZIOREK, J. Possibilities of transfer process data from PLC to Cloud platforms based on IoT. IFAC-PapersOnLine, v. 51, n. 6, p. 156-161, 2018.
[6] BEHNKE, I.; AUSTAD, H. Real-Time Performance of Industrial IoT Communication Technologies: A Review. IEEE Access, 2023.
[7] ALTUS SISTEMAS DE AUTOMAÇÃO. Base de Conhecimento: Saiba como utilizar o protocolo MQTT na Série Nexto. São Leopoldo: Altus. Disponível em: https://www.altus.com.br. Acesso em: 13 ago. 2026.
[8] SIEMENS AG. SIMATIC IOT2050: Operating Instructions. Munique: Siemens, 2023. Disponível em plataforma técnica da Siemens. Acesso em: 13 ago. 2026.

Anexos



Bancada de trabalho

CÓDIGO: ET0743

AUTOR: HECTOR HENDRIO GOMES ARAUJO

COAUTOR: DANIEL LOPES MARTINS

COAUTOR: HEITOR MEDEIROS FLORENCIO

ORIENTADOR: ADRIAO DUARTE DORIA NETO

TÍTULO: Estudo de Técnicas de Visão Estereoscópica Embarcada em Dispositivos IoT para Aumento da Confiabilidade no Monitoramento de Variáveis em Plantas Industriais

Resumo

Este trabalho apresenta uma abordagem não invasiva para a medição de nível em tanques cilíndricos baseada em visão computacional estereoscópica e inteligência artificial embarcada. O sistema utiliza duas câmeras e uma rede neural convolucional siamesa para estimar diretamente o nível do líquido a partir de um par de imagens, sem reconstrução geométrica 3D explícita. O modelo foi embarcado no microcontrolador ESP32-S3 com TensorFlow Lite for Microcontrollers e quantização Int8. Os resultados demonstraram erro absoluto médio (MAE) de 2,93 mm na faixa de 30 cm, superando a abordagem monocular. A quantização reduziu o modelo em 73% (404 KB), alcançando tempo de ciclo médio de 657 ms, validando a viabilidade técnica para monitoramento industrial de baixo custo.

Palavras-chave: Visão computacional; Redes siamesas; TinyML; Medição não invasiva.

TITLE: Study of Embedded Stereoscopic Vision Techniques in IoT Devices to Enhance Variable Monitoring Reliability in Industrial Plants

Abstract

This work presents a non-invasive approach for level measurement in cylindrical tanks based on stereoscopic computer vision and embedded artificial intelligence. The system employs two cameras and a Siamese convolutional neural network to directly estimate the liquid level from a pair of images, without explicit 3D geometric reconstruction. The model was deployed on the ESP32-S3 microcontroller using TensorFlow Lite for Microcontrollers and Int8 quantization. The experimental results demonstrated a mean

absolute error (MAE) of 2.93 mm over a 30 cm range, outperforming the monocular approach. Quantization reduced the model size by 73% (404 KB) while achieving an average cycle time of 657 ms, validating the technical feasibility for low-cost industrial monitoring.

Keywords: Computer vision; Siamese networks; TinyML; Non-invasive measurement.

Introdução

A medição de nível de tanques é um elemento central em processos industriais, sendo essencial para a garantia da segurança operacional, do controle eficiente e da otimização de recursos. Tradicionalmente, essa tarefa é realizada por sensores dedicados por contato, como sensores de pressão diferencial que inferem a altura do líquido a partir da coluna de pressão hidrostática (LIPTAK, 2003). Contudo, essas tecnologias consolidadas apresentam limitações operacionais severas em ambientes com alta corrosividade, presença de incrustações, espumas, vapores ou materiais abrasivos, fatores que degradam os instrumentos e elevam os custos de manutenção (FREITAS, 2013).

No paradigma da Indústria 4.0 e da Internet das Coisas (IoT), cresce a demanda por soluções não invasivas, de baixo custo e com manutenção reduzida. A visão computacional embarcada destaca-se como alternativa viável ao inferir variáveis físicas a partir de imagens ópticas externas, eliminando qualquer contato com o fluido. Trabalhos recentes aplicam redes neurais convolucionais (CNNs) para regressão visual de nível (YANG et al., 2024). Entretanto, arranjos com uma única câmera (visão monocular) são altamente vulneráveis a erros de paralaxe decorrentes de desalinhamentos angulares entre a lente e o menisco da água (BOBOVNIK et al., 2021).

A visão computacional estereoscópica contorna essa vulnerabilidade ao capturar informações espaciais complementares de profundidade a partir de múltiplos pontos de observação (BARNARD; FISCHLER, 1982; SZELISKI, 2022). Enquanto abordagens tradicionais utilizam estereopsia e casamento de blocos para gerar mapas de disparidade e posterior reconstrução 3D explícita (ZBONTAR; LECUN, 2016; YANG et al., 2024) — o que demanda alto custo computacional —, este trabalho propõe uma abordagem end-to-end. Utiliza-se uma arquitetura de CNN siamesa que correlaciona diretamente as duas perspectivas ópticas com o valor escalar de nível.

Paralelamente, o avanço de técnicas de TinyML e frameworks otimizados, como o TensorFlow Lite for Microcontrollers, aliado a microcontroladores de baixo custo com aceleração vetorial e memórias expandidas, como o ESP32-S3, possibilita a inferência local com baixa latência e total independência de nuvem (DAVID et al., 2021; JACOB et al., 2018).

O objetivo deste trabalho é desenvolver e validar experimentalmente um sistema de medição de nível não invasivo baseado em visão estereoscópica e CNN siamesa implementado em ESP32-S3, avaliando o impacto da quantização Int8 na acurácia e na latência em relação à visão monocular.

Metodologia

1. Aparato Experimental

A bancada de testes consistiu em um kit didático Quanser acoplado, utilizando o Tanque 1 (cilíndrico, acrílico transparente, faixa útil de 0 a 300 mm). A água foi recirculada por meio de bomba controlada via PID no ESP32 (Figura 1, em Anexos). Como sistema sensorial óptico, empregou-se um arranjo estereoscópico composto por duas câmeras OV3660 (campo de visão de 65°), espaçadas com uma linha de base (baseline) de 11 cm entre eixos ópticos, posicionadas a uma distância frontal de 50 cm e altura de 60 cm em relação à base da bancada (Figura 2, em Anexos).

A arquitetura de processamento foi distribuída em três nós ESP32-S3 (Dual-Core LX7, 16 MB Flash, 8 MB PSRAM). Dois dispositivos atuaram como nós de captura de imagem RGB e o terceiro como nó concentrador e processador. O nó central requisitou os quadros via protocolo HTTP sobre rede local Wi-Fi, armazenou os dados nos buffers da PSRAM, alimentou o modelo de aprendizado profundo e transmitiu a predição via interface Serial/UART.

2. Construção do Dataset

Foram coletadas amostras ao longo de toda a extensão de 30 cm do tanque (intervalos de 1 cm). Para cada nível, adquiriram-se 10 pares estéreo: 5 amostras em regime estacionário (bomba desligada) e 5 amostras sob perturbação superficial dinâmica (bomba ligada com oscilações no menisco), totalizando 300 pares estéreo (600 imagens). A referência do nível foi aferida pela escala milimétrica do painel traseiro do aparato. As imagens foram capturadas em resolução de 128x128 pixels RGB em modo de exposição automática e sob variações de luminosidade ambiente natural e artificial ao longo do dia, assegurando robustez óptica. O dataset foi particionado aleatoriamente em 60% para treino (180 pares), 20% para validação (60 pares) e 20% para teste (60 pares).

3. Arquitetura da Rede Siamesa e Treinamento

O modelo foi projetado com dois ramos convolucionais idênticos em paralelo, com compartilhamento total de pesos (weight sharing) (Figura 3, em Anexos). Cada ramo convolucional foi composto por camadas de convolução 2D intercaladas com Max Pooling para extração de mapas de características invariantes e redução dimensional. As saídas de ambos os ramos foram vetorizadas (Flatten) e combinadas por uma camada de concatenação (Concatenate), seguida por camadas densas (Fully Connected) com ativação linear na saída para regressão contínua.

O treinamento foi executado em ambiente TensorFlow/Keras por 150 épocas, tamanho de lote (batch size) de 32, otimizador Adam com taxa de aprendizado de 0,0005 e função de perda Erro Quadrático Médio (MSE).

4. Otimização e Embarque (TinyML)

Após o treinamento, o modelo foi exportado e convertido para formato TFLite com quantização híbrida pós-treinamento: pesos e operações internas convertidos para inteiros de 8 bits (Int8) e tensores de entrada mantidos em Float32. O modelo quantizado foi integrado ao microcontrolador ESP32-S3 via framework TensorFlow Lite for Microcontrollers (LiteRT), avaliando-se o tamanho de memória ocupada, tempo de ciclo e métricas de regressão (MSE, MAE e RMSE) no conjunto de teste.

Resultados e Discussões

1. Desempenho do Aprendizado e Acurácia de Predição

Durante o treinamento, as curvas de Erro Absoluto Médio (MAE) e de perda (Loss - MSE) exibiram convergência suave e estabilização progressiva ao longo das 150 épocas, sem divergência acentuada entre treino e validação, indicando adequação da capacidade da rede em relação à complexidade do dataset (Figura 4, em Anexos). O modelo Float32 alcançou no treino um Loss de 9,41 e MAE de 2,52 mm; na validação, Loss de 11,48 e MAE de 2,80 mm; e, no conjunto de teste, Loss de 11,78 e MAE de 2,73 mm.

A análise comparativa entre os valores reais da escala e os preditos pela rede demonstrou forte alinhamento ao longo da reta ideal de 45° (Figura 5, em Anexos). O estudo dos erros residuais indicou dispersão contida na faixa central da coluna de líquido (Figura 6, em Anexos). Os maiores desvios residuais observados foram pontuais (+12,8 mm e -18,2 mm), concentrados nos níveis superiores (acima de 25 cm). Esse comportamento é atribuído à redução da margem de enquadramento superior da câmera associada a reflexos ópticos na borda do reservatório e às oscilações superficiais sob refluxo da bomba.

2. Impacto da Quantização Int8 e Desempenho no ESP32-S3

A aplicação da quantização pós-treinamento para Int8 resultou em expressiva otimização de recursos de hardware. O modelo original em ponto flutuante (Float32) ocupava 1.495 KB de memória, enquanto o modelo quantizado em Int8 foi reduzido para 404 KB, representando uma diminuição de 73% no espaço de armazenamento Flash/PSRAM.

A degradação métrica decorrente da redução de precisão foi desprezível diante do ganho em eficiência computacional:

Float32 Original: MSE = 11,78 | MAE = 2,73 mm | RMSE = 3,43 mm

Int8 Quantizado: MSE = 12,46 | MAE = 2,93 mm | RMSE = 3,53 mm

O aumento no MAE foi de apenas 0,20 mm (7,3%), mantendo o erro médio abaixo de 3 mm em uma coluna d'água de 300 mm (erro relativo inferior a 1% do fundo de escala).

A avaliação do tempo de ciclo de processamento em tempo real no ESP32-S3 — compreendendo as requisições HTTP locais, transferência das matrizes para a PSRAM e execução da inferência da rede convolucional siamesa — registrou tempo médio total de 657 ms. O sistema demonstrou capacidade de atualização dinâmica com frequência superior a 1,5 Hz, atendendo amplamente aos requisitos de controle de tanques de processos e reservatórios industriais.

3. Comparação com a Abordagem Monocular

Para verificar o benefício da estereopsia frente ao arranjo tradicional de uma câmera, a abordagem monocular foi reavaliada sob idênticas condições de bancada, hiperparâmetros e quantização Int8:

Visão Monocular (Int8): MSE = 28,15 | MAE = 4,52 mm | RMSE = 5,31 mm

Visão Estereoscópica Siamesa (Int8): MSE = 12,46 | MAE = 2,93 mm | RMSE = 3,53 mm

A abordagem estereoscópica proporcionou uma redução de 35,2% no MAE e de 33,5% no RMSE quando comparada à configuração monocular. Esse ganho quantitativo valida a hipótese de que a redundância de perspectiva e a percepção implícita de profundidade fornecidas pelos dois ramos convolucionais mitigam as distorções causadas por erro de paralaxe e variações de iluminação especular, tornando a regressão significativamente mais precisa sem a necessidade de etapas explícitas e custosas de triangulação geométrica 3D.

Conclusão

Este trabalho demonstrou a viabilidade técnica do uso de visão computacional estereoscópica associada a redes neurais convolucionais siamesas para a medição não invasiva de nível de líquidos em dispositivos de borda IoT. A arquitetura proposta aprendeu a regressão direta do nível a partir de dois pontos de vista, alcançando Erro Absoluto Médio (MAE) de 2,93 mm e superando a visão monocular em mais de 35% de precisão sob as mesmas condições experimentais.

O emprego de técnicas de TinyML viabilizou a redução de 73% no tamanho do modelo (404 KB) com quantização Int8, permitindo sua completa execução no microcontrolador ESP32-S3. O tempo total de resposta de 657 ms por ciclo confirmou o potencial da solução para controle e monitoramento industrial em tempo real com hardware de baixo custo.

Como trabalhos futuros, propõe-se expandir o dataset para contemplar variações angulares dinâmicas das câmeras, fluidos com turbidez e colorações distintas, bem como conduzir validações cruzadas com coletas temporais totalmente desacopladas, aumentando a capacidade de generalização do modelo em campo.

Referências

BARNARD, S. T.; FISCHLER, M. A. Computational stereo. ACM Computing Surveys (CSUR), v. 14, n. 4, p. 553–572, 1982.

BOBOVNIK, G. et al. Liquid level measurement based on computer vision and parallax error correction. Measurement, v. 186, p. 110196, 2021.

DAVID, R. et al. TensorFlow Lite Micro: Embedded machine learning on TinyML systems. Proceedings of Machine Learning and Systems, v. 3, p. 800–811, 2021.

FREITAS, C. G. Sensor de nível para líquidos corrosivos baseado em guias de onda ópticos. 2013. Dissertação (Mestrado em Engenharia Elétrica) – Universidade Federal do Rio Grande do Norte, Natal, 2013.

JACOB, B. et al. Quantization and training of neural networks for efficient integer-arithmetic-only inference. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018. p. 2704–2713.

LIPTAK, B. G. Instrument Engineers' Handbook: Process Measurement and Analysis. 4. ed. Boca Raton: CRC Press, 2003. v. 1.

SZELISKI, R. Computer Vision: Algorithms and Applications. 2. ed. Cham: Springer, 2022.

YANG, X. et al. Deep learning-based stereoscopic computer vision for sea level measurement. Ocean Engineering, v. 293, p. 116645, 2024.

ZBONTAR, J.; LECUN, Y. Stereo matching by training a convolutional neural network to compare image patches. Journal of Machine Learning Research, v. 17, n. 1, p. 2287–2318, 2016.

Anexos



Figura 1 - Planta experimental

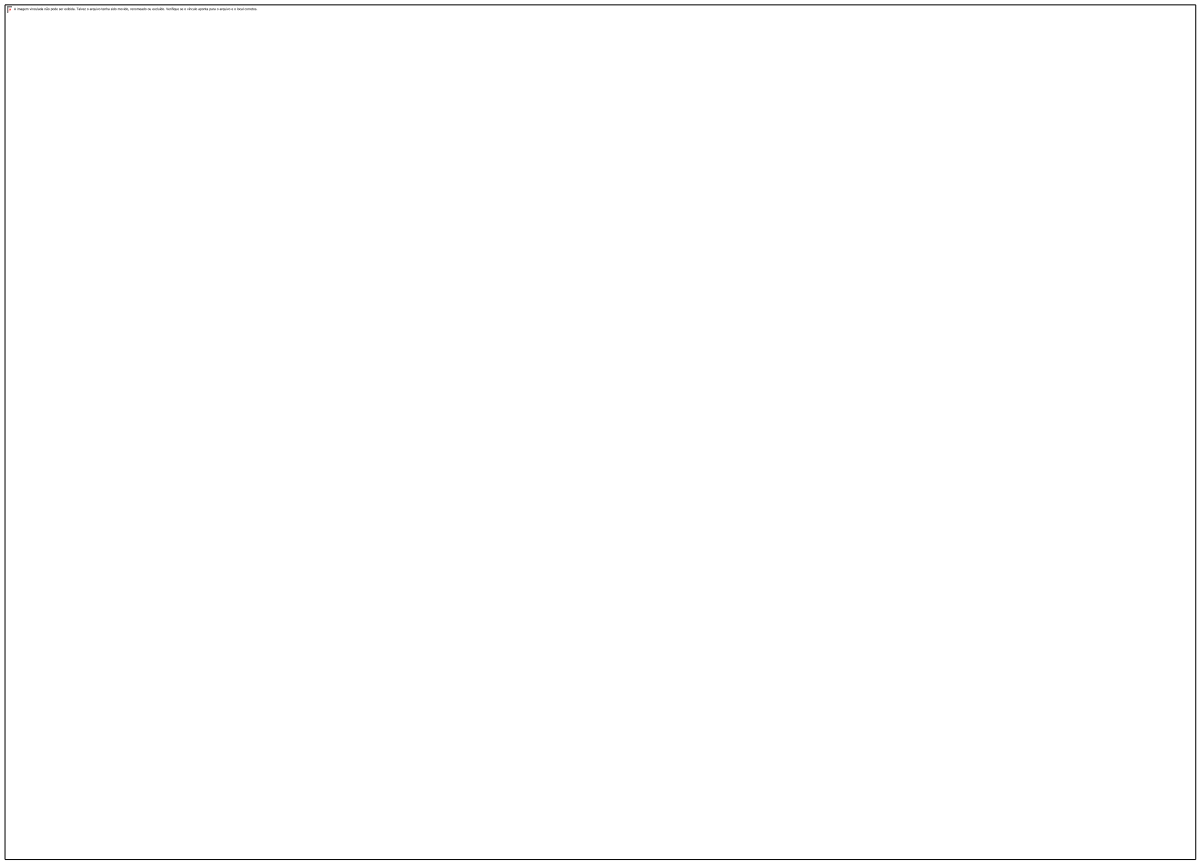


Figura 2 - Estrutura das duas câmeras



Figura 3 - Arquitetura da rede neural siamesa

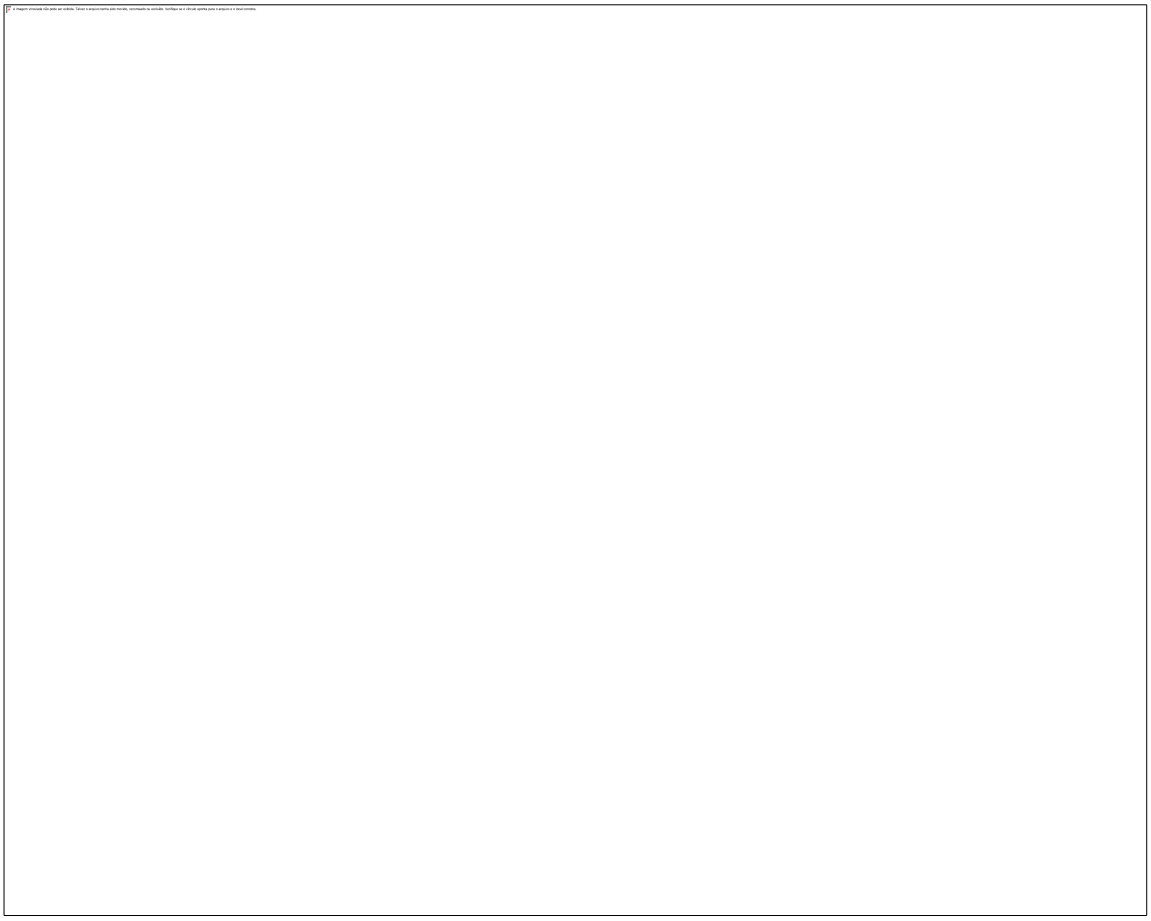


Figura 4 - Gráfico do treinamento

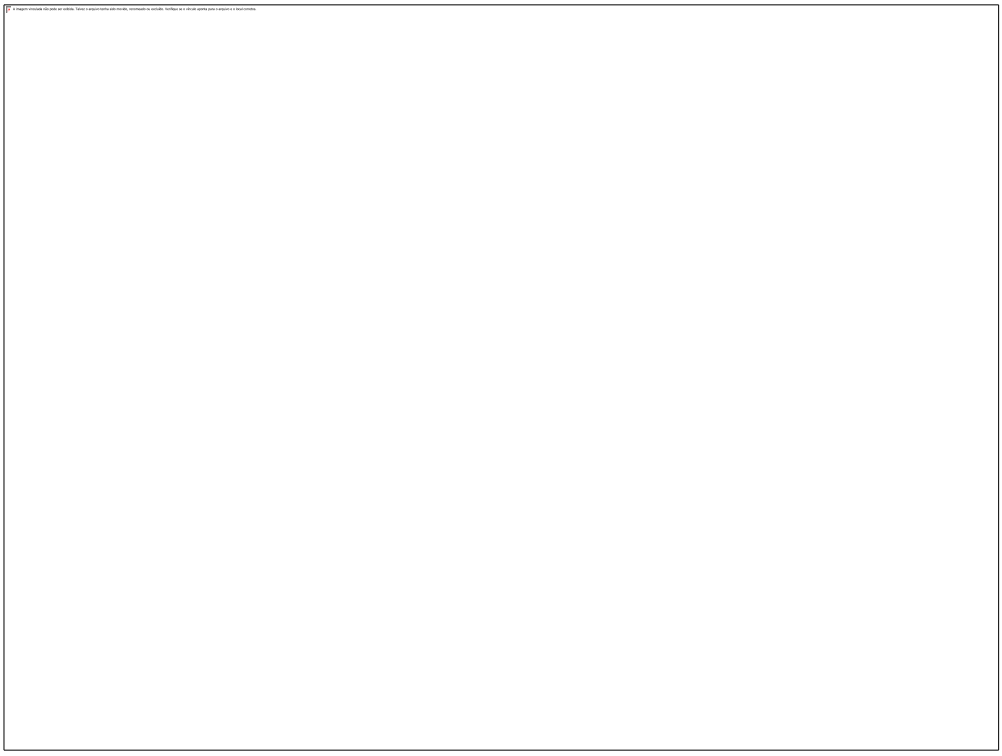


Figura 5 - Gráfico previsto versus real

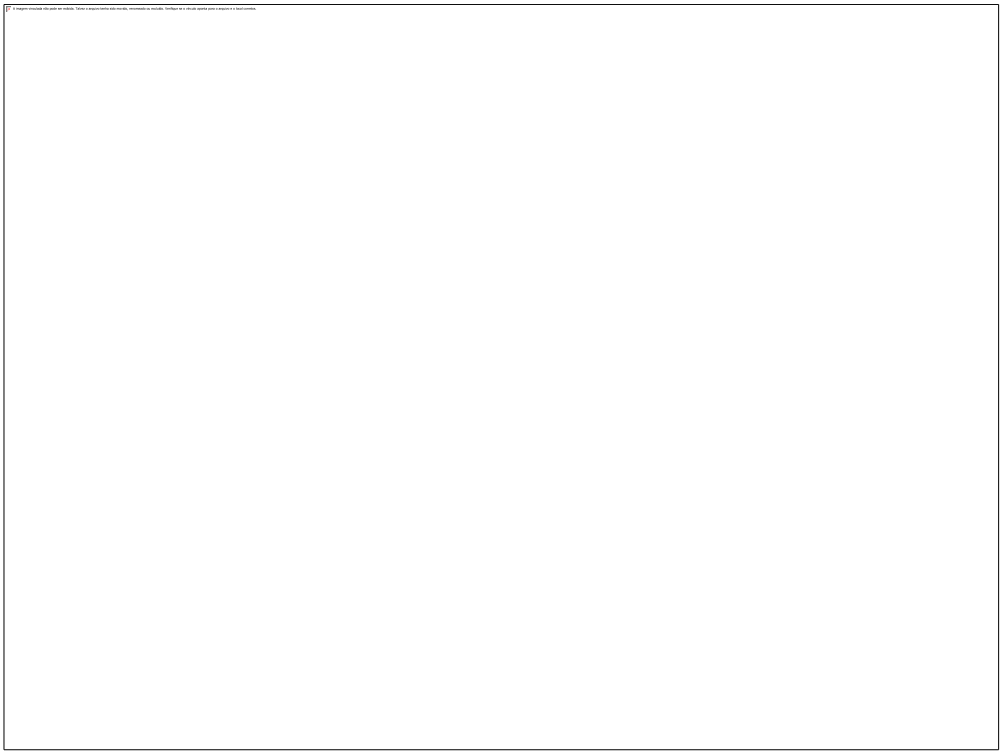


Figura 6 - Gráfico erro versus nível

CÓDIGO: ET0871

AUTOR: FELIPE DANTAS DE SOUZA MACEDO

ORIENTADOR: ROGER KREUTZ IMMICH

TÍTULO: Vulnerabilidades, Ataques e Mitigações em Smart Contracts no Contexto de IoT: Uma Análise da Literatura

Resumo

Smart contracts são empregados em aplicações de Internet das Coisas (IoT) para automatizar decisões, registrar eventos e mediar interações entre dispositivos, serviços e blockchains. Entretanto, falhas no código, na execução contratual ou nas camadas que condicionam seu estado podem comprometer segurança, privacidade e confiabilidade. Este trabalho apresenta uma análise da literatura sobre vulnerabilidades e ataques a smart contracts em contextos IoT, técnicas de análise, detecção, verificação, mitigação ou defesa e o papel do próprio contexto IoT. A busca em IEEE, Web of Science, ScienceDirect e Scopus consolidou 319 registros e levou 22 estudos à etapa de extração. Para esta versão, a síntese foi restrita a oito estudos com leitura integral e extração concluídas. Os achados abrangem ataques ao consenso, vulnerabilidades aritméticas e de código, contratos maliciosos, falhas de interação e controle de acesso, exposição de privacidade e manipulação de aleatoriedade. As respostas incluem limiares de consenso, bibliotecas preventivas, aprendizado profundo, verificação formal, assinaturas em anel, geração híbrida de aleatoriedade e padrões arquiteturais. As evidências são heterogêneas e indicam pouca validação fim a fim entre dispositivos físicos e execução contratual.

Palavras-chave: Smart contracts; IoT; segurança; análise da literatura

TITLE: Vulnerabilities, Attacks, and Mitigations in Smart Contracts in IoT Contexts: A Literature Analysis

Abstract

Smart contracts are used in Internet of Things (IoT) applications to automate decisions, record events, and mediate interactions among devices, services, and blockchains. However, flaws in contract code, execution, or underlying layers that condition contract state may compromise security, privacy, and reliability. This study presents a literature analysis of vulnerabilities and attacks affecting smart contracts in IoT contexts, the proposed analysis, detection, verification, mitigation, and defense techniques, and the role played by IoT itself. Searches in IEEE, Web of Science, ScienceDirect, and Scopus consolidated 319 records and led 22 studies to the extraction stage. For this version, the synthesis was restricted to eight studies whose full-text reading and extraction were

completed. Findings include consensus attacks, arithmetic and code vulnerabilities, malicious contracts, interaction and access-control failures, privacy exposure, and manipulable randomness. Countermeasures include consensus thresholds, preventive libraries, deep learning, formal verification, ring signatures, hybrid randomness generation, and architectural patterns. The evidence is heterogeneous and shows limited end-to-end validation from physical devices to smart-contract execution.

Keywords: Smart contracts; IoT; security; literature analysis

Introdução

Smart contracts são programas executados em infraestruturas blockchain capazes de aplicar regras e automatizar transações sem que cada operação dependa de uma autoridade central. Em cenários de Internet das Coisas (IoT), esses contratos podem receber eventos produzidos no mundo físico, controlar acesso, registrar estados e coordenar serviços distribuídos. Essa aproximação conecta, porém, ambientes com propriedades distintas: de um lado, código replicado e persistente em blockchain; de outro, dispositivos, sensores, gateways e redes sujeitos a restrições de recursos, falhas físicas e condições variáveis (FOTIOU; POLYZOS, 2018; GUO, 2025).

A segurança dessa integração não se limita a bugs tradicionais de programação. Erros aritméticos podem afetar o estado de contratos; falhas de lógica podem persistir após a implantação; a transparência de blockchains públicas pode expor informações sensíveis; contratos podem depender de entradas físicas ou aleatórias manipuláveis; e ataques à camada de consenso podem interferir nas transações reconhecidas pela plataforma. As respostas técnicas também são diversas, incluindo bibliotecas preventivas, análise e classificação de código, verificação formal, mecanismos criptográficos e modificações arquiteturais.

Outro problema é distinguir quando IoT realmente condiciona a segurança e quando aparece apenas como domínio de aplicação. Em uma rede Home IoT com poucos participantes, por exemplo, o número reduzido de nós altera diretamente a viabilidade de um ataque majoritário (MOON et al., 2022). Em contraste, overflow e underflow permanecem vulnerabilidades genéricas de contratos Solidity mesmo quando os contratos avaliados pertencem a aplicações IoT (KHOR et al., 2020). Essa distinção evita atribuir especificidade ao domínio apenas pela presença do termo IoT.

Diante desse cenário, o objetivo deste trabalho é caracterizar vulnerabilidades e ataques que afetam smart contracts em contextos IoT, identificar as técnicas propostas para analisá-los ou tratá-los, examinar de que forma IoT condiciona o problema, a solução ou a avaliação e apontar padrões e lacunas no conjunto efetivamente analisado. O texto resulta de um levantamento bibliográfico estruturado e de uma síntese qualitativa dos estudos cuja análise foi concluída, sem pretensão de representar integralmente o conjunto de publicações inicialmente selecionado.

Metodologia

O levantamento bibliográfico foi organizado com base no PICOC como instrumento de planejamento da busca. A população compreendeu smart contracts e tecnologias associadas, incluindo Solidity, chaincode, Ethereum, EVM e aplicações descentralizadas. A intervenção reuniu termos relacionados a vulnerabilidades, ataques, exploits, ameaças, falhas, análise de segurança, detecção, verificação formal, auditoria, mitigação e defesa. O contexto abrangeu IoT e cenários associados, como IIoT, sistemas ciberfísicos, redes de sensores, smart grid, smart home, smart city, dispositivos embarcados, edge e gateways. O componente Comparison não foi utilizado.

Foram pesquisadas as bases IEEE, Web of Science, ScienceDirect e Scopus. As strings combinaram três blocos conceituais: smart contracts e plataformas relacionadas; vulnerabilidades, ataques ou mecanismos de segurança; e IoT ou contextos correlatos. A sintaxe foi adaptada a cada base. Foram considerados trabalhos em inglês, publicados entre 2015 e 2026 como artigos de periódico ou trabalhos completos de conferência, conforme os filtros disponíveis em cada fonte.

Os critérios de inclusão exigiram que o estudo investigasse falha ou ataque ao smart contract, ou à plataforma quando o efeito comprometesse execução, estado, validade transacional ou consenso relevante aos contratos; apresentasse vínculo substantivo com IoT; apresentasse mecanismo técnico de análise, detecção, verificação, mitigação ou defesa; e fosse estudo primário, completo, revisado por pares e elegível no período definido. Foram excluídos trabalhos em que o smart contract era apenas instrumento para proteger outro ativo sem comprometimento de sua execução ou estado, estudos em que IoT aparecia apenas superficialmente, revisões e surveys, materiais documentais inelegíveis, estudos sem tratamento técnico da segurança de smart contracts e publicações retratadas.

A triagem consolidou 319 registros e manteve 22 estudos como candidatos à etapa de leitura e extração. Esses 22 não correspondem, porém, ao conjunto efetivamente analisado neste texto. A ordem de trabalho foi definida de forma operacional: foram priorizados os artigos cujo texto integral podia ser consultado e conferido conjuntamente durante o processo assistido por IA, o que fez com que alguns estudos aos quais o pesquisador já possuía acesso institucional fossem adiados. Outros textos tiveram o acesso localizado ou restabelecido apenas no encerramento do trabalho. Além disso, pelo menos um estudo inicialmente selecionado deixou de satisfazer o recorte após leitura integral. Diante do prazo, não se tentou completar retrospectivamente todos os 22: para manter um procedimento uniforme e evitar respostas construídas apenas a partir de resumos, a síntese foi fechada com oito estudos que haviam passado por leitura integral, formulação das respostas às RQs e validação contra o texto. Portanto, a diferença entre 22 e oito reflete sobretudo incompletude da etapa de extração e decisões de priorização, e não falta de acesso aos demais artigos.

A extração foi orientada pelas seguintes questões: RQ1 - Quais vulnerabilidades e ataques afetam smart contracts em contextos IoT e por quais mecanismos eles se manifestam? RQ2 - Quais técnicas de análise, detecção, verificação, mitigação ou defesa foram propostas, e quais vulnerabilidades ou ataques elas tratam? RQ3 - De que maneira o contexto IoT condiciona o problema de segurança, a solução proposta ou sua

avaliação nos estudos incluídos? RQ4 - Que padrões de cobertura e quais lacunas aparentes emergem da relação entre vulnerabilidades, ataques e técnicas propostas, considerando os contextos ou componentes IoT quando aplicável?

Para cada estudo sintetizado, foram registrados o problema de segurança, a contribuição técnica, o papel de IoT, a forma de avaliação, os resultados e as limitações relevantes. A RQ4 foi respondida por síntese transversal, e não individualmente. Os experimentos originais não foram reproduzidos: resultados de simulações, protótipos, análise de gas, classificadores, verificação formal e provas analíticas foram interpretados conforme reportados pelos autores, preservando as diferenças metodológicas.

Ferramentas de IA generativa tiveram participação substancial em todo o desenvolvimento do trabalho, e não apenas na revisão linguística do texto. Elas foram usadas na formulação inicial do problema e do protocolo, na construção e adaptação das strings de busca, na organização dos critérios de inclusão e exclusão, na pré-classificação e triagem dos 319 registros, na estruturação das matrizes de trabalho, na explicação de conceitos técnicos, na localização de evidências dentro dos artigos, na conferência das respostas às questões de pesquisa e na redação e revisão do manuscrito. A triagem foi conduzida de forma híbrida: a IA produziu classificações e justificativas preliminares, enquanto o pesquisador revisou os resultados e dedicou maior atenção aos casos fronteiriços; em parte dos lotes, por limitação de tempo, classificações consideradas claras foram aceitas com conferência humana menos aprofundada. Na etapa de extração usada nesta síntese, o pesquisador leu os textos integrais dos oito estudos, formulou interpretações preliminares e utilizou a IA para confrontá-las com o artigo, localizar trechos de suporte e apontar possíveis extrapolações. A decisão final sobre inclusão e interpretação permaneceu com o pesquisador, mas a dependência de IA foi metodologicamente relevante e constitui uma ameaça à validade, especialmente na triagem e na interpretação de estudos tecnicamente complexos.

Resultados e Discussões

RQ1 - Vulnerabilidades, ataques e mecanismos de manifestação

Os oito estudos sintetizados mostram que os problemas de segurança se distribuem por diferentes camadas. No nível da infraestrutura, Moon et al. (2022) estudam um ataque de 51% em uma plataforma blockchain para Home IoT. O problema não é um bug do código contratual, mas o controle majoritário do consenso, capaz de interferir na formação de blocos e nas transações reconhecidas pela plataforma. Sua relevância para smart contracts decorre do fato de que a camada de consenso condiciona quais operações e estados são aceitos.

No código Solidity, Khor et al. (2020) concentram-se em integer overflow e integer underflow e discutem ainda o problema denominado pelos autores como unprotected function, associado ao uso de if() em operações que falham sem o comportamento de reversão obtido com require(). Huo (2025), por sua vez, trabalha com um conjunto mais amplo de classes para detecção automática: reentrância, integer overflow, integer

underflow, predictable variable attack e abnormal state. Shah et al. (2023) tratam contratos vulneráveis ou maliciosos de forma binária, sem discriminar no experimento principal categorias tradicionais de vulnerabilidade.

Outros estudos deslocam o foco da sintaxe do código para a interação e para entradas externas. Guo (2025) procura verificar falhas de controle de acesso e vulnerabilidades sutis que podem emergir da sequência de chamadas e da interação entre múltiplos contratos, sem restringir a contribuição a uma taxonomia tradicional de bugs. Chodavarapu et al. (2024) tratam do risco de manipulação ou viés da aleatoriedade consumida por dApps quando participantes responsáveis por fornecer entropia são comprometidos.

A transparência da blockchain também aparece como fonte de risco. Mahmoud et al. (2018) mostram que, em seguros acionados por sensores IoT, a publicidade dos dados on-chain pode permitir associar um evento físico ao titular da apólice, produzindo perda de privacidade. Fotiou e Polyzos (2018) discutem de modo mais geral que código, estado e histórico publicamente observáveis limitam o armazenamento de informações secretas e que a imutabilidade torna falhas de lógica implantadas difíceis de corrigir. Nesse último caso, trata-se de análise conceitual e de diretrizes, e não de vulnerabilidade reproduzida experimentalmente.

Assim, a RQ1 não revela uma única classe dominante no conjunto sintetizado. Os achados abrangem consenso, aritmética, lógica contratual, interação entre contratos, privacidade e qualidade de entradas externas. A superfície de segurança relevante, portanto, não termina no bytecode: propriedades da plataforma e da ponte entre o mundo físico e o contrato também podem determinar o resultado da execução.

RQ2 - Técnicas de análise, detecção, verificação, mitigação ou defesa

As respostas técnicas acompanham essa diversidade. Para o ataque majoritário em Home IoT, Moon et al. (2022) propõem que o usuário defina um limiar de consenso superior ao padrão considerado no estudo. Quando a concordância não alcança o limiar, as transações são descartadas, o bloco não é criado e a rede é interrompida até nova autenticação. A simulação em MATLAB com 11 nós indica redução da probabilidade de comprometimento à medida que o limiar aumenta, mas evidencia um trade-off de disponibilidade, pois a rede também pode ser interrompida na ausência de ataque.

Para vulnerabilidades aritméticas, Khor et al. (2020) avaliam SafeMath e propõem uma Improved SafeMath com maior cobertura declarada e menor consumo de gas. Os autores relatam nove operações cobertas, duas condições gerais de verificação e economia de 26 unidades de gas por transação em relação à biblioteca original. Mythril e Oyente deixaram de reportar overflow e underflow nos contratos analisados após a integração das bibliotecas, mas isso não constitui prova de ausência completa de falhas. SmartCheck, Securify, Mythril e Oyente são ferramentas utilizadas na avaliação, e não contribuições propostas pelo estudo.

Os trabalhos de detecção por aprendizado de máquina seguem outra estratégia. Huo (2025) utiliza uma Tree-based Convolutional Neural Network aprimorada sobre a árvore sintática do contrato, com um mecanismo de continuous binary tree para

representar subárvores de tamanhos variáveis. Shah et al. (2023) utilizam LSTM, GRU e ANN sobre representações derivadas de opcodes para classificar contratos como vulneráveis ou não vulneráveis. Embora ambos reportem desempenho elevado, seus desenhos, classes e conjuntos de dados não são equivalentes; por isso, valores de acurácia e F1 não permitem estabelecer um ranking direto entre as técnicas.

Guo (2025) segue a linha de verificação formal. O comportamento de contratos Solidity é modelado por Transition Systems e propriedades em Computation Tree Logic, com tradução para o model checker Verds. A proposta verifica tanto contratos individuais quanto interações e sequências de chamadas, buscando falhas que poderiam não aparecer em análises isoladas. A avaliação inclui casos de controle de acesso e cadeia de suprimentos, além de medidas de tempo e complexidade; entre as limitações estão explosão de estados, sobrecarga computacional e dependência da linguagem e da ferramenta.

Nas propostas criptográficas, Mahmoud et al. (2018) integram assinaturas em anel ligáveis ao contrato de seguro. O mecanismo permite verificar que um sensor autorizado gerou o evento sem revelar qual sensor do conjunto foi responsável, enquanto a propriedade de linkability permite detectar reutilização da mesma chave secreta. O protótipo Ethereum avalia custo de gas e mostra redução substancial da verificação com contratos pré-compilados. Chodavarapu et al. (2024) propõem um Hybrid Random Number Generator que agrega fontes de aleatoriedade provenientes de múltiplos dispositivos IoT com Trusted Execution Environments, combinando mecanismos criptográficos e agregação on-chain; a segurança analítica depende de ao menos uma fonte honesta para evitar viés no resultado.

Fotiou e Polyzos (2018) propõem explicitamente um skeleton-based design: um contrato base mantém a interface e referencia contratos de implementação, permitindo implantar uma nova implementação e atualizar o apontamento quando uma falha é descoberta. Os autores também recomendam manter dados sensíveis off-chain com hashes on-chain, empregar análise automática e prever um kill switch, mas essas recomendações não devem ser confundidas com técnicas necessariamente originais do artigo. O estudo é conceitual e não apresenta benchmark empírico para o padrão skeleton.

RQ3 - Papel do contexto IoT

O papel de IoT varia significativamente. Em alguns estudos, retirar o contexto IoT modificaria o próprio problema ou a solução. Moon et al. (2022) justificam o risco do ataque majoritário pelo pequeno número de nós em uma rede doméstica. Mahmoud et al. (2018) dependem do vínculo entre um sensor físico, o evento de dano e o titular do seguro para caracterizar a perda de privacidade. Chodavarapu et al. (2024) utilizam dispositivos IoT com TEE como fontes físicas distribuídas de aleatoriedade. Fotiou e Polyzos (2018) estruturam suas recomendações considerando dispositivos com recursos limitados, necessidade de gateways e a impossibilidade de um smart contract atuar diretamente no mundo físico.

Em Guo (2025), o cenário IoT torna relevantes interações entre vários dispositivos, usuários e contratos para controle de acesso, embora Transition Systems, CTL e model

checking sejam técnicas reutilizáveis fora de IoT. Em Shah et al. (2023), o contrato ocupa uma posição específica no fluxo de dados: antes de os dados IoT serem registrados, a execução contratual participa de sua validação, de modo que um contrato malicioso pode degradar a confiança do pipeline. Ainda assim, os classificadores empregados para contratos não dependem, em sua mecânica interna, de características exclusivas de IoT.

Dois casos evidenciam a diferença entre contexto substantivo e técnica genérica. Khor et al. (2020) selecionam contratos de blockchains direcionadas a aplicações IoT, mas overflow, underflow e SafeMath permanecem mecanismos gerais de Solidity/Ethereum. Huo (2025) incorpora dispositivos, gateways e camadas IoT à arquitetura geral, enquanto o detector baseado em AST/TCNN pode, em princípio, operar sobre contratos de outros domínios. Assim, a presença de IoT deve ser interpretada causalmente: ela pode moldar a ameaça, a fonte de dados, a arquitetura ou a avaliação, mas nem sempre altera o mecanismo técnico central.

RQ4 - Padrões de cobertura e lacunas

Quatro padrões emergem do conjunto sintetizado. Primeiro, há fragmentação entre camadas de segurança. Os estudos não se concentram somente em vulnerabilidades de código: consenso, privacidade, interação entre contratos e qualidade de entradas externas também aparecem como pontos de falha. Isso favorece uma visão fim a fim, na qual a correção do contrato é necessária, mas não suficiente para garantir o comportamento seguro de uma aplicação IoT.

Segundo, as técnicas propostas também são heterogêneas. Bibliotecas preventivas e padrões arquiteturais atuam antes ou durante a implantação; detectores por aprendizado profundo procuram classificar contratos; verificação formal examina propriedades e sequências de execução; mecanismos criptográficos tratam anonimato ou aleatoriedade; e mudanças no consenso atuam na plataforma. Essa diversidade torna inadequado comparar todos os estudos por uma única métrica.

Terceiro, as avaliações utilizam evidências pouco padronizadas: simulações probabilísticas, consumo de gas, métricas de classificação, casos de model checking, análises de ataque, provas analíticas e discussões conceituais. Mesmo quando dois trabalhos reportam acurácia ou custo, diferenças de dataset, classe, plataforma e ameaça limitam comparações diretas. No conjunto analisado, não foi identificado um benchmark comum capaz de reproduzir, sob o mesmo modelo de ameaça, o caminho completo desde o dispositivo ou sensor até o estado final do contrato.

Por fim, permanecem lacunas nas fronteiras físico-digitais. Algumas propostas assumem sensores resistentes a adulteração, TEE confiável, gateways ou pelo menos uma fonte honesta; outras não avaliam de forma integrada o comprometimento desses componentes com vulnerabilidades do contrato. Também há propostas conceituais sem validação empírica e estudos de aprendizado de máquina com limitações de rastreabilidade de dados, rótulos ou correspondência entre classes. Esses achados apontam para a necessidade de avaliações reprodutíveis que combinem vulnerabilidades contratuais, entradas físicas adversariais, restrições de recursos e custos da execução blockchain.

Conclusão

A síntese de oito estudos mostra que a segurança de smart contracts em contextos IoT é um problema transversal. Foram identificados ataques ao consenso, vulnerabilidades aritméticas e de lógica, contratos maliciosos, problemas de interação e controle de acesso, riscos de privacidade e manipulação de aleatoriedade. As respostas incluem limiares configuráveis de consenso, bibliotecas preventivas, detectores baseados em aprendizado profundo, verificação formal, assinaturas em anel ligáveis, geração híbrida de aleatoriedade e padrões de atualização arquitetural.

O contexto IoT não exerce o mesmo papel em todos os trabalhos. Em alguns casos, ele altera diretamente o risco, a fonte de confiança ou a arquitetura; em outros, apenas situa uma técnica que continua essencialmente genérica para smart contracts. Essa distinção é importante para interpretar o escopo das contribuições e evitar atribuir especificidade ao domínio apenas pela presença de IoT no cenário de aplicação.

A principal limitação deste trabalho é a incompletude da análise. Embora 22 estudos tenham permanecido após a triagem como candidatos à extração, somente oito foram incorporados à síntese porque apenas esses tiveram leitura integral, respostas às RQs e validação concluídas segundo o mesmo procedimento até o fechamento. Essa redução não significa que os outros 14 fossem necessariamente inacessíveis: vários eram acessíveis ao pesquisador, alguns foram deliberadamente adiados pela ordem de priorização adotada e outros tiveram o acesso identificado apenas no final. Por isso, as frequências e lacunas observadas não devem ser generalizadas para os 22 estudos selecionados na triagem. Soma-se a isso o uso substancial de IA em etapas de planejamento, triagem, interpretação e redação, o que pode introduzir erros de classificação ou leitura apesar das conferências humanas realizadas. Ainda assim, os oito estudos analisados permitem descrever, de forma exploratória, a diversidade de problemas e mecanismos encontrados e apontar questões que merecem investigação posterior.

Referências

CHODAVARAPU, R.; KARANJAI, R.; FAN, X.; SHI, W.; XU, L. Adding All Flavors: A Hybrid Random Number Generator for dApps and Web3. In: Stabilization, Safety, and Security of Distributed Systems (SSS), 2024. DOI: 10.1007/978-3-031-74498-3_21.

FOTIOU, N.; POLYZOS, G. C. Smart Contracts for the Internet of Things: Opportunities and Challenges. In: 2018 European Conference on Networks and Communications (EuCNC), 2018. p. 256-260. DOI: 10.1109/EuCNC.2018.8443212.

GUO, Z. Blockchain-enhanced smart contracts for formal verification of IoT access control mechanisms. Alexandria Engineering Journal, v. 118, p. 315-324, 2025. DOI: 10.1016/j.aej.2024.12.109.

HUO, X. Blockchain-Based Distributed Network Security Architecture with Smart Contract Vulnerability Detection Using Improved Tree CNN. Informatica, v. 49, n. 17, 2025. DOI: 10.31449/inf.v49i17.8050.

KHOR, J.; MASAMA, M. A.; SIDOROV, M.; LEONG, W.; LIM, J. An Improved Gas Efficient Library for Securing IoT Smart Contracts against Arithmetic Vulnerabilities. ACM International Conference Proceeding Series, 2020. DOI: 10.1145/3384544.3384577.

MAHMOUD, O.; KOPP, H.; ABDELHAMID, A. T.; KARGL, F. Applications of Smart-Contracts: Anonymous Decentralized Insurances with IoT Sensors. In: 2018 IEEE International Congress on Cybermatics, 2018. DOI: 10.1109/Cybermatics_2018.2018.00254.

MOON, H.-S.; SONG, J.; SHIN, H.; JANG, J. Home IoT device management blockchain platform using smart contracts and a countermeasure against 51% attacks. ACM International Conference Proceeding Series, 2022. DOI: 10.1145/3512353.3512381.

SHAH, H.; SHAH, D.; JADAV, N. K.; GUPTA, R.; TANWAR, S.; ALFARRAJ, O.; TOLBA, A.; RABOACA, M. S.; VERDES, M. Deep Learning-Based Malicious Smart Contract and Intrusion Detection System for IoT Environment. Mathematics, v. 11, n. 2, art. 418, 2023. DOI: 10.3390/math11020418.

CÓDIGO: ET1020

AUTOR: ISRAEL MARTINS VALCAREGGI

COAUTOR: LORENZO SANTOS FURTADO

ORIENTADOR: RENAN CIPRIANO MOIOLI

TÍTULO: Aperfeiçoamento de Arquiteturas de Redes Neurais Pulsadas para Processamento de Sinais de Eletrocardiograma (ECG)

Resumo

O seguinte trabalho aborda a aplicação de Redes Neurais Pulsadas (Spiking Neural Networks — SNNs) na classificação de sinais de eletrocardiograma (ECG), com foco na identificação de arritmias cardíacas. O estudo investiga a utilização de uma arquitetura SNN para processar sinais de ECG como se- quências temporais de disparos, explorando técnicas de pré-processamento, normalização e codificação dos sinais. Como modelo de referência, também é utilizado um algoritmo Random Forest, permitindo comparar uma aborda- gem tradicional de aprendizado de máquina com a rede neural pulsada. Além disso, são avaliadas diferentes estratégias para lidar com o desbalanceamento das classes, incluindo os dados originais, undersampling e técnicas de aug- mentation aplicadas à classe de arritmia. Os modelos são avaliados por meio de métricas como acurácia, precisão, recall e F1-score, com atenção especial à capacidade de identificar corretamente os casos de arritmia. Os resultados permitem analisar o desempenho das diferentes abordagens e as limitações das estratégias de balanceamento aplicadas à classificação de sinais biomédicos.

Palavras-chave: Redes Neurais Pulsadas. Eletrocardiograma. Arritmias Cardíacas.

TITLE: Enhancement of Spiking Neural Network Architecturesfor Electrocardiogram (ECG) Signal Processing

Abstract

This study investigates the application of Spiking Neural Networks (SNNs) to the classification of electrocardiogram (ECG) signals, focusing on the detec- tion of cardiac arrhythmias. A SNN architecture is developed to process ECG signals as temporal sequences of spikes, exploring different preprocessing, nor- malization, and signal encoding techniques. A Random Forest model is also employed as a reference method, allowing comparison between a conventional 1 machine learning approach and the

spiking neural network. In addition, dif- ferent strategies for addressing class imbalance are evaluated, including the use of the original dataset, undersampling, and data augmentation applied to the arrhythmia class. The models are evaluated using metrics such as accu- racy, precision, recall, and F1-score, with particular emphasis on the ability to correctly identify arrhythmia cases. The results provide an analysis of the performance of the different approaches and the limitations of class balancing strategies in the classification of biomedical signals.

Keywords: Keywords: Spiking Neural Networks. Electrocardiogram. Cardiac Arrhythm

Introdução

As doenças cardiovasculares estão entre as principais causas de mortalidade no mundo [5], tornando a identificação precoce de alterações na atividade cardíaca uma importante questão de saúde. Entre essas alterações, as arritmias cardíacas correspondem a irregula- ridades no ritmo dos batimentos e podem estar associadas a diferentes condições clínicas. O eletrocardiograma (ECG) é um dos principais exames utilizados para o monitoramento da atividade elétrica do coração, fornecendo informações que podem auxiliar na identifi- cação dessas alterações. Entretanto, a análise de grandes volumes de sinais de ECG pode ser uma tarefa complexa e demandar tempo e conhecimento especializado. Nesse contexto, técnicas de aprendizado de máquina vêm sendo exploradas como fer- ramentas de apoio à análise automatizada de sinais biomédicos. Entre essas técnicas, as Redes Neurais Pulsadas (Spiking Neural Networks — SNNs) apresentam uma abor- dagem inspirada no funcionamento do sistema nervoso biológico, na qual as informações são representadas por eventos discretos denominados disparos (spikes). Diferentemente das redes neurais convencionais, as SNNs incorporam explicitamente a dimensão tempo- ral dos dados, característica que pode ser relevante para o processamento de sinais que apresentam comportamento dinâmico, como os sinais de ECG. Um dos desafios presentes na classificação de arritmias é o desbalanceamento entre as classes. Em conjuntos de dados de ECG, é comum que registros correspondentes a ritmos normais sejam significativamente mais numerosos que registros associados a arritmias. Esse desequilíbrio pode levar os modelos de classificação a favorecerem a classe majoritária, resultando em elevada acurácia geral, mas em baixa capacidade de identificação dos casos de arritmia. Dessa forma, a avaliação de métricas como precisão, recall e F1-score torna-se 2 fundamental para analisar adequadamente o desempenho dos modelos. Diante desse problema, este trabalho tem como objetivo desenvolver e avaliar uma SNN para a classificação de sinais de ECG, investigando sua capacidade de distinguir registros normais de registros associados a arritmias. Como referência, também é empre- gado um modelo Random Forest, permitindo comparar a abordagem baseada em redes neurais pulsadas com um método tradicional de aprendizado de máquina. Além disso, são investigadas diferentes estratégias para lidar com o desbalanceamento das classes, incluindo o uso dos dados originais, undersampling e técnicas de data augmentation. A avaliação das abordagens é realizada por meio de métricas de classificação, com ênfase na identificação da classe de arritmia. Dessa forma, busca-se não apenas analisar a acurácia dos modelos, mas também compreender os impactos das diferentes estratégias de preparação dos dados sobre a capacidade de detecção dos casos minoritários. Espera-se, assim, contribuir para

a investigação do uso de SNNs no processamento de sinais biomédi- cos e para a compreensão de suas possibilidades e limitações na classificação automatizada de arritmias cardíacas.

Metodologia

O desenvolvimento deste trabalho foi realizado utilizando a MIT-BIH Arrhythmia Database [4, 3], uma base de dados amplamente utilizada em estudos de processamento e
classificação de sinais de eletrocardiograma (ECG). A base é composta por 48 registros, acompanhados de sinais de ECG e suas respectivas anotações de batimentos. Inicialmente,
foi realizada uma análise exploratória dos registros, considerando características como frequência de amostragem, número de derivações, duração dos sinais e distribuição dos símbolos presentes nas anotações. Os registros possuem frequência de amostragem de 360 Hz e duas derivações de ECG. Também foi analisada a distribuição das anotações, permitindo identificar o desbalanceamento entre os eventos classificados como normais e
aqueles associados a diferentes tipos de arritmia.
Para a construção do conjunto de dados, cada anotação de batimento foi utilizada

como ponto central para a extração de uma janela temporal de 180 amostras. Foram con-
sideradas 90 amostras anteriores e 90 amostras posteriores ao ponto central da anotação.

Janelas que ultrapassassem os limites do registro foram descartadas. Para os experi- mentos, foi utilizada a primeira derivação do sinal, correspondente à derivação MLII. Os

diferentes símbolos das anotações foram agrupados em duas classes: os símbolos N, R e L foram classificados como Normal (N), enquanto os demais símbolos foram agrupados como Arritmia (A). Dessa forma, o conjunto completo foi composto por 225.158 janelas,
sendo 180.726 da classe Normal e 44.432 da classe Arritmia.

Para evitar que janelas provenientes de um mesmo registro estivessem simultaneamente

nos conjuntos de treinamento e teste, os experimentos finais utilizaram uma divisão ba- seada nos registros. Foram selecionados 22 registros para treinamento e 23 registros para

teste, resultando em 52.505 janelas para treinamento e 51.155 para teste. Essa estratégia foi adotada para obter uma avaliação mais adequada da capacidade de generalização dos modelos para registros não observados durante o treinamento. Como experimento inicial,
também foi realizada uma divisão aleatória das janelas utilizando 20% dos dados para

teste e estratificação das classes. Entretanto, essa abordagem apresentou desempenho praticamente perfeito e não foi utilizada como avaliação final, pois poderia permitir a presença de janelas do mesmo registro nos conjuntos de treinamento e teste.

Para a classificação utilizando aprendizado de máquina tradicional, foi desenvolvido um modelo Random Forest [1]. Cada janela de ECG foi utilizada diretamente como um vetor de 180 características. O modelo foi configurado com 100 árvores de decisão, utilizando random_state=42 e processamento paralelo.

Para o desenvolvimento da Spiking Neural Network (SNN), as janelas de ECG foram submetidas a uma etapa de normalização individual. Em cada janela, os valores foram transformados para o intervalo entre 0 e 1 utilizando os valores mínimo e máximo da própria janela. Em seguida, foi aplicada uma codificação estocástica baseada em Bernoulli
[2], na qual cada amostra normalizada foi utilizada como probabilidade de geração de um
disparo. Assim, cada janela de 180 amostras foi convertida em uma sequência temporal de spikes ao longo de 180 passos de tempo. Na implementação com o mlGeNN, a entrada
foi representada por um único neurônio, sendo os disparos armazenados de acordo com seus tempos e identificadores neuronais.

A SNN foi implementada utilizando a biblioteca mlGeNN [8], baseada no framework GeNN [7]. A arquitetura possui uma camada de entrada, uma camada oculta com 64 neurônios e uma camada de saída com dois neurônios, correspondentes às classes Normal
e Arritmia. A camada oculta utiliza neurônios Leaky Integrate-and-Fire (LIF), enquanto a camada de saída utiliza neurônios Leaky Integrate. As conexões são densas, permitindo o processamento temporal das sequências de spikes.

O treinamento da SNN foi realizado utilizando o algoritmo EventProp [6], com a função de perda sparse_categorical_crossentropy e o otimizador Adam. Foram testadas

diferentes configurações de taxa de aprendizado durante a experimentação, sendo selecionada a configuração utilizada nos experimentos finais. O treinamento foi realizado

durante 15 épocas, com embaralhamento dos dados e utilização de 10% dos dados de treinamento para validação.

Devido ao desbalanceamento entre as classes, também foram investigadas estratégias de preparação dos dados, incluindo a utilização da distribuição original, o undersampling

da classe Normal e técnicas de data augmentation aplicadas à classe Arritmia. O desempenho dos modelos foi avaliado utilizando acurácia, precisão, recall e F1-score, além da

matriz de confusão. Foi dada atenção especial ao recall e ao F1-score da classe Arritmia, uma vez que a predominância da classe Normal pode resultar em elevada acurácia mesmo
quando uma parcela relevante dos casos de arritmia não é identificada.

Resultados e Discussões

A partir da etapa de preparação dos dados, foram realizados experimentos para avaliar o desempenho da Spiking Neural Network (SNN) na classificação das janelas de ECG entre as classes Normal (N) e Arritmia (A). A avaliação foi realizada por meio de acurácia, precisão, recall e F1-score, com atenção especial aos resultados da classe Arritmia, devido ao desbalanceamento observado na base de dados.
Como primeira avaliação, foi utilizado um conjunto de dados sem aplicação de técnicas de balanceamento ou aumento artificial das amostras. Nesse experimento, a SNN foi treinada utilizando a distribuição original dos dados. O conjunto de teste foi composto por 51.155 janelas, sendo 44.033 pertencentes à classe Normal e 7.122 à classe Arritmia.

O modelo apresentou acurácia de 87,74%. Para a classe Normal, foram obtidos precisão de 0,89, recall de 0,98 e F1-score de 0,93. Para a classe Arritmia, entretanto, os resultados foram inferiores, com precisão de 0,67, recall de 0,24 e F1-score de 0,35. Esses resultados indicam que, embora a acurácia geral seja relativamente elevada, a capacidade de identificação dos casos de arritmia ainda é limitada. O recall de 0,24 indica que uma parcela significativa dos exemplos pertencentes à classe Arritmia não foi corretamente identificada pelo modelo.

Esse comportamento está relacionado à distribuição desigual das classes. No conjunto de teste, aproximadamente 86% das amostras pertencem à classe Normal. Dessa forma, uma acurácia elevada pode ser obtida mesmo quando o modelo apresenta baixa capacidade de reconhecimento da classe minoritária. Portanto, para este problema, métricas como recall e F1-score da classe Arritmia são mais informativas do que a acurácia isoladamente.

Com o objetivo de verificar o efeito do balanceamento, foi realizado um segundo experimento utilizando undersampling da classe Normal. Nessa abordagem, a quantidade de exemplos da classe majoritária foi reduzida para igualar aproximadamente o número de amostras das duas classes no conjunto de treinamento. O objetivo foi diminuir a influência da classe Normal durante o aprendizado e verificar se a rede passaria a apresentar maior sensibilidade aos casos de arritmia.

Nesse experimento, observou-se uma redução significativa da acurácia, que passou para aproximadamente 65,59%. Em contrapartida, o recall da classe Arritmia aumentou para aproximadamente 53%. Esse resultado evidencia um compromisso entre o desempenho global e a capacidade de detecção da classe minoritária. A redução da quantidade de exemplos normais fez com que a rede deixasse de favorecer tão fortemente essa classe, aumentando sua capacidade de reconhecer arritmias.
Embora a redução da acurácia possa inicialmente indicar uma piora do modelo, a

análise das métricas por classe mostra que o undersampling produziu uma mudança rele-
vante no comportamento da SNN. Para um sistema voltado à identificação de alterações

em sinais de ECG, a capacidade de detectar corretamente os casos de arritmia é particu-larmente importante. Assim, o aumento do recall da classe Arritmia deve ser considerado

juntamente com a perda de desempenho observada na classe Normal.

Em seguida, foi investigada uma estratégia de data augmentation aplicada à classe Ar-ritmia. Foram geradas novas amostras a partir dos sinais existentes por meio de pequenas

perturbações nos dados, buscando aumentar a quantidade de exemplos disponíveis para a classe minoritária sem reduzir diretamente o conjunto de exemplos da classe Normal. Nesse experimento, a quantidade de exemplos de Arritmia no treinamento foi aumentada
de 6.681 para 13.362, enquanto a classe Normal permaneceu com 45.824 exemplos. Apesar do aumento da quantidade de exemplos da classe minoritária, os resultados

não apresentaram melhora na detecção de arritmias. A SNN obteve acurácia de aproxi-madamente 86,52%, porém o recall da classe Arritmia foi de apenas 0,04 e o F1-score foi

de aproximadamente 0,07. Ao mesmo tempo, o recall da classe Normal alcançou 1,00.

Esse comportamento indica que o modelo continuou fortemente inclinado à classificação
das amostras como Normal.

A comparação entre os experimentos demonstra que o aumento da quantidade absoluta

de exemplos de Arritmia não foi suficiente para solucionar o problema de desbalancea-mento. Embora a proporção da classe minoritária no treinamento tenha aumentado, a

classe Normal ainda representava a maioria das amostras. Dessa forma, o data augmen-tation utilizado não produziu o mesmo efeito observado com o undersampling sobre a

capacidade de identificação da classe Arritmia.
A Tabela 1 apresenta uma comparação entre as estratégias avaliadas para a SNN.

Tabela 1: Desempenho da SNN sob diferentes estratégias de preparação dos dados.
Estratégia Acurácia Recall (A) F1-score (A)
Sem balanceamento 87,74% 24% 0,35
Undersampling 65,59% 53% 0,30
Data augmentation 86,52% 4% 0,07

Os resultados mostram que não existe uma única métrica capaz de representar adequadamente o desempenho dos modelos neste problema. O experimento sem balanceamento

apresentou a maior acurácia entre as estratégias avaliadas, porém identificou apenas uma
parcela limitada dos casos de arritmia. O undersampling apresentou a maior capacidade

de identificação da classe Arritmia, medida pelo recall, mas provocou uma redução expressiva na acurácia geral. Já o data augmentation manteve uma acurácia próxima à

obtida no baseline, mas apresentou o pior resultado para a classe Arritmia.

Como método de comparação, também foi desenvolvido um classificador Random Forest.

Inicialmente, foi realizada uma divisão aleatória das janelas em conjuntos de trei-

namento e teste, mantendo a proporção das classes por meio de estratificação. Nesse primeiro experimento, o modelo apresentou desempenho aparentemente próximo de 100%.

Entretanto, essa avaliação não foi utilizada como comparação principal, pois as janelas

provenientes de um mesmo registro de ECG poderiam estar simultaneamente nos conjuntos de treinamento e teste, permitindo que características específicas dos registros fossem

compartilhadas entre os dois conjuntos.

Para uma comparação mais adequada, foi utilizada uma divisão baseada nos registros da MIT-BIH Arrhythmia Database. Os registros foram separados em conjuntos de

treinamento e teste, de modo que as janelas de um determinado registro permanecessem exclusivamente em um dos conjuntos. O conjunto de treinamento resultou em 52.505 janelas, enquanto o conjunto de teste apresentou 51.155 janelas.
Com essa divisão por registros, o Random Forest apresentou acurácia de 87,53%. Para a classe Arritmia, foram obtidos precisão de 0,55, recall de 0,63 e F1-score de 0,58. Para
a classe Normal, os valores foram precisão de 0,94, recall de 0,92 e F1-score de 0,93. A matriz de confusão apresentou 4.481 arritmias corretamente classificadas, 2.641 arritmias

classificadas como Normal, 3.740 amostras Normais classificadas como Arritmia e 40.293
classificações corretas da classe Normal.

Esses resultados indicam que o Random Forest apresentou maior capacidade de identificação da classe Arritmia do que a SNN no experimento sem técnicas de balanceamento,

principalmente quando considerado o recall da classe minoritária. Entretanto, a comparação deve ser interpretada considerando as diferenças entre as abordagens e, principalmente, a forma de divisão dos dados. O resultado do Random Forest também evidencia

a importância de evitar a mistura de janelas provenientes do mesmo registro entre treinamento e teste, uma vez que essa situação pode produzir estimativas excessivamente

otimistas de desempenho.

De forma geral, os experimentos demonstram que o principal desafio encontrado não está apenas na classificação dos sinais, mas também no tratamento do desbalanceamento

entre as classes. A elevada quantidade de exemplos Normais favoreceu modelos que priori-
zam essa classe, resultando em acurácias elevadas mesmo quando uma parcela considerável

das arritmias não é detectada. Por esse motivo, o recall e o F1-score da classe Arritmia foram considerados indicadores fundamentais na interpretação dos resultados.
Os experimentos também mostram que diferentes estratégias de preparação dos dados

produzem efeitos distintos. O undersampling foi a estratégia que proporcionou o maior au-
mento no recall da classe Arritmia entre os experimentos realizados com a SNN, enquanto

o data augmentation avaliado não apresentou benefício para essa métrica. Esses resultados sugerem que o simples aumento da quantidade de exemplos da classe minoritária não

garante uma melhora na capacidade de classificação, sendo necessário considerar também
a distribuição relativa das classes e as características das transformações aplicadas aos sinais.

Assim, os resultados obtidos estabelecem uma base para a avaliação das estratégias de balanceamento e para a comparação entre a abordagem baseada em redes neurais pulsadas e métodos tradicionais de aprendizado de máquina. A análise evidencia que, para aplicações de classificação de arritmias, a avaliação deve considerar conjuntamente o desempenho global e a capacidade de identificar corretamente a classe minoritária, evitando conclusões baseadas exclusivamente na acurácia.

Conclusão

Os resultados obtidos neste trabalho demonstraram a viabilidade da utilização de uma Spiking Neural Network (SNN) para a classificação de sinais de eletrocardiograma da MIT-BIH Arrhythmia Database em duas classes: Normal e Arritmia. A utilização de codificação estocástica permitiu representar as janelas de ECG como sequências temporais de disparos, possibilitando o processamento dos sinais por uma rede neural pulsada.

Os experimentos evidenciaram, entretanto, que o desbalanceamento entre as classes constitui um dos principais desafios para o problema estudado. No experimento sem aplicação de técnicas de balanceamento, a SNN apresentou acurácia de 87,74%, mas obteve recall de apenas 24% para a classe Arritmia. Esse resultado demonstra que uma acurácia elevada não necessariamente representa uma boa capacidade de detecção da classe minoritária, uma vez que o modelo apresentou forte tendência à classificação das amostras como Normal.

A aplicação de undersampling modificou esse comportamento, elevando o recall da classe Arritmia para 53%. Por outro lado, houve uma redução da acurácia para aproximadamente 65,59%, evidenciando um compromisso entre a identificação da classe minoritária e o desempenho global do modelo. Já a estratégia de data augmentation, apesar de aumentar a quantidade de exemplos de Arritmia no conjunto de treinamento, não produziu melhoria na detecção dessa classe. Nesse experimento, a acurácia foi de aproximadamente 86,52%, enquanto o recall de Arritmia foi de apenas 4%.

A comparação com o Random Forest também mostrou resultados relevantes. Utilizando uma divisão dos dados baseada nos registros da base, evitando a distribuição de janelas do mesmo registro entre treinamento e teste, o modelo apresentou acurácia de 87,53%, recall de 63% e F1-score de 0,58 para a classe Arritmia. Dessa forma, o Random Forest apresentou, nos experimentos realizados, maior capacidade de identificação da classe Arritmia do que a SNN sem balanceamento.

De maneira geral, os resultados indicam que a avaliação de modelos para classificação de arritmias não deve ser baseada exclusivamente na acurácia. O recall e o F1-score da

classe Arritmia mostraram-se métricas fundamentais para avaliar a capacidade dos mode-
los de identificar corretamente os casos de interesse. Os experimentos também demonstra-
ram que diferentes estratégias de preparação dos dados podem produzir comportamentos significativamente distintos na classificação.

Referências

[1] BREIMAN, Leo. Random forests. Machine Learning, v. 45, p. 5–32, 2001. DOI: 10.1023/A:1010933404324. Disponível em: https://doi.org/10.1023/A:1010933404324. Acesso em: 27 ago. 2026.
[2] ESHRAGHIAN, Jason K. et al. Training spiking neural networks using lessons from deep learning. Proceedings of the IEEE, v. 111, n. 9, p. 1016–1054, 2023. DOI: 10.1109/JPROC.2023.3308088. Disponível em: https://arxiv.org/abs/2109.12894. Acesso em: 29 ago. 2026.

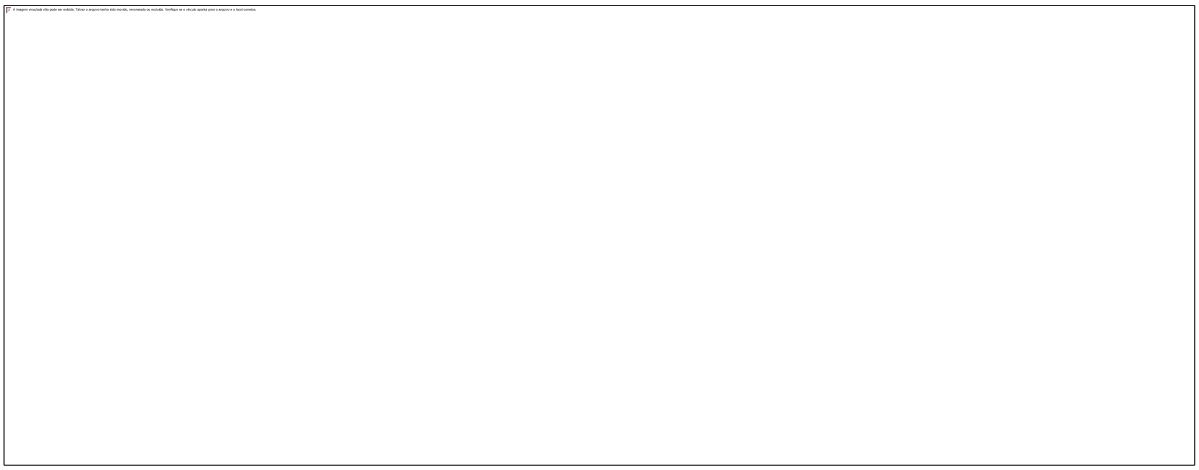
[3] MOODY, George B.; MARK, Roger G. The impact of the MIT-BIH arrhythmia data-
base. IEEE Engineering in Medicine and Biology Magazine, v. 20, n. 3, p. 45–50, 2001.

DOI: 10.1109/51.932724. Disponível em: https://doi.org/10.1109/51.932724. Acesso em: 27 ago. 2026.
[4] PHYSIOBANK. MIT-BIH Arrhythmia Database. PhysioNet, 2005. Disponível em: https://physionet.org/content/mitdb/1.0.0/. Acesso em: 27 ago. 2026.
[5] WORLD HEALTH ORGANIZATION. Cardiovascular diseases (CVDs). Geneva: WHO, 2025. Disponível em: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds). Acesso em: 29 ago. 2026.

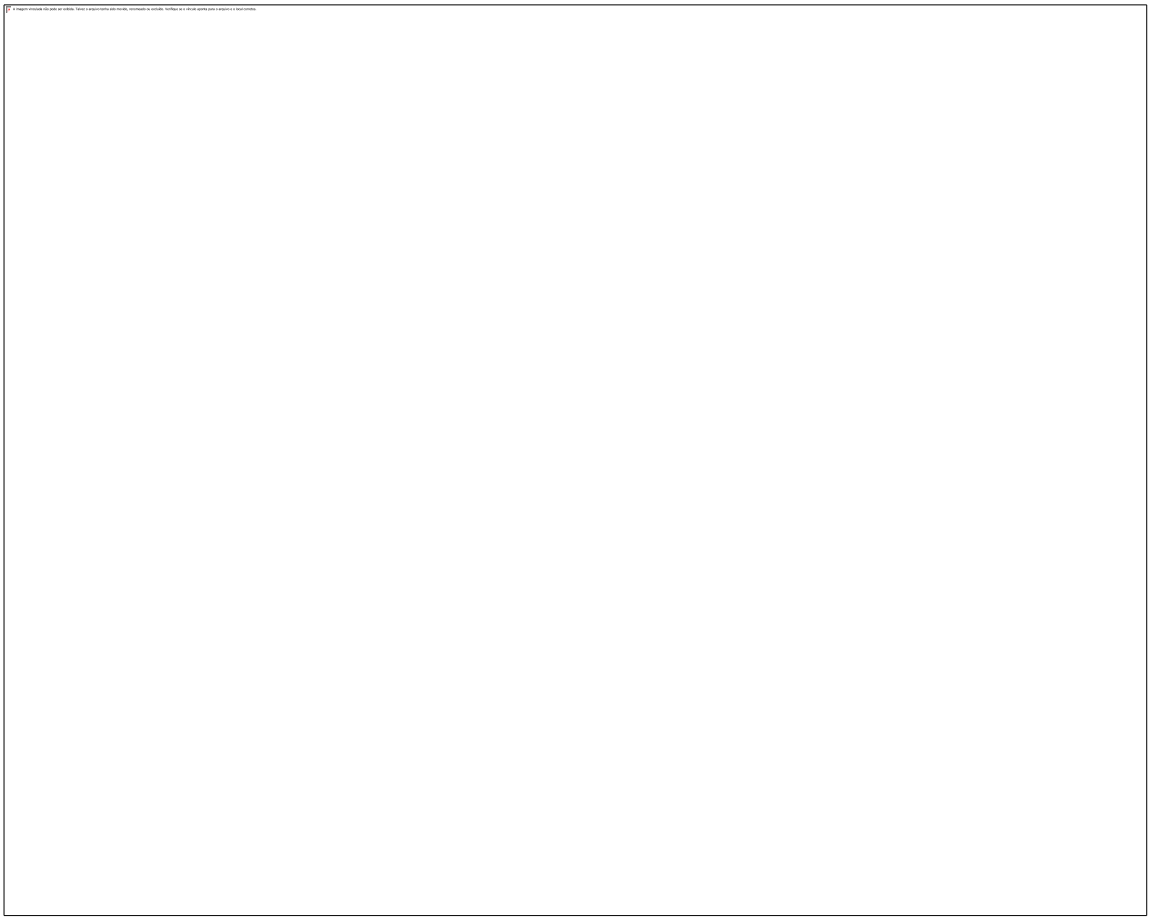
[6] WUNDERLICH, Timo C.; PEHLE, Christian. Event-based backpropagation can com-
pute exact gradients for spiking neural networks. Scientific Reports, v. 11, art. 12829,

2021. DOI: 10.1038/s41598-021-91786-z. Disponível em: https://doi.org/10.1038/s41598-021-91786-z. Acesso em: 27 ago. 2026.
[7] YAVUZ, Esin; TURNER, James; NOWOTNY, Thomas. GeNN: a code generation framework for accelerated brain simulations. Scientific Reports, v. 6, art. 18854, 2016. DOI: 10.1038/srep18854. Disponível em: https://doi.org/10.1038/srep18854. Acesso em: 27 ago. 2026.
[8] ML-GENN. Training networks: EventProp. mlGeNN Documentation. Disponível em: https://ml-genn.readthedocs.io/en/latest/usage/training_networks.html. Acesso em: 27 ago. 2026.
[9] GENN TEAM. GeNN: GPU-enhanced Neuronal Networks. GitHub. Disponível em: https://github.com/genn-team/genn. Acesso em: 27 ago. 2026.

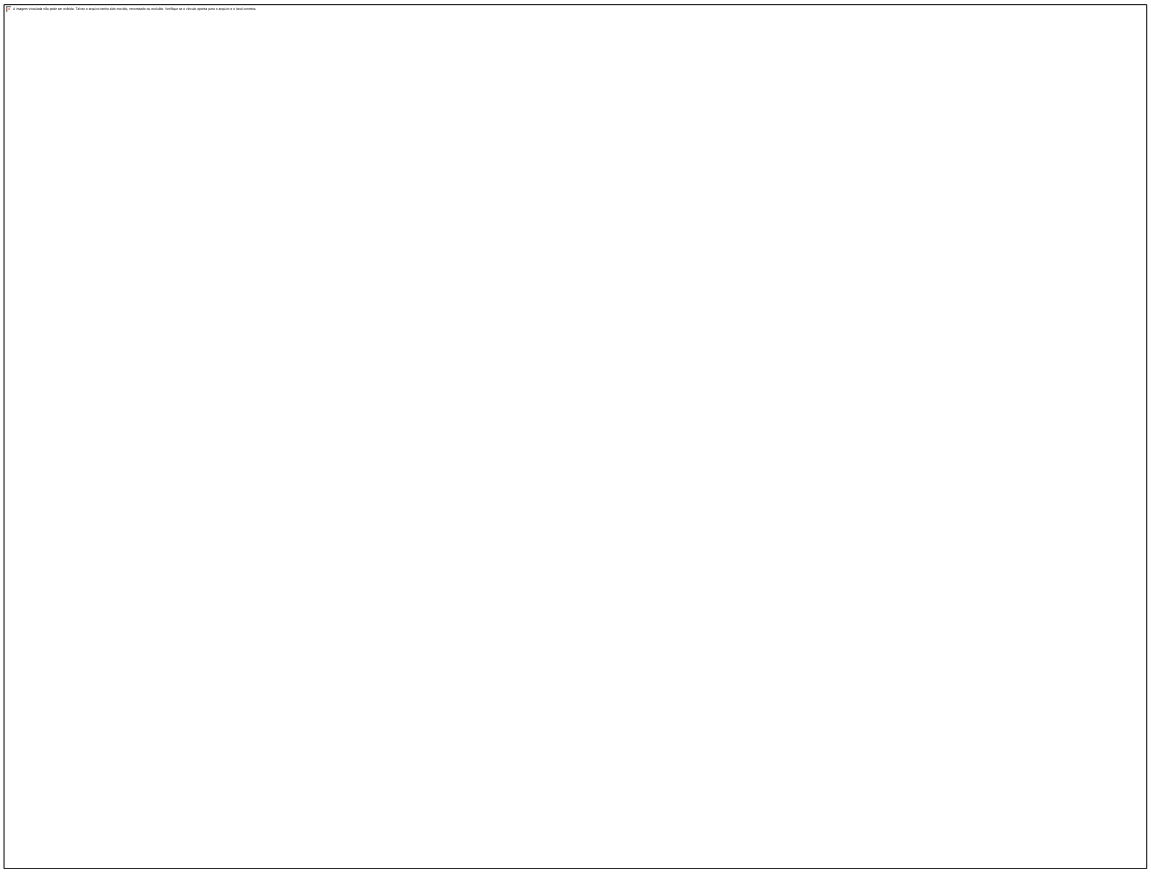
Anexos



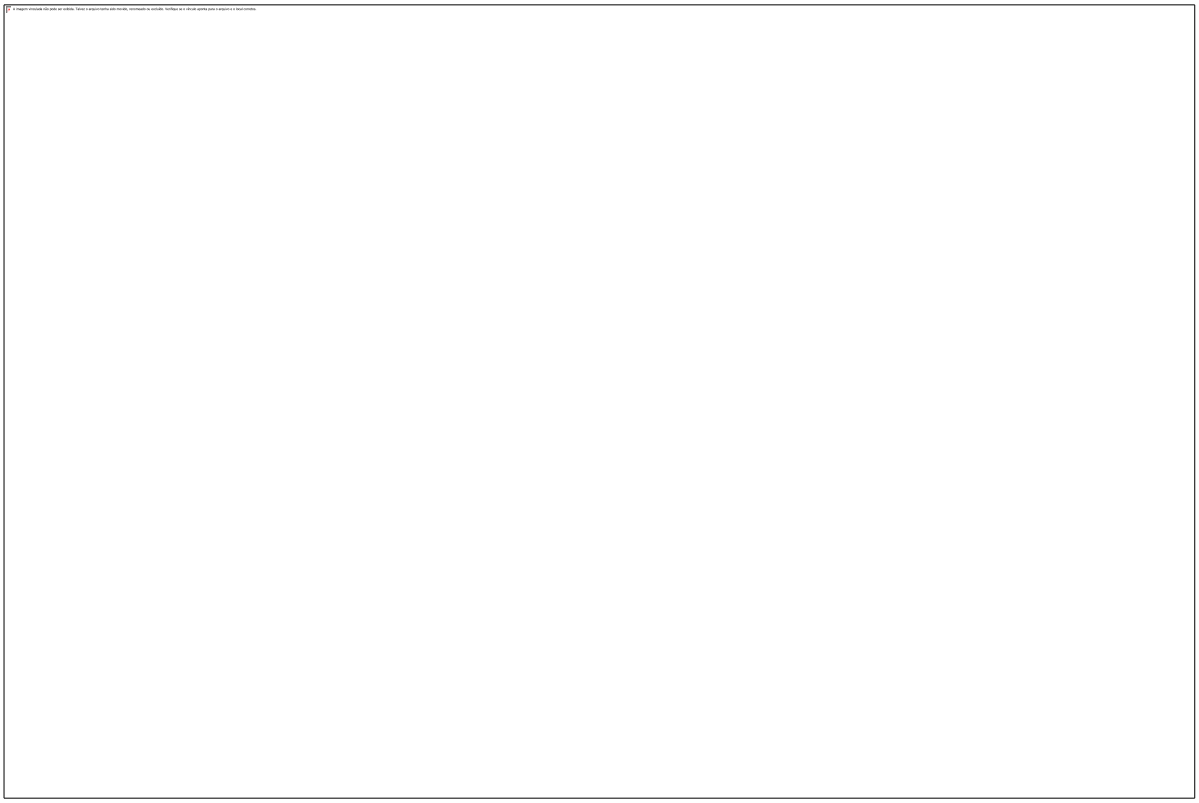
Exemplo de sinal de ECG



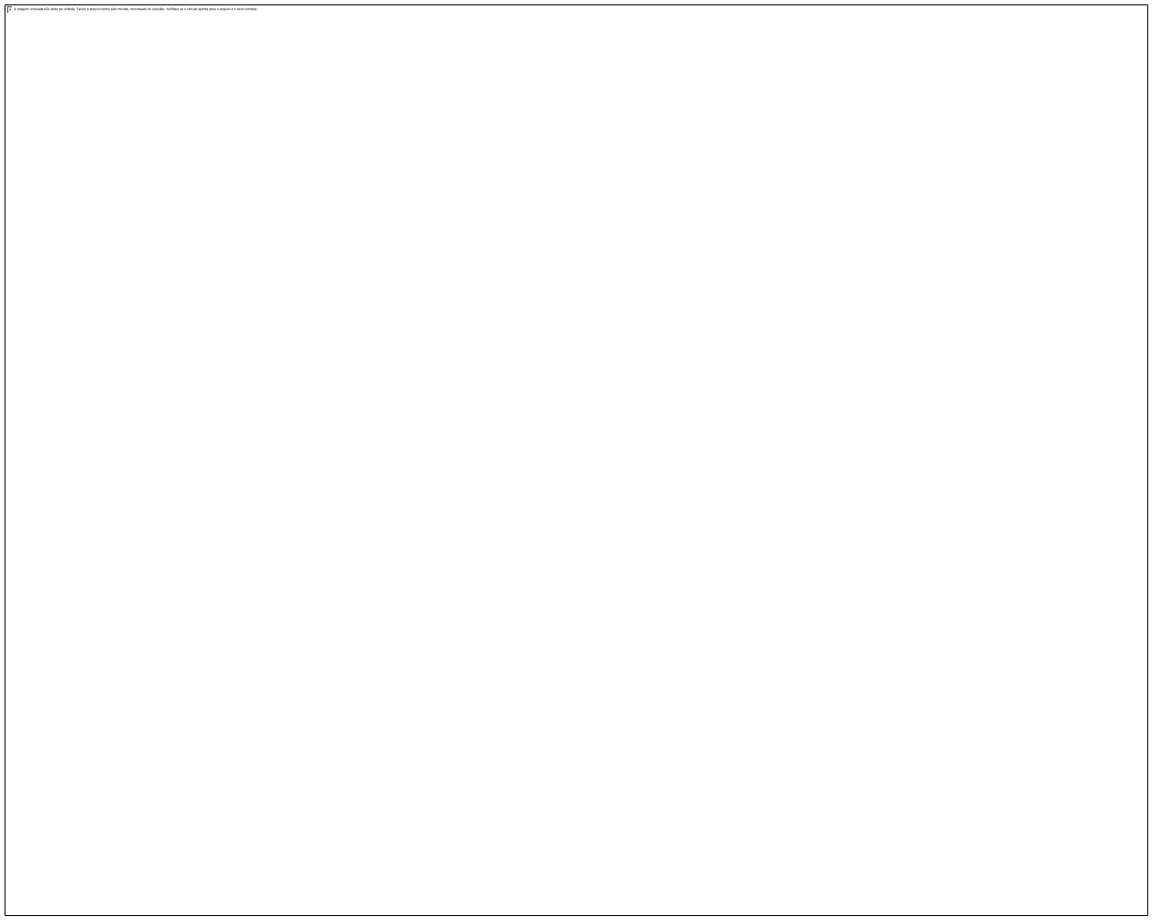
Matriz de confusão - Baseline



Matriz de confusão - undersampling



Evoluçao da acurácia no treinamento utilizando Undersampling



Matriz de confusão - Augmentation

CÓDIGO: ET1052

AUTOR: LORENZO SANTOS FURTADO

COAUTOR: ISRAEL MARTINS VALCAREGGI

ORIENTADOR: RENAN CIPRIANO MOIOLI

TÍTULO: Implantação de Redes Neurais Pulsadas no Processamento de Imagens Biomédicas

Resumo

Este trabalho investiga a aplicação de Redes Neurais Pulsadas (SNNs, Spiking Neural Networks) à classificação binária de imagens dermatológicas do benchmark DermaMNIST, utilizando o framework mlGeNN sobre hardware GPU NVIDIA. O problema consiste em distinguir lesões malignas de benignas a partir de imagens de 64×64 pixels, abordado sob duas estratégias: (i) treinamento direto com o algoritmo e-prop e sua variante EventProp, com otimização de hiperparâmetros via Optuna (TPE e CMA-ES); e (ii) conversão ANN/SNN a partir de uma CNN treinada com TensorFlow (abordagem top-down). A melhor acurácia obtida com a abordagem direta foi de 71,95% com EProp e 71,39% com EventProp em treino, enquanto a abordagem top-down com CNN atingiu 79,00% no conjunto de teste. O trabalho documenta os principais desafios encontrados: instabilidade de treinamento, restrições de memória de vídeo, limitações do ring buffer do EventProp com imagens RGB, e a curva de aprendizado associada ao ecossistema mlGeNN.

Palavras-chave: Redes Neurais Pulsadas. IA. Melanoma. DermaMNIST. Optuna

TITLE: Implementation of Spiking Neural Networks in Biomedical Image Processing

Abstract

This study investigates the application of Spiking Neural Networks (SNNs) to the binary classification of dermatological images from the DermaMNIST benchmark, using the mlGeNN framework on NVIDIA GPU hardware. The task involves

distinguishing between malignant and benign lesions using 64×64 pixel images, approached via two strategies: (i) direct training using the e-prop algorithm and its variant, EventProp, with hyperparameter optimization via Optuna (TPE and CMA-ES); and (ii) ANN-to-SNN conversion from a CNN trained with TensorFlow (a top-down approach). The best accuracy achieved with the direct approach was 71.95% with e-prop and 71.39% with EventProp during training, whereas the top-down CNN approach reached 79.00% on the test set. The study documents the key challenges encountered: training instability, video memory constraints, limitations of the EventProp ring buffer with RGB images, and the learning curve associated with the mlGeNN ecosystem.

Keywords: Spiking Neural Networks. AI. Melanoma. DermaMNIST. Optuna

Introdução

As Redes Neurais Pulsadas representam a chamada terceira geração de modelos de redes neurais artificiais, proposta originalmente por Maass (1997). Diferentemente das redes convencionais que processam valores reais contínuos em cada camada, as SNNs comunicam informação exclusivamente por meio de eventos discretos no tempo, os spikes, o que as aproxima do funcionamento observado em neurônios biológicos. Essa característica confere às SNNs duas vantagens teóricas fundamentais para aplicações de tempo real e sistemas embarcados: eficiência energética, pois os cálculos só ocorrem quando há spikes, e capacidade de processar padrões temporais de forma natural, aproveitando a dimensão do tempo como recurso computacional.

O framework mlGeNN (Machine Learning com GeNN) traduz esses conceitos para hardware GPU de forma acessível, compilando modelos de SNNs para código CUDA via GeNN (GPU-enhanced Neuronal Networks). Na versão 2.3.1 utilizada neste trabalho, o framework oferece dois algoritmos de aprendizado supervisionado: o e-prop, baseado em traços de elegibilidade, e o EventProp, que calcula gradientes exatos orientados a eventos. Ambos evitam o dispendioso backpropagation through time (BPTT), tornando o treinamento mais eficiente em termos de memória para sequências temporais longas.

O dataset escolhido, DermaMNIST, é um benchmark de classificação multiclasse de lesões dermatológicas com 7 classes, derivado do ISIC Archive. Neste trabalho adotou-se uma binarização do problema, separando lesões malignas (Melanoma — classe 4) das benignas (Benign keratosis-like lesions — classe 2), gerando um problema binário com aproximadamente 1.548 amostras de treino e 460 de teste após o processo de filtragem.

A motivação para iniciar pela configuração mais simples possível, replicar a rede MLP que classifica bem o MNIST e reajustar os hiperparâmetros para o DermaMNIST, seguiu a lógica de um benchmark bottom-up: compreender as limitações do modelo mais básico antes de aumentar a complexidade. Ao longo do trabalho essa estratégia revelou desafios específicos do problema (desbalanceamento severo, similaridade visual

entre classes, restrições de memória GPU) que fundamentaram as decisões de arquitetura e algoritmo descritas nas seções seguintes.

Metodologia

O DermaMNIST foi carregado em resolução 64×64 pixels via biblioteca MedMNIST, em escala de cinza. A binarização das 7 classes originais em maligno/benigno. Como as classes binarizadas apresentavam desbalanceamento moderado (769 vs. 779 amostras após filtragem), foi aplicado data augmentation via transformações geométricas (rotação aleatória entre −60° e 60°) e adição de ruído gaussiano (μ=0, σ=1), ambas aplicadas com probabilidade igual sobre amostras da classe minoritária. O conjunto de teste foi mantido com a distribuição original sem augmentation, separado previamente para evitar data leakage. A inversão das imagens (255 − x) foi aplicada antes da codificação, de forma que as lesões, originalmente escuras, gerassem spikes mais cedo no encoding por latência, priorizando a informação clinicamente relevante.

Para converter as imagens em sequências temporais de spikes, foi utilizada a função log_latency_encode_data do mlGeNN, que implementa a codificação por latência logarítmica conforme a equação:

$T(x) = \tau_{eff} \cdot \log(x / (x - \theta))$ se $x > \theta$, ∞ caso contrário

onde τ_{eff} é a constante de tempo e θ é o limiar de intensidade de pixel. Pixels mais brilhantes (após inversão, correspondentes às lesões) geram spikes mais cedo, sendo priorizados temporalmente pela rede. Os valores τ_{eff} = 19,45 ms e θ = 8 foram otimizados para EProp via Optuna; para EventProp os melhores valores encontrados foram τ_{eff} = 28,86 ms e θ = 168 (escala 0–255 após inversão).

Usamos duas arquiteturas principais, ambas totalmente feedfoward, onde todas as camadas são conectadas diretamente entre si. Na primeira camada da rede cada neurônio corresponde a um pixel da imagem e irá receber o spike no tempo de acordo com o resultado da equação de codificação de cada imagem. Para o algoritimo de treinamento e-prop (Bellec et al., 2020) usamos uma camada escondida e para EventProp (Wunderlich & Pehle, 2021) duas camadas intermediárias. A camada escondida é conectada à saída de dois neurônios, um para cada rótulo, a predição é com base no neurônio que acumulou maior potencial de membrana durante o experimento.

O e-prop é um algoritmo biologicamente plausível que calcula gradientes de forma online durante o forward pass, sem necessidade de backward pass. O princípio central são os traços de elegibilidade: cada sinapse mantém uma variável eFiltered que decai exponencialmente com a mesma constante de tempo τ_{mem} do neurônio LIF:

$eFiltered = (eFiltered \times \alpha) + e$, $\alpha = \exp(-dt / \tau_{mem})$

A atualização do peso acumula o produto do traço pelo sinal de erro da camada de saída (ΔG += eF × E_post), mais um termo de regularização que penaliza desvios da taxa alvo de disparo. Por não armazenar histórico de estados, a memória do e-prop é constante independentemente do número de timesteps. Sua principal limitação é que os gradientes são aproximados, e cada conexão precisa de seu próprio traço, tornando-o incompatível com pesos compartilhados (como convoluções).

O EventProp calcula gradientes exatos através de um backward pass orientado a eventos. Durante o forward pass, os tempos de spike de cada neurônio são armazenados em um ring buffer de tamanho fixo max_spikes por neurônio por amostra. No backward pass, esses tempos são percorridos em ordem reversa para calcular os gradientes. Isso elimina a necessidade de armazenar todos os estados intermediários como no BPTT, mas impõe uma restrição crítica: se um neurônio disparar mais vezes que max_spikes, os spikes excedentes são silenciosamente descartados, corrompendo o gradiente.

A principal vantagem do EventProp sobre o e-prop é suportar múltiplas camadas escondidas sem a restrição de conexão direta à saída, e produzir gradientes exatos que tendem a convergir de forma mais estável. A principal desvantagem é o consumo de memória proporcional a batch_size × n_neurônios × max_spikes × 4 bytes, o que torna o uso com imagens RGB de alta resolução proibitivo para GPUs com pouca VRAM.

A busca de hiperparâmetros foi conduzida com a biblioteca Optuna (ref), testando dois algoritmos de amostragem: o TPE (Tree-structured Parzen Estimator), que modela cada hiperparâmetro independentemente e é eficiente para buscas exploratórias amplas; e o CMA-ES (Covariance Matrix Adaptation Evolution Strategy), que aprende correlações entre parâmetros e é mais eficiente para refinar regiões já identificadas como promissoras. Em todos os experimentos foi utilizado o MedianPruner com número de warmup steps=5, que interrompe trials cujo desempenho esteja abaixo da mediana nas épocas iniciais, economizando tempo de cluster. Os resultados de cada trial foram salvos incrementalmente em arquivo CSV via callback.

Em paralelo à estratégia de treinamento direto (bottom-up), foi investigada a conversão de uma CNN treinada com TensorFlow/Keras para uma SNN equivalente no mlGeNN (abordagem top-down). A CNN foi treinada com imagens RGB 64×64, arquitetura Conv(32) > Pool > Conv(64) > Pool Conv(128) > Pool Dense (256) > Dense(2), usando exclusivamente ReLU como ativação — pré-requisito para conversão via neurônios integra-dispara, que aproximam o comportamento da ReLU no domínio de spikes.

Os pesos treinados foram salvos camada a camada e carregados no mlGeNN, substituindo as camadas pelas equivalentes disponíveis no pacote. A inferência SNN utiliza o InferenceCompiler com evaluate_timesteps=500, e o readout spike_count na camada de saída, contando spikes ao longo da simulação em vez de acumular potencial de membrana.

Resultados e Discussões

A Figura 1 evidencia uma trajetória de melhora consistente ao longo dos experimentos. O ponto de partida, uma rede MLP simples com os hiperparâmetros padrão usados para MNIST, resultou em acurácia próxima ao acaso (50%), revelando que os parâmetros otimizados para MNIST não são diretamente transferíveis para DermaMNIST. A maior contribuição isolada para a melhoria de acurácia foi a inclusão de τ_eff e θ como hiperparâmetros no Optuna (Exp. 2 - 68,37%), revelando que a qualidade da codificação de entrada tem impacto tão significativo quanto a arquitetura da rede. O valor $\theta \approx 8$ encontrado pelo TPE, correspondendo a quase todos os pixels gerando spikes, pode parecer contra-intuitivo, mas após a inversão das imagens esse valor faz com que apenas pixels muito escuros (pele intacta, menos informativa) sejam silenciados, enquanto toda a variação da lesão é preservada.

A Figura 2 ilustra como o parâmetro max_example_timesteps interage diretamente com a janela temporal de aprendizado do e-prop. O traço de elegibilidade decai com a mesma constante de tempo dos neurônios LIF: para $\tau_mem = 37{,}41$ ms, um spike ocorrido no timestep 0 contribui com apenas ≈7% para a atualização de pesos no timestep 80, e praticamente zero no timestep 200. Isso explica a melhora de 64% para 76% de acurácia ao aumentar de 80 para 200 timesteps: a janela maior permite que mais pixels, especialmente os de baixa intensidade que disparam tarde, contribuam para o aprendizado.

Este resultado é uma manifestação direta de um dos desafios fundamentais do e-prop: o problema de temporal credit assignment. Spikes que ocorrem cedo na apresentação de um exemplo têm seus traços praticamente extintos quando o erro é calculado ao final da simulação. O EventProp mitiga parcialmente esse problema com gradientes exatos, mas a mesma limitação física existe: informação temporal muito distante do momento de decisão é difícil de atribuir corretamente.

A transição do EProp para o EventProp foi motivada principalmente pela necessidade de suportar duas camadas escondidas. O EPropCompiler exige que toda camada escondida esteja diretamente conectada à saída, uma restrição que tornaria impraticável uma arquitetura profunda sem skip connections. O EventProp não tem essa restrição, permitindo arquiteturas sequenciais convencionais.

Em termos de acurácia final, ambas as abordagens convergem para resultados similares (≈71%), o que é consistente com a literatura que indica que gradientes exatos do EventProp oferecem vantagem principalmente em tarefas com estrutura temporal complexa. Para classificação de imagens estáticas convertidas em spikes por latência, a vantagem de gradientes exatos é menos pronunciada.

A principal limitação prática do EventProp revelada neste trabalho é o ring buffer e o consumo de memória ao tentar usar imagens RGB de 64 pixels (12.288 neurônios de entrada), o max_spikes necessário para não corromper os gradientes, proporcional ao número máximo de spikes por amostra, resultou em consumo estimado de ≈40 GB de VRAM apenas para o ring buffer da camada.

Um padrão recorrente em quase todos os experimentos foi a instabilidade de treinamento: a rede frequentemente alternava entre épocas de alta performance e colapso para ≈50%. Esse comportamento, visível na Figura 4 e 5, é característico de

datasets pequenos (≈1.548 amostras de treino) combinados com learning rates relativamente altos. Com apenas 24 batches por época (batch_size=64), cada atualização de gradiente é muito ruidosa, e a rede pode perder em poucas épocas o que levou dezenas para aprender.

O mecanismo de Checkpoint do mlGeNN foi fundamental para contornar esse problema: ao salvar os pesos da melhor época, a inferência final usa a configuração de maior acurácia de validação independentemente do comportamento posterior. A adição do data augmentation contribuiu para reduzir a instabilidade nas primeiras épocas, mas o overfitting rápido (treino com 7000 amostras augmentadas vs. teste com 460 amostras reais) continuou sendo um desafio central.

A abordagem top-down demonstrou ser a mais eficiente em termos de acurácia final: a CNN treinada com TensorFlow atingiu 79% no conjunto de teste (precision=0,81/0,78, recall=0,76/0,82 para benigno/maligno respectivamente), superando em ≈8 pontos percentuais a melhor abordagem de treinamento direto. Esse resultado é esperado dado que: (i) o backpropagation convencional calcula gradientes exatos com informação global da rede; (ii) imagens RGB preservam informação cromática clinicamente relevante, especialmente na distinção de melanomas; e (iii) a CNN com pool reduz progressivamente a dimensionalidade espacial, mantendo apenas as características mais discriminativas.

A tentativa de usar imagens RGB com EventProp revelou que o consumo de VRAM do ring buffer escala cubicamente com a resolução e o número de canais. A GPU não suporta imagens RGB 64×64 com max_spikes adequado. Imagens RGB 28×28 couberam em memória mas não aprenderam, sugerindo que a resolução mínima viável com RGB para este problema é superior ao que a GPU disponível permite.

Incompatibilidade de versão de biblioteca: O mlGeNN 2.3.1 não possui suporte nativo a MaxPool2D, bias em camadas convolucionais, ou BatchNormalization, o que limita a fidelidade da conversão ANN/SNN. Versões mais recentes do mlGeNN (2.4+) incluem suporte expandido a camadas convolucionais, mas não estavam disponíveis no ambiente do cluster no período da pesquisa.

Instabilidade de treinamento: O dataset de apenas 1.548 amostras de treino é insuficiente para convergência estável com SNNs, que tipicamente exigem mais dados que ANNs equivalentes devido à aproximação dos gradientes. O augmentation amenizou mas não eliminou o problema.

Ecossistema ainda em maturação: O mlGeNN é um framework acadêmico em desenvolvimento ativo. A documentação é escassa, alguns parâmetros têm comportamento diferente do esperado (como o pool_size no AvgPoolConv2D usando padding valid em vez de same), e erros de segmentação sem mensagem de diagnóstico foram frequentes durante a fase de conversão CNN/SNN.

Conclusão

Este trabalho demonstrou que SNNs treinadas com e-prop e EventProp via mlGeNN são capazes de classificar imagens dermatológicas com acurácia superior ao acaso em um problema clinicamente relevante, atingindo até 71,95% com treinamento direto. A otimização sistemática de hiperparâmetros via Optuna foi essencial para alcançar esse resultado — a acurácia sem otimização ficou abaixo de 65%.

A comparação entre EProp e EventProp revelou que, para este problema específico (imagens estáticas convertidas em spikes por latência), a qualidade do gradiente (aproximado vs. exato) tem impacto menor do que esperado, e os resultados foram similares. A principal vantagem do EventProp neste contexto foi suportar arquiteturas mais profundas.

A abordagem top-down (CNN/SNN) demonstrou superioridade em acurácia (79%), mas requer validação completa da SNN resultante. A diferença de 7–8 pontos percentuais sugere que, para problemas de classificação de imagens estáticas com datasets pequenos, o treinamento ANN convencional seguido de conversão pode ser mais eficiente que o treinamento direto de SNNs com os algoritmos atuais.

Como perspectivas futuras, destacam-se: (i) validação completa da conversão CNN/SNN com medição do número de spikes e eficiência energética estimada; (ii) investigação com o dataset ISIC completo em maior resolução, aproveitando GPUs com maior VRAM; (iii) exploração de encodings alternativos ao latency encoding para imagens RGB; e (iv) comparação com modelos de referência da literatura para o DermaMNIST (ResNet-18, EfficientNet) para contextualizar as limitações atuais das SNNs.

Referências

BELLEC, G. et al. A solution to the learning dilemma for recurrent networks of spiking neurons. Nature Communications, v. 11, n. 3625, 2020.

MAASS, W. Networks of spiking neurons: the third generation of neural network models. Neural Networks, v. 10, n. 9, p. 1659–1671, 1997.

YANG, J. et al. MedMNIST v2: A Large-Scale Lightweight Benchmark for 2D and 3D Biomedical Image Classification. Scientific Data, v. 10, n. 41, 2023.

WUNDERLICH, T. C.; PEHLE, C. Event-based backpropagation can compute exact gradients for spiking neural networks. Scientific Reports, v. 11, n. 12829, 2021.

AKIBA, T. et al. Optuna: A Next-generation Hyperparameter Optimization Framework. KDD 2019. arXiv:1907.10902.

TURNER, J. P.; KNIGHT, J. C.; SUBRAMANIAN, A.; FURBER, S. mlGeNN: accelerating SNN inference using GPU-enabled neural networks. 2022.

BERGSTRA, J.; BENGIO, Y. Random Search for Hyper-Parameter Optimization. JMLR, v. 13, p. 281–305, 2012.

Anexos

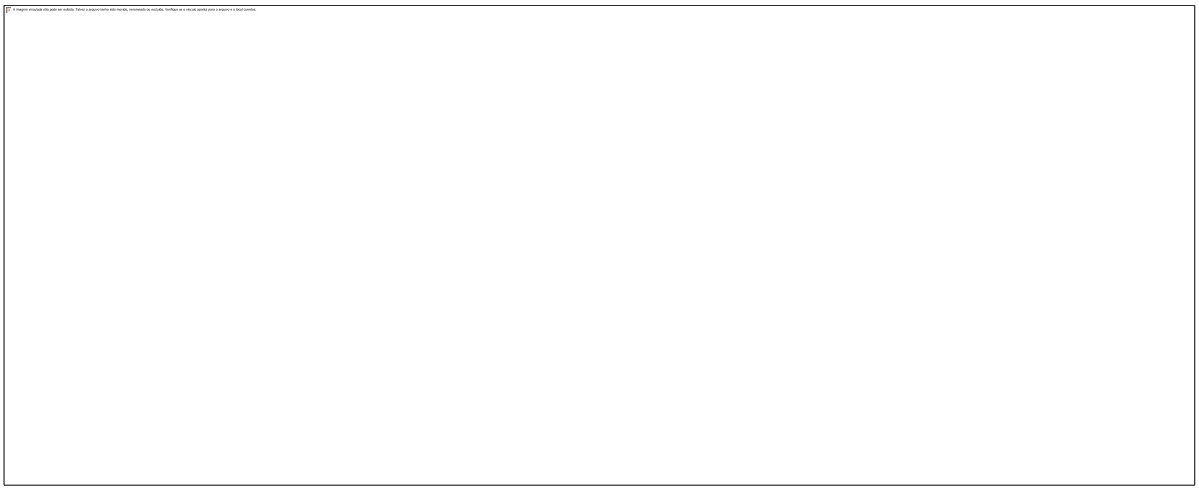


Figura 1 — Evolução da acurácia de validação ao longo dos experimentos de otimização. As barras azuis correspondem aos experimentos com EProp, as verdes ao EventProp, e a vermelha à abordagem CNN - SNN.



Figura 2 — Decaimento exponencial do traço de elegibilidade ao longo dos timesteps, para dois valores de max_example_timesteps.

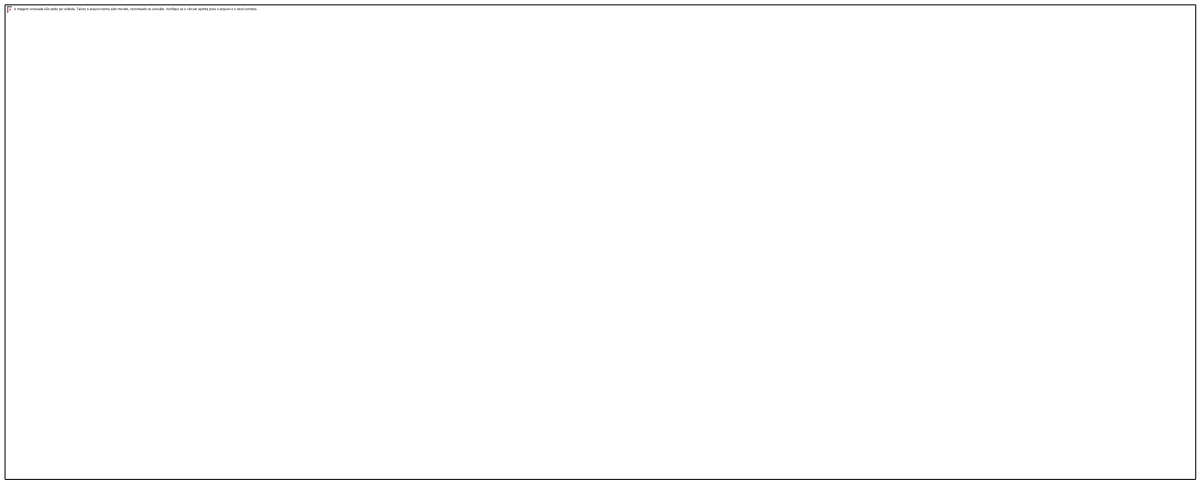


Figura 3 — Distribuição de spikes por threshold para imagens RGB 28×28. À esquerda, média e máximo de spikes por amostra. À direita, número de amostras sem nenhum spike. O threshold ideal equilibra cobertura de spikes e restrições de memória.



Figura 4 — Curva de acurácia de treino vs. validação para configuração Eprop com data augmentation. A rede entra em regime de overfitting.

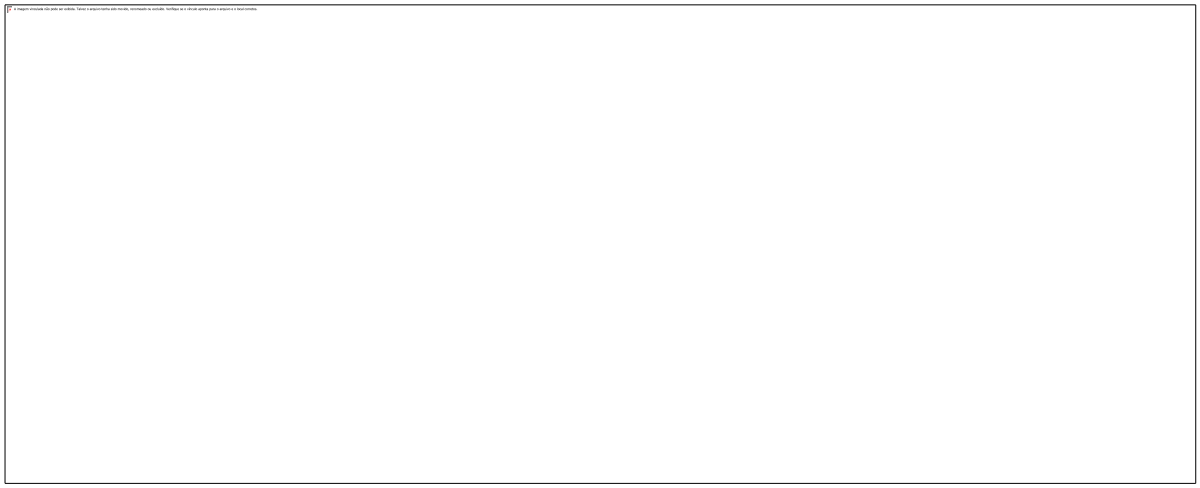


Figura 5 — Curva de acurácia de treino vs. validação para a melhor configuração EventProp. A rede atinge o pico de validação (71,95%) na época 3 e depois entra em regime de overfitting seguido de colapso parcial.

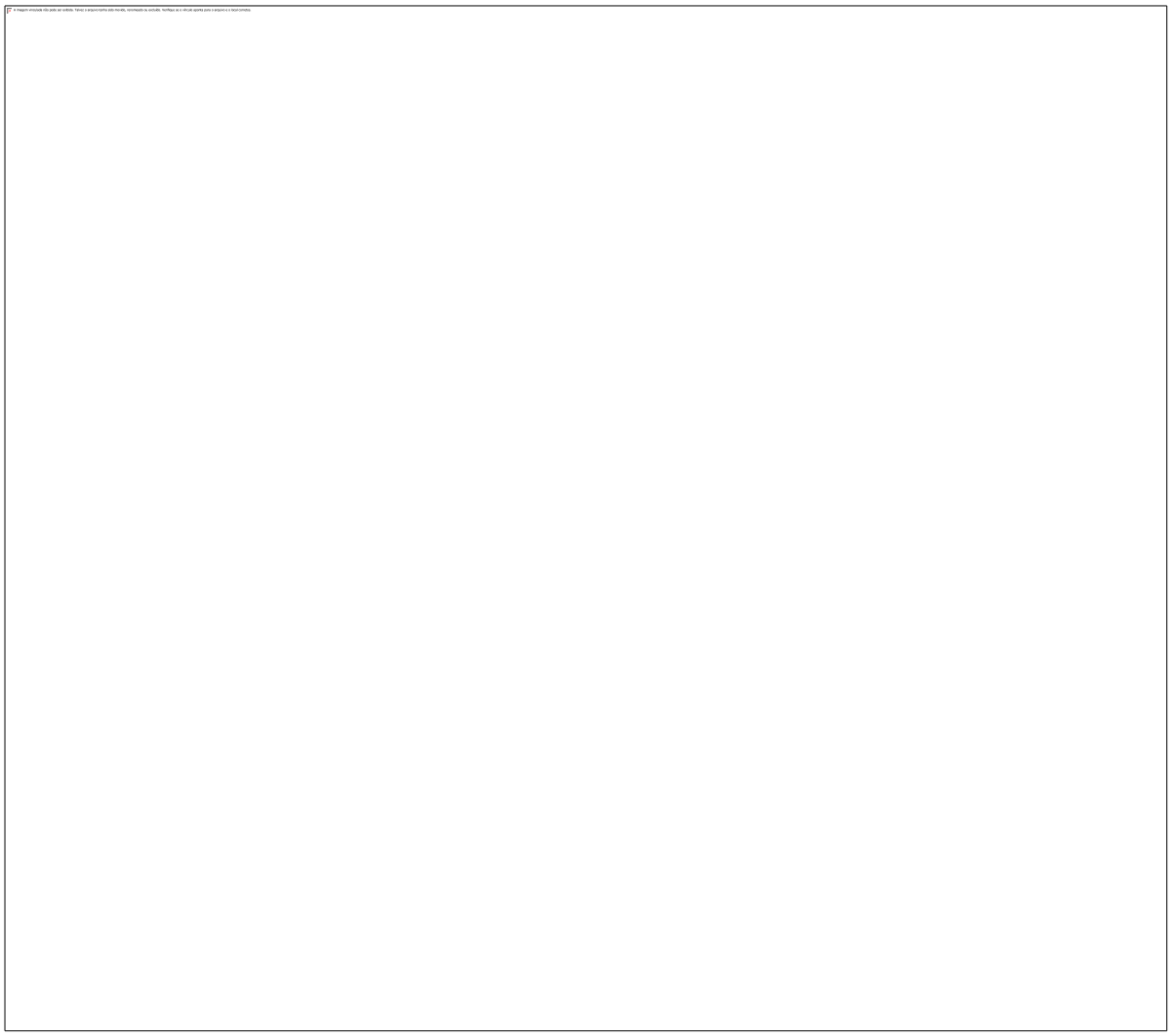


Figura 6 — Matriz de confusão do treinamento com eprop e data augmentation.

CÓDIGO: ET1085

AUTOR: MARIANA TIMBO DE OLIVEIRA

COAUTOR: ISRAEL MARTINS VALCAREGGI

ORIENTADOR: PATRICK CESAR ALVES TERREMATTE

TÍTULO: IA Explicável para Detecção Precoce de Resistência Antimicrobiana: Uma Abordagem para Redução do Risco Clínico

Resumo

A Inteligência Artificial (IA) vem se consolidando como ferramenta promissora na área da saúde, contribuindo para auxílio a diagnósticos mais rápidos e para a tomada de decisão clínica assistida por dados. Nesse cenário, destaca-se a Resistência Antimicrobiana (RAM), um dos maiores desafios de saúde pública mundial, responsável por cerca de 1,27 milhão de mortes anuais, além de agravar o prognóstico de pacientes com doenças infecciosas devido ao uso inadequado de antibióticos. A detecção de resistência por cultura microbiológica é demorada, dificultando decisões terapêuticas em caráter de urgência, o que evidencia a necessidade de soluções de apoio à decisão médica. Diante disso, esta pesquisa propôs o desenvolvimento de modelos de aprendizado de máquina supervisionado para prever o risco de resistência antimicrobiana, utilizando dados hospitalares da base MIMIC-IV, com variáveis demográficas, clínicas, microbiológicas e derivadas. Foram avaliados os modelos Random Forest, Regressão Logística e LightGBM, com otimização de hiperparâmetros via Optuna, técnicas de balanceamento de classes e interpretabilidade por meio do SHAP. O modelo Random Forest apresentou o melhor desempenho, com AUC-ROC de aproximadamente 0,88, sendo o histórico de infecção prévia, o número de diagnósticos distintos e a classificação Gram do microrganismo os fatores mais influentes nas previsões.

Palavras-chave: Resistência Antimicrobiana; Aprendizado de Máquina; SHAP; MIMIC-IV

TITLE: Explainable AI for early detection of Antimicrobial Resistance: An Approach to Reduce Clinical Risk

Abstract

Artificial Intelligence (AI) has consolidated its position as a promising tool in healthcare, contributing to faster diagnostic support and data-driven clinical decision-making. In this scenario, Antimicrobial Resistance (AMR) stands out as one of the greatest global public health challenges, responsible for approximately 1.27 million annual deaths, in addition to worsening the prognosis of patients with infectious diseases due to the inappropriate use of antibiotics. Resistance detection through microbiological culture is time-consuming, complicating urgent therapeutic decisions and highlighting the need for medical decision-support solutions. To address this, this study proposed the development of supervised machine learning models to predict the risk of antimicrobial resistance using hospital data from the MIMIC-IV database, incorporating demographic, clinical, microbiological, and derived variables. The Random Forest, Logistic Regression, and LightGBM models were evaluated, utilizing hyperparameter optimization via Optuna, class balancing techniques, and interpretability through SHAP. The Random Forest model achieved the best performance, with an AUC-ROC of approximately 0.88, where prior infection history, the number of distinct diagnoses, and the microorganism's Gram classification were the most influential factors in the predictions.

Keywords: Antimicrobial Resistance; Machine Learning; SHAP; MIMIC-IV

Introdução

A Inteligência Artificial (IA) tem transformado diversas áreas do conhecimento, e na saúde seu impacto se mostra particularmente relevante, ao possibilitar o processamento de grandes volumes de dados clínicos para apoiar diagnósticos, prognósticos e decisões terapêuticas com maior agilidade e precisão. Modelos de aprendizado de máquina, em especial, têm se destacado na identificação de padrões complexos em registros hospitalares, oferecendo suporte a profissionais de saúde em cenários nos quais o tempo de resposta é crítico para o desfecho clínico do paciente.

Nesse contexto, insere-se o problema da Resistência Antimicrobiana (RAM), atualmente entre as dez maiores causas de morte no mundo. O uso frequente e inadequado de antibióticos, associado a prescrições incompatíveis com o quadro clínico do paciente, figura entre os principais fatores que impulsionam o desenvolvimento de resistência bacteriana. Esse fenômeno compromete diretamente a eficácia dos tratamentos, contribuindo para piores prognósticos e para um risco aumentado de desfechos negativos, sendo responsável por cerca de 1,27 milhão de mortes anuais em todo o mundo. Além do impacto clínico, a RAM gera um ônus econômico significativo, ao tornar os tratamentos mais longos, complexos e onerosos.

A detecção da resistência bacteriana é tradicionalmente realizada por meio de testes de cultura microbiológica. Entretanto, esse processo é demorado, o que o torna pouco

viável para tratamentos de urgência e para a rotina de prescrição médica. Assim, torna-se necessário o desenvolvimento de soluções capazes de apoiar a decisão médica antes mesmo da disponibilização completa dos resultados de suscetibilidade antimicrobiana. Nesse cenário, o uso de técnicas de aprendizado de máquina supervisionado se mostra promissor, permitindo prever o risco de resistência bacteriana a determinados fármacos e analisar padrões presentes nos dados clínicos, com potencial de contribuir para decisões mais precisas, reduzir falhas terapêuticas e apoiar o uso mais adequado de antimicrobianos.

Diante desse cenário, este trabalho teve como objetivo avaliar modelos supervisionados de aprendizado de máquina para a predição de resistência antimicrobiana, a partir de registros de internação hospitalar da base de dados MIMIC-IV. Foram utilizadas variáveis demográficas, clínicas, microbiológicas e derivadas, sendo a variável-alvo definida como a condição de resistência ou suscetibilidade do isolado bacteriano ao antimicrobiano testado. Os modelos Random Forest, Regressão Logística e LightGBM foram avaliados e comparados, com aplicação de estratégias de balanceamento de classes e otimização de hiperparâmetros por meio da biblioteca Optuna, visando aprimorar o desempenho preditivo. Adicionalmente, foram empregadas técnicas de IA explicável, com destaque para o método SHAP (SHapley Additive exPlanations), para interpretar as saídas dos modelos e identificar os fatores clínicos mais associados ao risco de resistência.

Metodologia

O desenvolvimento da pesquisa seguiu três etapas principais, como está na figura do fluxograma do trabalho em anexo: preparação dos dados, treinamento dos modelos e avaliação de desempenho.

Os dados foram obtidos da base MIMIC-IV, desenvolvida pelo Massachusetts Institute of Technology em parceria com o Beth Israel Deaconess Medical Center (BIDMC), que reúne informações clínicas detalhadas de pacientes atendidos em Unidades de Terapia Intensiva (UTIs) e enfermarias, incluindo registros demográficos, resultados de exames laboratoriais e prescrições médicas. A extração foi realizada por meio do Google BigQuery, utilizando consultas em SQL para combinar a tabela de eventos microbiológicos (microbiologyevents) com informações demográficas (patients), dados de internação (admissions), registros de prescrição (prescriptions) e de permanência em UTI (icustays).

Após aplicação dos critérios de inclusão e exclusão, o conjunto final utilizado contou com 293.847 registros de suscetibilidade microbiológica, referentes a 23.333 pacientes únicos, com idade entre 18 e 90 anos. Pacientes acima de 90 anos foram excluídos devido aos protocolos de desidentificação da base, que alteram cronologias e limitam idades acima de 89 anos para conformidade com a norma HIPAA. A análise foi feita

sobre as suscetibilidades relativas aos 16 antimicrobianos mais frequentemente prescritos da base trabalhada.

Além da extração, foram feitos estudos das que seriam trabalhadas, entre elas variáveis utilizadas contemplaram quatro grupos: (i) contexto demográfico e de internação com idade, gênero, tipo de admissão, permanência em UTI e tempo de internação até a coleta; (ii) histórico clínico, com a variável que determina se houve ou não uma infecção prévia em internações anteriores e número de diagnósticos distintos anteriores; (iii) características microbiológicas, com o tipo de amostra do teste, microrganismo identificado na amostra (255 espécies), classificação Gram do microrganismo e antimicrobiano testado (16 fármacos); e (iv) a variável-alvo, binária, indicando resistência (1) ou suscetibilidade (0). A classificação Gram, não é disponível diretamente na base. Ela foi obtida por meio de mapeamento manual entre os microrganismos e sua classificação Gram-positiva ou Gram-negativa, com base em referências microbiológicas. Na análise exploratória, observou-se desbalanceamento da variável-alvo (~79% suscetibilidade e ~21% resistência), com predominância de bactérias Gram-negativas (~75%) e também uma distribuição equilibrada entre gêneros.

O presente trabalho optou por um modelo único, treinado com a combinação de 255 microrganismos e 16 antimicrobianos, em vez de modelos individuais por microrganismo, de modo que o usuário informe o microrganismo, o antimicrobiano e as características clínicas do paciente para obter a previsão de suscetibilidade.

No experimento, foram avaliados três modelos de aprendizado de máquina de classificação binária, sendo eles a Regressão Logística, que modela a probabilidade de pertencimento à classe por meio de função sigmoide; Random Forest, composto por múltiplas árvores treinadas de forma independente, com amostragem de dados e seleção aleatória de atributos a cada divisão, controlando a variância e reduzindo o overfitting; e LightGBM, baseado em boosting, no qual modelos são treinados sequencialmente para corrigir erros dos anteriores, capturando relações não lineares.

Durante o processo, aplicou-se a otimização de hiperparâmetros com a busca de hiperparâmetros realizada com a biblioteca Optuna, que utiliza modelos probabilísticos para explorar o espaço de busca de forma mais eficiente que grid ou random search. A otimização foi conduzida sobre o conjunto de treino, com validação cruzada de cinco folds e a AUC-ROC média como métrica-objetivo, ao longo de 100 tentativas por modelo. O Random Forest apresentou o maior ganho (AUC de 0,86 para 0,88), seguido do LightGBM (0,84 para 0,86) e da Regressão Logística (0,78 para 0,80), com as melhores configurações encontradas para cada modelo (por exemplo, para o Random Forest: n_estimators=311, max_depth=30, min_samples_split=3, min_samples_leaf=1, max_features=None).

Para mitigar o viés em favor da classe majoritária no desbalanceamento das classes, foram aplicadas técnicas de undersampling (Random Under-Sampling, que reduz aleatoriamente exemplos da classe majoritária) e oversampling (SMOTE - Synthetic Minority Over-sampling Technique, que gera exemplos sintéticos da classe minoritária a partir da similaridade entre pontos existentes, pela biblioteca Imbalanced-learn). Para evitar vazamento de dados, o SMOTE foi aplicado exclusivamente ao conjunto de

treino, após a divisão treino-teste, mantendo o conjunto de teste com sua distribuição original.

Os modelos foram avaliados por meio de Acurácia, Precisão, Recall, F1-score e AUC-ROC, sendo esta última adotada como métrica principal por ser menos sensível ao desbalanceamento das classes.

Após o treinamento, para tornar os modelos mais interpretáveis, transparentes e confiáveis, foi aplicado o método SHAP (SHapley Additive exPlanations), que permite analisar a contribuição de cada variável para as previsões, em níveis global e individual. A análise foi conduzida sobre um subconjunto de 1.000 amostras do conjunto de teste, de modo a reduzir o custo computacional mantendo representatividade.

Os experimentos acima foram conduzidos utilizando os recursos do Centro de Alto Desempenho Computacional (NPAD/UFRN), com ambiente de execução composto por dois processadores Intel Xeon de 64 threads e uma GPU NVIDIA Tesla T4 com 15 GB de memória dedicada. Na implementação, foram utilizadas as bibliotecas Scikit-learn, SHAP, LightGBM e Imbalanced-learn, em linguagem Python.

Resultados e Discussões

Avalicação de desempenho dos modelos

Os três modelos (Random Forest, Regressão Logística e LightGBM) foram avaliados nos cenários com dados originais (imbalanced), com técnicas de desbalanceamento (Random Under-Sampling e com SMOTE), considerando as métricas de Acurácia, Precisão, Recall, F1-score e AUC-ROC. Em anexo está a figura da tabela de todos os resultados de desempenho dos modelos.

O Random Forest com dados originais apresentou o melhor desempenho geral, alcançando AUC-ROC de 0,88 no conjunto de dados original desbalanceado, além da maior acurácia (0,85) e precisão (0,77) entre todos os cenários testados. O LightGBM obteve desempenho intermediário, com AUC-ROC de 0,86 no cenário original, enquanto a Regressão Logística apresentou o menor desempenho, com AUC-ROC de 0,80, refletindo sua natureza linear diante de um problema com relações mais complexas entre variáveis.

As estratégias de balanceamento impactaram os modelos de formas distintas. De modo geral, o Random Under-Sampling aumentou expressivamente o recall, por exemplo, de 0,40 para 0,80 no Random Forest, indicando maior capacidade de identificar corretamente as amostras resistentes. Entretanto, esse ganho veio acompanhado de redução na acurácia e, principalmente, na precisão (de 0,77 para 0,50 no Random Forest), o que aumenta o número de falsos positivos. Do ponto de vista clínico, um recall mais alto pode ser desejável, pois reduz a chance de classificar erroneamente um caso resistente como suscetível; contudo, a queda de precisão eleva o risco de uso

desnecessário de antimicrobianos de amplo espectro. O SMOTE, por sua vez, apresentou resultados intermediários entre o cenário original e o undersampling, com ganho de recall mais moderado e menor perda de precisão.

Apesar do maior recall obtido com o Random Under-Sampling, essa abordagem não foi considerada a mais adequada como modelo principal, pois a redução aleatória de exemplos da classe majoritária pode descartar informações clinicamente relevantes presentes nos dados originais. De forma semelhante, embora o SMOTE tenha apresentado resultados competitivos, a geração de perfis clínicos sintéticos pode introduzir ruído artificial e distorcer a distribuição epidemiológica real da resistência antimicrobiana. Por esses motivos, priorizou-se o Random Forest treinado sobre o conjunto de dados original, por preservar a prevalência clínica real e por ter demonstrado capacidade inerente de lidar com o desbalanceamento, apresentando o melhor equilíbrio entre AUC-ROC, acurácia, precisão e robustez preditiva.

Comparação com a literatura

Ao comparar o desempenho obtido (AUC-ROC de aproximadamente 0,88) com estudos relacionados, observa-se que valores mais altos podem ser alcançados em cenários mais específicos. No estudo de Silveira (2023), no qual modelos foram treinados separadamente para cada antimicrobiano avaliado, o Random Forest atingiu AUC-ROC de até 0,98 para a predição de resistência à vancomicina, com valores entre 0,81 e 0,84 para os demais fármacos. Esses resultados refletem um cenário de classificação mais restrito, no qual cada modelo aprende padrões associados a um único fármaco, reduzindo a complexidade da tarefa. Já o presente trabalho propõe uma abordagem unificada, capaz de prever simultaneamente múltiplas combinações entre microrganismos e antimicrobianos, o que aumenta significativamente a complexidade do problema.

Entre os estudos analisados, o trabalho de Rosa et al. (2024) é o metodologicamente mais próximo desta pesquisa, por também utilizar dados derivados da base MIMIC (ainda que de uma versão anterior) e propor um modelo geral para predição de resistência antimicrobiana. Naquele estudo, o melhor desempenho foi obtido pelo XGBoost, com AUC de 0,84 no conjunto original e 0,84 após aplicação do SMOTE, enquanto o Random Forest alcançou valores de AUC entre 0,78 e 0,79, inferiores aos obtidos neste trabalho. Essa diferença pode estar relacionada ao conjunto de atributos utilizado: enquanto Rosa et al. empregaram majoritariamente variáveis demográficas e microbiológicas básicas (idade, gênero, tipo de amostra, antimicrobiano testado e classificação Gram), este estudo incorporou também variáveis clínicas derivadas, como tempo de internação até a coleta, número de comorbidades e histórico de infecção prévia, o que parece contribuir para uma representação mais abrangente do contexto clínico do paciente e, consequentemente, para maior capacidade discriminativa dos modelos. Ressalva-se, contudo, que diferenças entre versões da base, procedimentos de pré-processamento e composição das coortes também podem ter contribuído para a diferença observada, dificultando isolar o efeito exclusivo das variáveis derivadas.

Em síntese, a análise comparativa demonstra que o modelo Random Forest desenvolvido neste trabalho apresenta desempenho preditivo robusto e competitivo, considerando a complexidade do cenário de predição proposto, simultâneo para

múltiplas combinações de microrganismo e antimicrobiano, ao invés de modelos individuais. Os resultados reforçam a importância de incorporar variáveis clínicas contextuais na modelagem da resistência antimicrobiana.

Análise de explicabilidade (SHAP)

A análise de interpretabilidade por meio do SHAP revelou padrões consistentes entre os três modelos avaliados, apesar das diferenças estruturais entre os algoritmos, o que sugere que os fatores identificados representam sinais clínicos robustos associados ao risco de resistência antimicrobiana. Os gráficos das importancias dos atributos sobre cada modelo estão nas figuras em anexo.

O histórico de infecção prévia (prior_infection) destacou-se como uma das variáveis mais importantes em todos os modelos: valores mais altos dessa variável foram associados a valores SHAP positivos, indicando que pacientes com infecções prévias apresentam maior probabilidade prevista de resistência. Esse resultado é consistente com achados de Lewin-Epstein et al. (2021), que também identificaram infecções prévias como fator relevante na predição de resistência bacteriana, embora o estudo referido enfatize especificamente infecções previamente resistentes, ambos os trabalhos convergem quanto à relevância de infecções recorrentes como preditor.

Indicadores clínicos também mostraram forte contribuição preditiva. O número de diagnósticos distintos (num_distinct_diagnoses) e o tempo de internação até a coleta da amostra (length_of_stay) figuraram entre as variáveis mais influentes. A multimorbidade está associada a maior exposição a antibióticos e a procedimentos invasivos, contexto classicamente relacionado ao desenvolvimento de resistência antimicrobiana (Chahine et al., 2026). Além disso, internações mais longas aumentam a probabilidade de colonização e infecção por patógenos multirresistentes associados ao ambiente hospitalar (Santos, 2004).

No âmbito microbiológico, a presença de Enterococcus faecium destacou-se como um dos principais preditores positivos de resistência em todos os modelos, achado clinicamente esperado, dado que essa espécie apresenta altas taxas de resistência intrínseca e adquirida, incluindo resistência à vancomicina e a múltiplas classes terapêuticas (Hollenbeck e Rice, 2012; Geraldes et al., 2022). Além disso, o E. faecium é classificado pela Organização Mundial da Saúde como patógeno de alta prioridade, com opções terapêuticas extremamente limitadas (World Health Organization, 2024).

Limitações

A principal limitação deste estudo é a ausência de validação externa dos modelos desenvolvidos: todo o processo de treinamento e avaliação foi conduzido exclusivamente com dados da base MIMIC-IV, o que pode limitar a generalização dos achados para outros contextos clínicos e introduzir vieses populacionais, institucionais e operacionais. Embora a validação externa seja uma etapa importante para avaliar a robustez do modelo, a obtenção de uma base equivalente é desafiadora, uma vez que, até o momento, não há conjuntos de dados públicos com combinação comparável de registros de suscetibilidade microbiológica, prescrições, histórico de internação e informações clínicas em escala semelhante à do MIMIC-IV.

Conclusão

Este trabalho investigou a aplicação de modelos de aprendizado de máquina para a predição de resistência antimicrobiana a partir de dados clínicos da base MIMIC-IV. O modelo Random Forest apresentou o melhor desempenho geral, com AUC-ROC de aproximadamente 0,88, superando ou apresentando resultado competitivo em relação a estudos semelhantes reportados na literatura, mesmo diante de um cenário de predição mais complexo, ao abranger simultaneamente múltiplas combinações de microrganismos e antimicrobianos.

A análise de explicabilidade com SHAP permitiu identificar as principais variáveis associadas às previsões dos modelos, destacando-se o histórico de infecção prévia, o número de diagnósticos distintos, o tempo de internação até a coleta da amostra e características microbiológicas, como a espécie do microrganismo e sua classificação Gram. Esses achados se mostraram consistentes entre os diferentes algoritmos avaliados e alinhados a evidências da literatura, reforçando a confiabilidade e a coerência clínica dos resultados obtidos.

Dessa forma, considerando a consistência dos resultados e sua fundamentação clínica, os modelos desenvolvidos demonstram potencial para apoiar a tomada de decisão médica, especialmente na identificação precoce do risco de resistência antimicrobiana. Uma vez validados clinicamente, esses achados podem contribuir para decisões de prescrição mais informadas e para o uso mais adequado de antibióticos, com potenciais benefícios para os desfechos terapêuticos dos pacientes.

Como principal limitação, destaca-se a ausência de validação externa, uma vez que todo o processo foi conduzido exclusivamente com dados de uma única instituição. Como trabalhos futuros, pretende-se realizar a validação externa dos modelos em bases de dados de outras instituições, a fim de avaliar sua capacidade de generalização em diferentes contextos clínicos, além de explorar a incorporação de novas variáveis clínicas e a aplicação da abordagem em cenários de decisão em tempo real.

Referências

CHAHINE, E. B.; WATANABE, J. H.; DONG, M. T.; ANDRADE, R. M. Beyond resistance: prioritizing safety in antimicrobial stewardship for older adults. The Senior Care Pharmacist, v. 41, n. 3, p. 88-97, 2026.

GERALDES, C.; TAVARES, L.; GIL, S.; OLIVEIRA, M. Enterococcus virulence and resistant traits associated with its permanence in the hospital environment. Antibiotics, v. 11, n. 7, p. 857, 2022.

HE, H.; GARCIA, E. A. Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering, v. 21, n. 9, p. 1263-1284, 2009.

HOLLENBECK, B. L.; RICE, L. B. Intrinsic and acquired resistance mechanisms in enterococcus. Virulence, v. 3, n. 5, p. 421-433, 2012.

JOHNSON, A.; BULGARELLI, L.; POLLARD, T.; GOW, B.; MOODY, B.; HORNG, S.; CELI, L. A.; MARK, R. MIMIC-IV. PhysioNet, versão 3.1, 2024.

LEWIN-EPSTEIN, O.; BARUCH, S.; HADANY, L.; STEIN, G. Y.; OBOLSKI, U. Predicting antibiotic resistance in hospitalized patients by applying machine learning to electronic medical records. Clinical Infectious Diseases, v. 72, n. 11, p. e848-e855, 2021.

MINISTÉRIO DA SAÚDE (Brasil). Resistência antimicrobiana (RAM). Portal Gov.br, 2026.

ROSA, S. d. C. L.; CARVALHO, N. I.; DUANI, H. Machine learning for microbial resistance prediction. Journal of Health Informatics, v. 16, n. especial, 2024.

SANTOS, N. de Q. A resistência bacteriana no contexto da infecção hospitalar. Texto & Contexto - Enfermagem, v. 13, n. esp., p. 64-70, 2004.

SILVEIRA, B. P. Uso de aprendizado de máquina para predizer suscetibilidade antibiótica através de dados de pacientes hospitalizados. Dissertação (Mestrado) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 2023.

WORLD HEALTH ORGANIZATION. Global antimicrobial resistance and use surveillance system (GLASS) report 2023. Genebra: WHO, 2023.

WORLD HEALTH ORGANIZATION. WHO updates list of drug-resistant bacteria most threatening to human health. WHO News Release, 2024.

Anexos



Fluxograma da metodologia



Gráfico da importância de atributos no modelo Random Forest



Gráfico da importância de atributos no modelo de Regressão logística

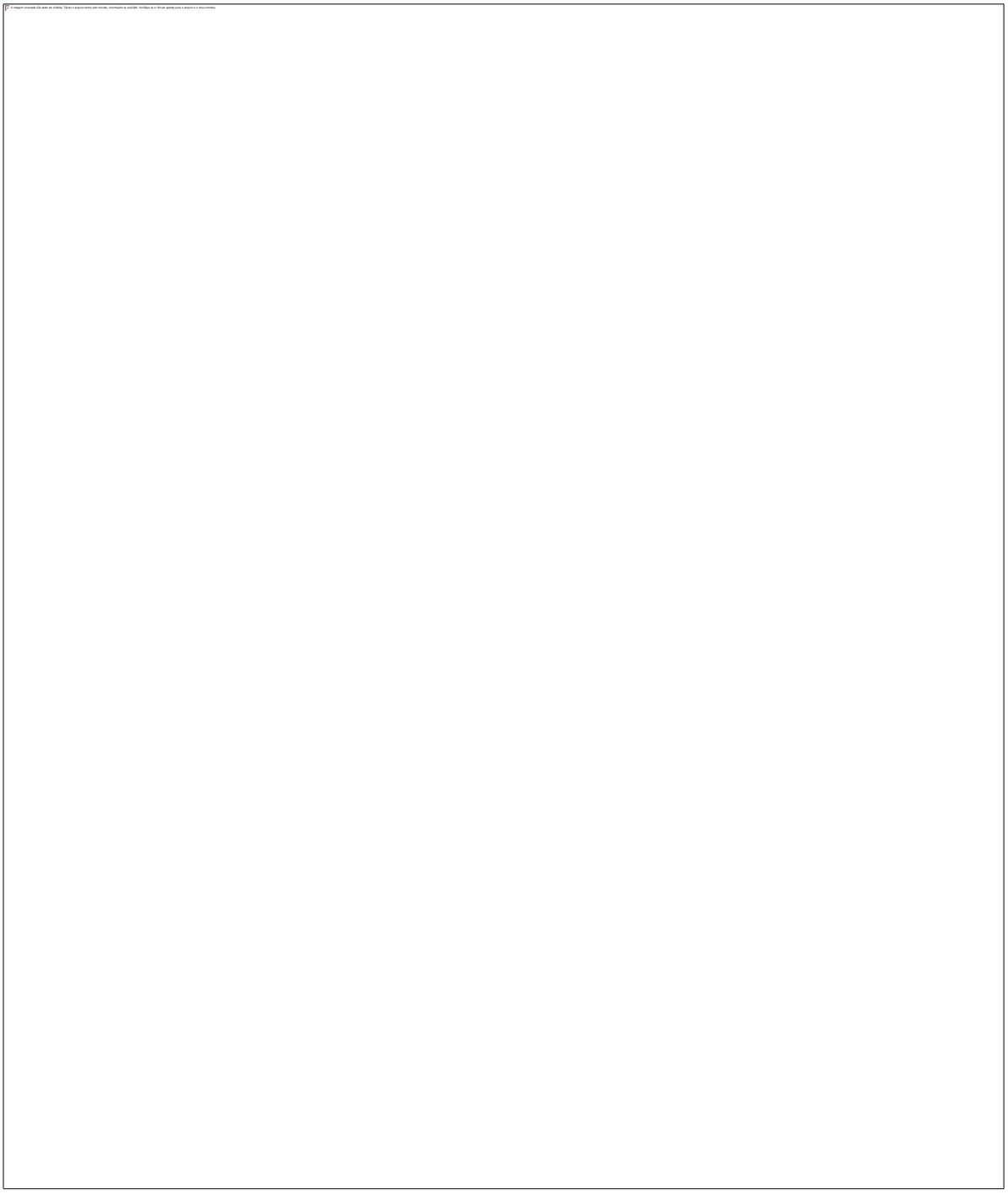


Gráfico da importância de atributos no modelo LightGBM

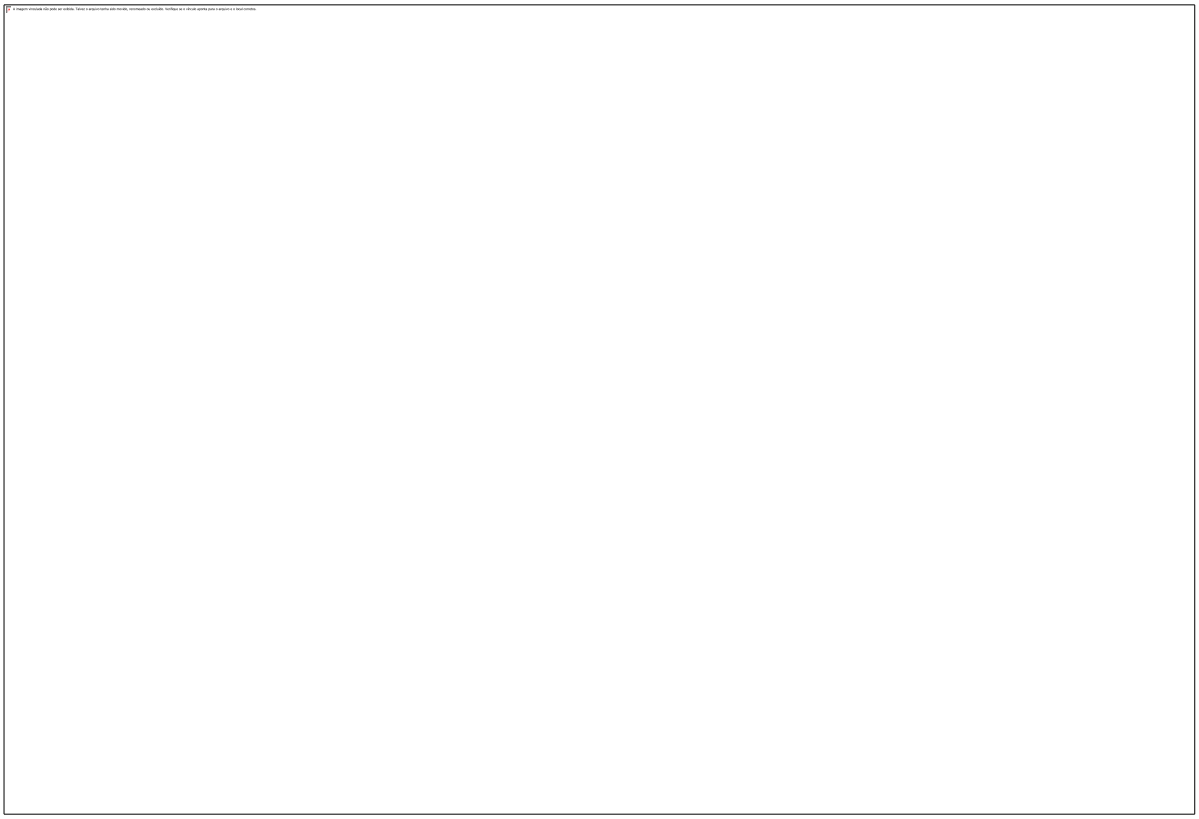


Figura da tabela de resultado dos modelos

CÓDIGO: ET1133

AUTOR: MARINA MEDEIROS DE ARAÚJO LEITE

COAUTOR: ALDO QUEIROZ CABRAL DA FONSECA TINOCO

COAUTOR: JOSÉ IVO SCHWADE ARAÚJO

ORIENTADOR: LEONARDO ENZO BRITO DA SILVA

TÍTULO: Um Estudo Sobre Estratégias de Codificação Posicional para Modelos Fusion ART Multicamadas em Tarefas de Classificação: Índices de Posição Normalizados e Máscara Binária

Resumo

As redes neurais baseadas na Teoria da Ressonância Adaptativa (ART) oferecem uma estrutura estável para o aprendizado incremental e baseado em correspondência match-based, mas seu custo computacional aumenta significativamente quando estendidas para arquiteturas multicamadas. Neste trabalho, investigamos estratégias para reduzir o custo computacional de modelos Fusion ART multicamadas supervisionados, preservando seu desempenho de classificação. Nossa abordagem baseia-se em restringir as entradas de canais de características entre camadas aos k maiores valores de ativação, aproveitando o mecanismo winner-takes-all intrínseco aos modelos ART. Demonstramos que a aplicação ingênua dessa estratégia leva a um overfitting severo e a uma generalização deficiente, especialmente em arquiteturas mais profundas. Em seguida, demonstramos como aplicar a codificação posicional a esse conjunto de dados truncado para solucionar tais problemas de overfitting, explorando duas arquiteturas distintas: uma com um canal dedicado ao processamento do vetor posicional e outra que apenas soma o vetor posicional ao vetor de características. Duas estratégias de processamento diferentes foram aplicadas à arquitetura com canal adicional no que diz respeito ao vetor posicional: normalização e tratamento com máscara binária.

Palavras-chave: Ressonância Adaptativa.Aprendizado Multicamadas.Codificação Posicional

TITLE: A Study on Positional Encoding Strategies for Multilayer Fusion ART Models in Classification Tasks: Normalized Position Indices and Binary Masks

Abstract

Adaptative Resonance Theory (ART) neural networks provide a stable framework for incremental and match-based learning, but their computational cost increases significantly when extended to multi-layer architectures. In this work, we investigate strategies to reduce the computational cost of supervised multi-layer Fusion ART models while preserving their classification performance. Our approach is based on restricting inter-layer feature-channel inputs to the top-k activation values, leveraging the winner-takes-all mechanism intrinsic to ART models. We demonstrate that naively applying this strategy leads to severe overfitting and poor generalization, especially in deeper architectures. We then demonstrate how we can apply positional encoding to this truncated dataset to solve those overfitting problems, exploring two different architectures: one with a channel dedicated to processing the positional vector and other only adding up the positional vector to the feature vector. Two different processing strategies were applied to the extra-channel architecture, with regard to the positional vector: normalization and binary mask treatment.

Keywords: Adaptive Resonance Theory.Multilayer Learning.Positional Encoding.

Introdução

A Teoria da Ressonância Adaptativa (ART) é uma abordagem sobre como o cérebro humano percebe informações novas e já vivenciadas (GROSSBERG, 1976a, 1976b), descrevendo o aprendizado de novos padrões como um processo baseado em match. Essa teoria culminou na idealização e no desenvolvimento de uma série de modelos de redes neurais (BRITO DA SILVA; ELNABARAWY; WUNSCH II, 2019), como, por exemplo, ART **(GROSSBERG, 1976a)** ,ARTMAP e Fuzzy ART **(CARPENTER; GROSSBERG, 1987; CARPENTER; GROSSBERG; REYNOLDS, 1991; CARPENTER; GROSSBERG; ROSEN, 1991)**. Um dos modelos da família de redes ART é a Fusion ART **(TAN; CARPENTER; GROSSBERG, 2007)**, que utiliza uma arquitetura multicanal para processar diferentes tipos de dados e trocar informações entre eles, visando capturar relações sutis entre fontes de dados distintas. A entrada de um modelo Fusion ART pode ser definida por Ick = (Ick1, Ick2, Ick3...Ickn), onde Icki ∈ [0, 1] denota a entrada i para o canal ck. É importante notar que a entrada não precisa estar presente em todos os canais; para os canais que não recebem entrada, os vetores de entrada são inicializados com valores iguais a 1. Os canais de um modelo Fusion ART podem ser compostos por qualquer modelo ART não supervisionado, e um modelo Fusion ART supervisionado pode ser obtido utilizando-se um canal para processar características e outro para processar rótulos, de forma semelhante à arquitetura ARTMAP. No entanto, esses modelos Fusion ART supervisionados podem ser organizados sequencialmente em camadas para formar um modelo de N camadas, onde a saída das funções de ativação do canal de características do modelo na camada n, para cada amostra, é fornecida como entrada para o canal de características do modelo na camada n + 1, mantendo-se os mesmos rótulos. Essa abordagem proporciona um aumento na acurácia, embora esse ganho venha acompanhado de um aumento no custo computacional (Figura 1).

Metodologia

O alto custo computacional decorre do fato de que o tamanho do vetor de valores de ativação é diretamente proporcional ao número total de nós do modelo, crescendo a cada camada, já que esses valores são calculados para cada vetor de pesos e para cada amostra.

No entanto, como o modelo Fusion ART seleciona a categoria mais provável aplicando uma estratégia winner-takes-all, ao reduzirmos as ativações na entrada do canal de características apenas aos k valores mais altos preservamos o mesmo vencedor com uma ocupação de memória muito menor.

Contudo, como essa abordagem extrai apenas os k maiores valores independentemente de qual categoria é a vencedora, e cada valor corresponde a uma característica da amostra, ela acaba trocando incorretamente características entre posições do conjunto de dados truncado. A ordem de apresentação das características passa a variar a cada camada e, com isso, perde-se a correspondência entre a posição que um valor ocupa no vetor truncado e a característica que o originou. **Codificação Posicional.** Recuperar essa correspondência exige transmitir ao modelo, junto de cada valor de ativação, a posição que ele ocupava antes do truncamento. Para isso,aplicamos uma estratégia de codificação posicional **(VASWANI et al., 2017)** utilizando o vetor r, que contém a informação posicional de cada valor de ativação. Observamos os resultados de duas abordagens distintas para representar esses dados posicionais:

1. **Índices Normalizados.** Para um valor de ativação arbitrário Tn em um vetor de ativação T de tamanho N, o elemento correspondente rn do vetor de índices r é definido por $rn = n / (N - 1)$.
2. **Máscara Binária.** Nesta abordagem, o vetor r foi definido para conter a informação posicional associada aos k valores de ativação mais altos de T e, a partir dele, construiu-se o vetor binário $B \in \{0,1\}^N$, tal que $B_j = 1$ se $j \in r$ e $B_j = 0$ se $j \notin r$. O vetor de ativações filtrado pelo top-k é preservado como entrada do canal de características, e B substitui r no canal posicional.

Arquiteturas

1. **Somar a informação de posição diretamente à entrada do canal de características.** Para cada amostra, após selecionar os k valores de ativação mais altos do vetor T, seus valores de codificação posicional correspondentes do vetor r são obtidos e normalizados. Assim, a representação final é construída por uma soma elemento a elemento entre os dois vetores: $x' = T(k) + r(k)$, onde T(k) contém os k valores de ativação selecionados e r(k) a informação posicional associada a essas ativações (Figura 2). Essa abordagem preserva a escala das ativações selecionadas e as relações estruturais entre características. Além disso, ela não necessariamente corrompe os vetores de ativação originais, já que vetores escolhidos aleatoriamente em espaços de alta dimensionalidade tendem a ser quase ortogonais **(BISHOP; BISHOP, 2023)**, permitindo que a rede processe informações posicionais e semânticas de maneira relativamente independente.
2. **Criar um novo canal no modelo para processar o vetor resultante.** Aproveitando o aspecto de múltiplas fontes de dados dos modelos Fusion ART, criamos um canal dedicado à informação posicional, separando x e r em dois

canais diferentes (Figura 3). Em seguida, aplicamos entre cada camada do modelo a função softmax com um parâmetro de temperatura T **(HINTON; VINYALS; DEAN, 2015)**, para melhor ponderar a importância dos valores de ativação mais altos: $\sigma(z)_i = \exp(z_i/T) / \sum_{j=1}^{K} \exp(z_j/T)$. Como utilizamos um Fuzzy ART em cada canal do Fusion ART supervisionado, aplicamos as técnicas de pré-processamento exigidas pelos modelos ART baseados em lógica fuzzy: normalização das características para o intervalo [0,1] e codificação por complemento das amostras, aumentando a entrada Ick com um vetor complementar tal que $Ick_i = 1 - Ick_i$.

CONFIGURAÇÃO EXPERIMENTAL

Hardware e Reprodutibilidade

Todos os experimentos foram conduzidos exclusivamente na infraestrutura de computação de alto desempenho do Núcleo de Processamento de Alto Desempenho (NPAD) da Universidade Federal do Rio Grande do Norte (UFRN). A partição utilizada apresenta sete nós computacionais com 2x **AMD Genoa 9654 DP/UP 96C/192T, 3 TB de RAM DDR5, 8x NVMe Micron 7450 PRO 3.8TB em RAID0 e Infiniband Connect X6 100Gb/s.**

Conjuntos de Dados

A avaliação utilizou dois conjuntos de dados de imagens de dígitos manuscritos, amplamente empregados em tarefas de aprendizado de máquina: o ORHD **(ALPAYDIN; KAYNAK, 1998)**, com 5.620 instâncias descritas por 64 características, e o USPS **(LECUN et al., 1989)**, com maior variabilidade, totalizando 9.298 amostras e 256 características por imagem. Ambos foram pré-processados conforme os requisitos dos modelos adotados, garantindo compatibilidade e comparação justa entre os experimentos.

Parâmetros e Configuração

Para nosso Fusion ART supervisionado, utilizamos um modelo Fuzzy ART em cada canal, com os seguintes parâmetros:

Canal de Rótulos: Gamma (γ) = 0,0 | Rho (ρ) = 1,0 | Alpha (α) = 10^{-3} | Beta (β) = 1,0
Canais de Características e de Codificação Posicional: Gamma (γ) = 0,5 | Rho (ρ) = 0,0 | Alpha (α) = 10^{-3} | Beta (β) = 1,0

Realizamos o experimento variando k de k = 1 a k = 5, além de testar utilizando todos os valores de ativação, e utilizando 6 camadas em cada iteração.

Resultados e Discussões

Baseline

Como dito anteriormente, o modelo Fusion ART seleciona a categoria mais provável para uma amostra aplicando uma estratégia winner-takes-all. Assim, se reduzirmos as ativações na entrada do canal de características apenas aos k valores mais altos, preservamos o mesmo vencedor com uma ocupação de memória muito menor.

No entanto, isso leva a um problema: como o modelo multicamadas extrai apenas os k valores de ativação mais altos independentemente da posição, especialmente ao lidar com dados fortemente dependentes da ordem de apresentação de suas características, podemos acabar com valores muito semelhantes para amostras completamente diferentes, confundindo consequentemente o modelo.

O impacto dessa abordagem ingênua — de simplesmente restringir os valores de entrada aos k valores de ativação mais altos — em diferentes profundidades de rede, para os conjuntos de dados USPS e ORHD, é ilustrado na Figura 4. Os resultados dos testes mostram uma queda acentuada na acurácia à medida que o número de camadas aumenta.

Quando todos os valores são utilizados (k = 0), o modelo alcança as acurácias de teste mais altas e mais estáveis para todos os números de camadas. Em ambos os conjuntos de dados, arquiteturas mais profundas inicialmente melhoram o desempenho, atingindo um pico com duas camadas, após o qual a acurácia de teste degrada gradualmente. Esse comportamento sugere que, embora o empilhamento de camadas Fusion ART aumente o poder representacional, a profundidade excessiva introduz uma superespecialização.

Em contrapartida, em todas as configurações truncadas, isto é, com k entre 1 e 5, a acurácia de treinamento permanece alta, frequentemente ultrapassando 95% mesmo em modelos mais profundos. Contudo, a acurácia de teste sofre um colapso acentuado após a primeira camada. Essa divergência entre treino e teste indica um claro caso de overfitting.

Esse fenômeno pode ser atribuído à perda de informação estrutural causada pela remoção das ativações. Embora o mecanismo winner-takes-all garanta que a categoria mais ativada seja preservada, descartar as ativações de menor ranking remove pistas contextuais necessárias para um bom desempenho.

De modo geral, esses resultados mostram que a seleção ingênua de top-k ativações compromete severamente a generalização.

Somando *r* à entrada do canal de características

As Figuras 5 e 8 apresentam uma comparação da acurácia de teste entre diferentes valores de k, considerando tanto a presença quanto a ausência da função softmax aplicada ao vetor de ativações top-k. Os resultados revelam que essa estratégia supera consistentemente o baseline apenas com softmax nas camadas mais profundas, para $k \geq 2$, estabilizando-se em torno de 0,65–0,80 de acurácia, em vez de colapsar. De modo

geral, os achados sugerem que adicionar r à entrada do canal de características mitiga a degradação de acurácia com a profundidade observada, com o benefício se tornando mais pronunciado à medida que k aumenta. Isso indica que a informação contextual proveniente de r ajuda a preservar o desempenho em arquiteturas mais profundas.

Em relação ao uso de memória, as Figuras 6 e 9 revelam que, em k = 1, o uso do baseline apenas com softmax colapsa para um crescimento quase nulo do número de categorias, enquanto a abordagem proposta, com ou sem softmax, mantém um número praticamente constante de categorias. Para $k \geq 2$, no entanto, esse padrão se inverte, e a variante com softmax sem codificação posicional atinge as maiores contagens de categorias, enquanto as variantes que incorporam a informação de r no canal de entrada apresentam um crescimento mais lento no número de categorias e se estabilizam em valores mais baixos. Em conjunto, esses resultados indicam que aumentar k amplifica a proliferação de categorias na ausência de codificação posicional, enquanto adicionar r ao canal de entrada restringe esse crescimento. Isso sugere que injetar informação posicional na entrada tem um efeito regularizador sobre o número de categorias distintas aprendidas, particularmente em contextos de maior tamanho.

Por fim, as Figuras 7 e 10 apresentam os tempos de treinamento para os conjuntos de dados ORHD e USPS revelam que, ao utilizar todos os valores de ativação, o tempo de treinamento cresce consideravelmente à medida que o número de camadas aumenta. Para k = 1, ambas as variantes da estratégia proposta exibem tempo de execução quase constante com o aumento da profundidade, sendo que o baseline baseado em softmax alcança o menor tempo de execução geral. Para $k \geq 2$, no entanto, a abordagem de adicionar o r ao input incorre em um aumento moderado no custo computacional em relação ao baseline em profundidades intermediárias; contudo, ela escala de forma mais favorável em profundidades maiores, eventualmente convergindo com — ou ficando abaixo — do tempo de execução do baseline na sexta camada, particularmente para valores maiores de k.

Criando um novo canal no modelo para processar o vetor *r*

Realizamos uma comparação considerando duas condições: o uso do softmax entre camadas e o uso da codificação posicional. Assim, obtivemos quatro configurações para cada uma das arquiteturas. Além disso, analisamos o cenário utilizando todos os valores de ativação, para 5 valores diferentes de k, e para 4 camadas do Fusion ART. As arquiteturas foram comparadas levando em consideração a presença e a ausência de codificação posicional.

Abordagem de Máscara Binária:

Observamos que a maior acurácia foi alcançada pela segunda camada do modelo que utiliza todos os valores de ativação, sem codificação posicional e sem a aplicação da função softmax. Este modelo será doravante referido como Modelo base.

As Figuras 11 e 14 mostram que, considerando tanto o conjunto de dados ORHD quanto o USPS, a abordagem não conseguiu superar a acurácia do modelo base, exceto quando todos os valores de ativação foram utilizados na segunda camada do conjunto de dados

ORHD. No entanto, quando o modelo multicamadas está restrito às top-k ativações, observamos que a presença da codificação posicional permite uma melhor preservação da sua performance; de modo inverso, quando ela está ausente percebe-se uma maior degradação da acurácia.

Também é notável que, à medida que o número de valores de ativação utilizados aumenta no intervalo de 1 a 5, a acurácia tende a se estabilizar entre 0,6 e 0,8. Além disso, a abordagem de Máscara Binária apresenta um desempenho geral positivo nesse aspecto, particularmente para o conjunto de dados ORHD, onde as diferenças foram mais pronunciadas.

Em relação ao uso de memória, a abordagem de Máscara Binária torna o modelo muito mais eficiente em termos de criação de categorias, especialmente para valores menores de k. O uso ou não da função softmax não fez diferença notável neste aspecto, como mostram as Figuras 12 e 15.

As Figuras 13 e 16 mostram o mesmo padrão de tempo de treinamento para os conjuntos de dados USPS e ORHD. Em praticamente todas as configurações, o baseline é a estratégia mais custosa, escalando de forma acentuada com a profundidade e atingindo os maiores tempos de execução em $k \geq 3$, exceto em $k = 1$, onde ela colapsa para a configuração mais barata, de aproximadamente 1 segundo por camada, uma vez removido o softmax. O uso da Máscara Binária mantém tempos de execução praticamente constantes conforme a profundidade aumenta, independentemente do uso de softmax, quando k está entre 2 e 5

Portanto, esses resultados indicam que, embora a Máscara Binária não supere consideravelmente a acurácia do modelo base, a abordagem oferece vantagens em eficiência computacional. Seus baixos tempos de treinamento em múltiplos valores de k, aliados ao crescimento restrito do número de categorias — sem as perdas de acurácia que ocorrem nas camadas mais profundas do baseline — sugerem que a Máscara Binária oferece o melhor equilíbrio entre desempenho preditivo e custo computacional, quando comparada à arquitetura de soma do vetor r ao canal de entrada.

Conclusão

No presente trabalho, investigamos estratégias para mitigar o custo computacional associado ao uso de modelos Fusion ART multicamadas em tarefas de classificação. Este estudo se mostra necessário porque, como evidenciado na Figura 1, o uso de múltiplas camadas pode aumentar a acurácia de teste, ao mesmo tempo em que eleva o tempo de treinamento e o uso de memória. Motivados por esse compromisso, propusemos e avaliamos duas abordagens para incorporar a informação posicional do

vetor r ao modelo: somar r diretamente à entrada do canal de características e introduzir um canal posicional dedicado com Máscara Binária.

Nenhuma das estratégias conseguiu superar consideravelmente a acurácia do modelo base. No entanto, cada uma delas ofereceu um equilíbrio distinto entre tempo de treinamento, uso de memória e resultados de acurácia. A abordagem de adição de r à entrada foi a mais eficaz em mitigar a degradação de acurácia observada com o aumento da profundidade, mantendo-se entre 0,65 e 0,80 para $k \geq 2$, embora ao custo de um aumento moderado no tempo de treinamento em profundidades intermediárias. Já a estratégia de Máscara Binária ofereceu um equilíbrio geral superior, alcançando contagens de categorias mais baixas e tempos de treinamento praticamente constantes na maioria dos valores de k, sem as perdas acentuadas de acurácia observadas no baseline base em profundidades maiores.

De modo geral, os resultados indicam que nenhuma estratégia se sobressai em todos os critérios avaliados. Portanto, a escolha da abordagem deve ser guiada pelas necessidades e restrições específicas da aplicação-alvo.

Referências

1. ALPAYDIN, E.; KAYNAK, C. Optical Recognition of Handwritten Digits [Dataset]. UCI Machine Learning Repository, 1998.
2. BISHOP, C. M.; BISHOP, H. Deep Learning: Foundations and Concepts. Cham: Springer, 2023.
3. BRITO DA SILVA, L. E.; ELNABARAWY, I.; WUNSCH II, D. C. A survey of adaptive resonance theory neural network models for engineering applications. Neural Networks, v. 120, p. 167–203, 2019.
4. CARPENTER, G. A.; GROSSBERG, S. A massively parallel architecture for a self-organizing neural pattern recognition machine. Computer Vision, Graphics, and Image Processing, v. 37, n. 1, p. 54–115, 1987.
5. CARPENTER, G. A.; GROSSBERG, S.; REYNOLDS, J. H. ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network. Neural Networks, v. 4, n. 5, p. 565–588, 1991.
6. CARPENTER, G. A.; GROSSBERG, S.; ROSEN, D. B. Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system. Neural Networks, v. 4, n. 6, p. 759–771, 1991.
7. GROSSBERG, S. Adaptive pattern classification and universal recoding: I. Parallel development and coding of neural feature detectors. Biological Cybernetics, v. 23, p. 121–134, 1976a.
8. GROSSBERG, S. Adaptive pattern classification and universal recoding: II. Feedback, expectation, olfaction, illusions. Biological Cybernetics, v. 23, p. 187–202, 1976b.
9. HINTON, G.; VINYALS, O.; DEAN, J. Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531, 2015.
10. LECUN, Y. et al. Backpropagation applied to handwritten zip code recognition. Neural Computation, v. 1, n. 4, p. 541–551, 1989.

11. TAN, A.-H.; CARPENTER, G. A.; GROSSBERG, S. Intelligence through interaction: Towards a unified theory for learning. In: Advances in Neural Networks — ISNN 2007. Berlin: Springer, 2007. p. 1094–1103.
12. VASWANI, A. et al. Attention is all you need. In: Advances in Neural Information Processing Systems (NeurIPS 2017), 2017. p. 5998–6008.LEGENDAS DAS FIGURAS:

Anexos

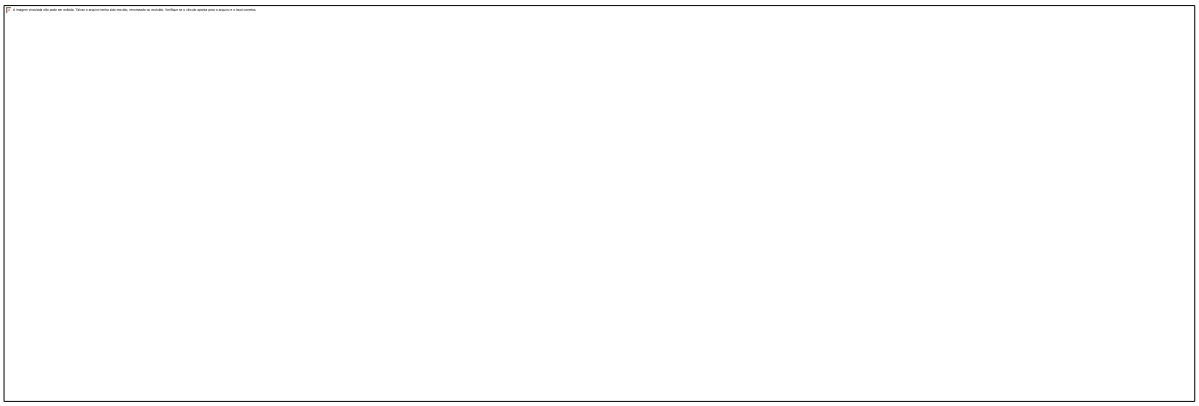


Figura 1. Comparação de desempenho e acurácia entre testes com 1 e 2 camadas, respectivamente, no conjunto de dados Optical Recognition of Handwritten Digits (ALPAYDIN; KAYNAK, 1998), com 5.620 amostras.

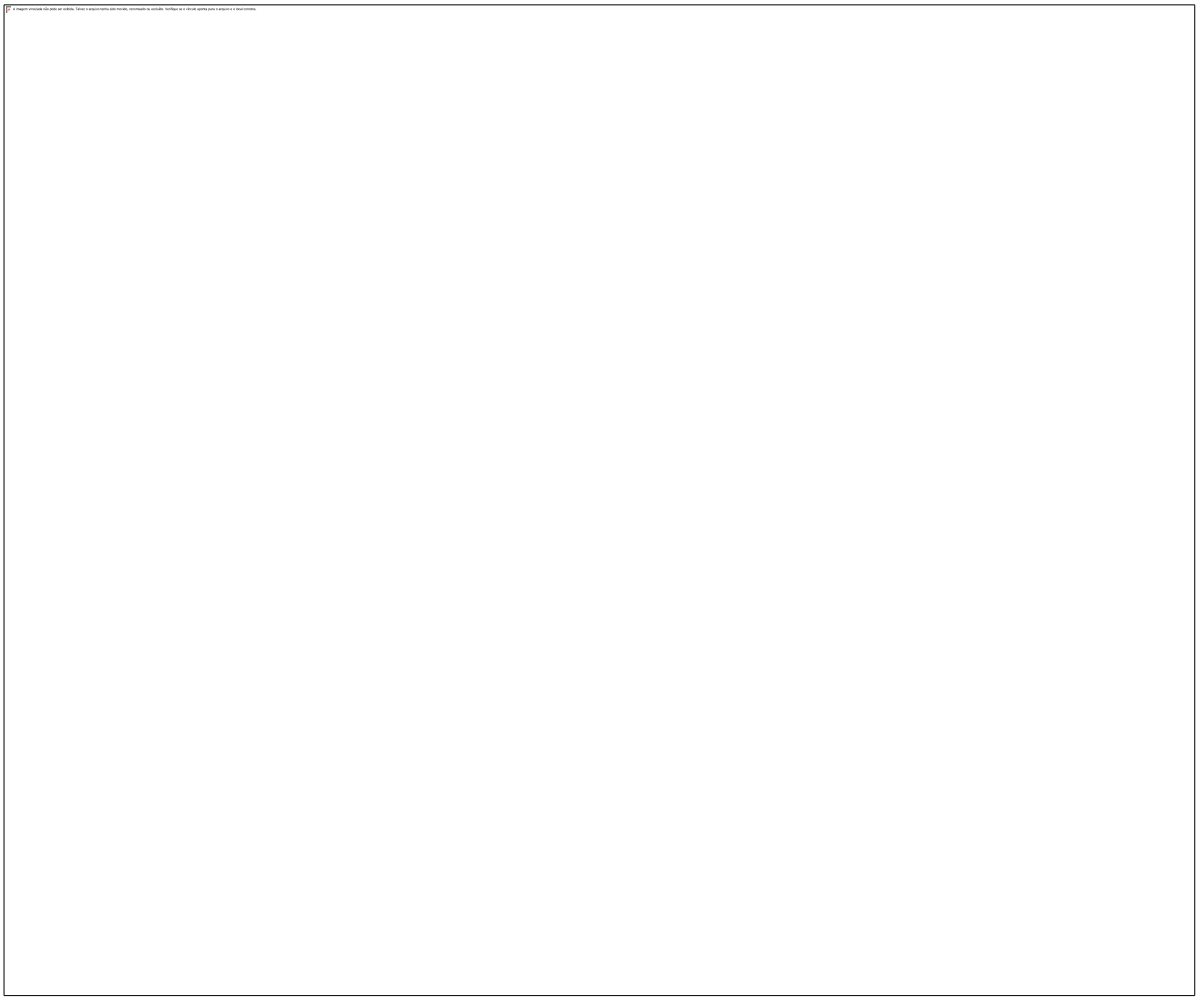


Figura 2. Diagrama da arquitetura que soma a informação de posição diretamente à entrada do canal de características.



Figura 3. Diagrama da arquitetura que possui um canal dedicado à informação de posição.

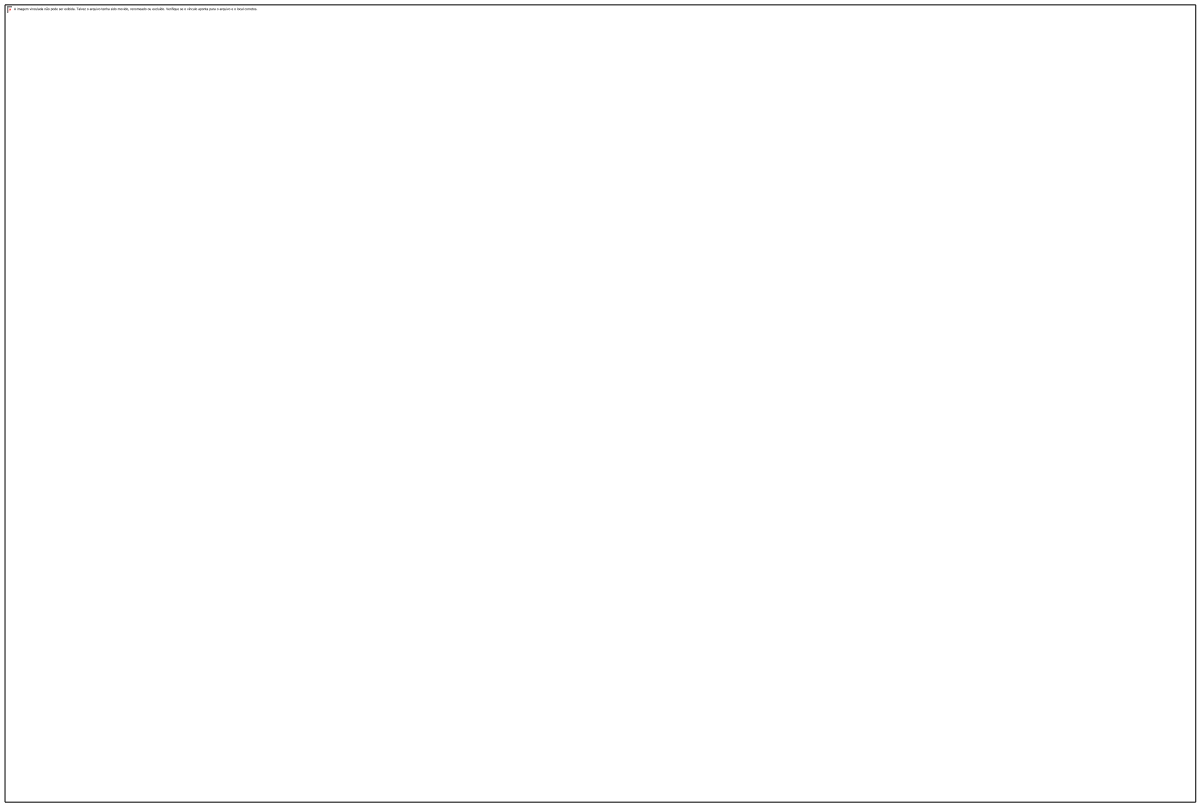


Figura 4. Acurácia em função do número de camadas para múltiplos valores de k nos conjuntos de dados USPS e ORHD. O caso k = 0 corresponde ao uso de todos os valores de ativação.

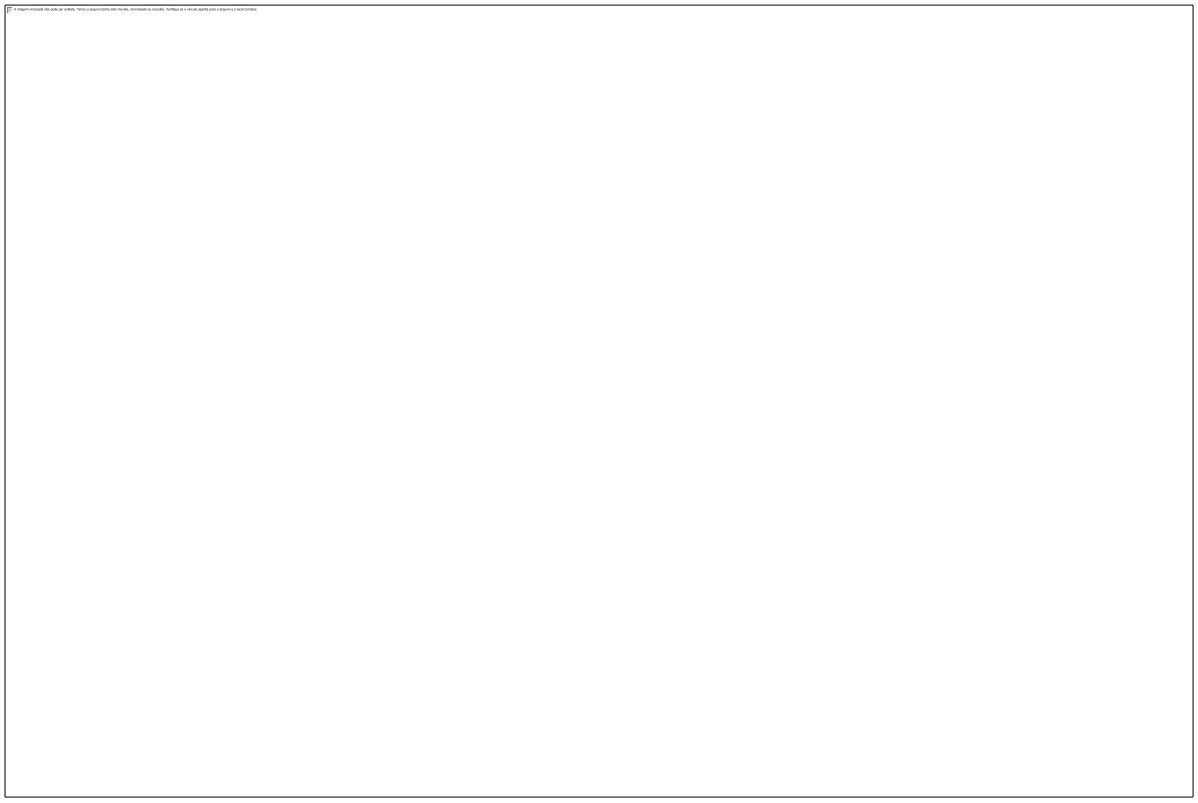


Figura 5. Comparação da acurácia de teste entre diferentes valores de k e diferentes estratégias de treinamento para o conjunto de dados ORHD, quando r é somado à entrada do canal de características.

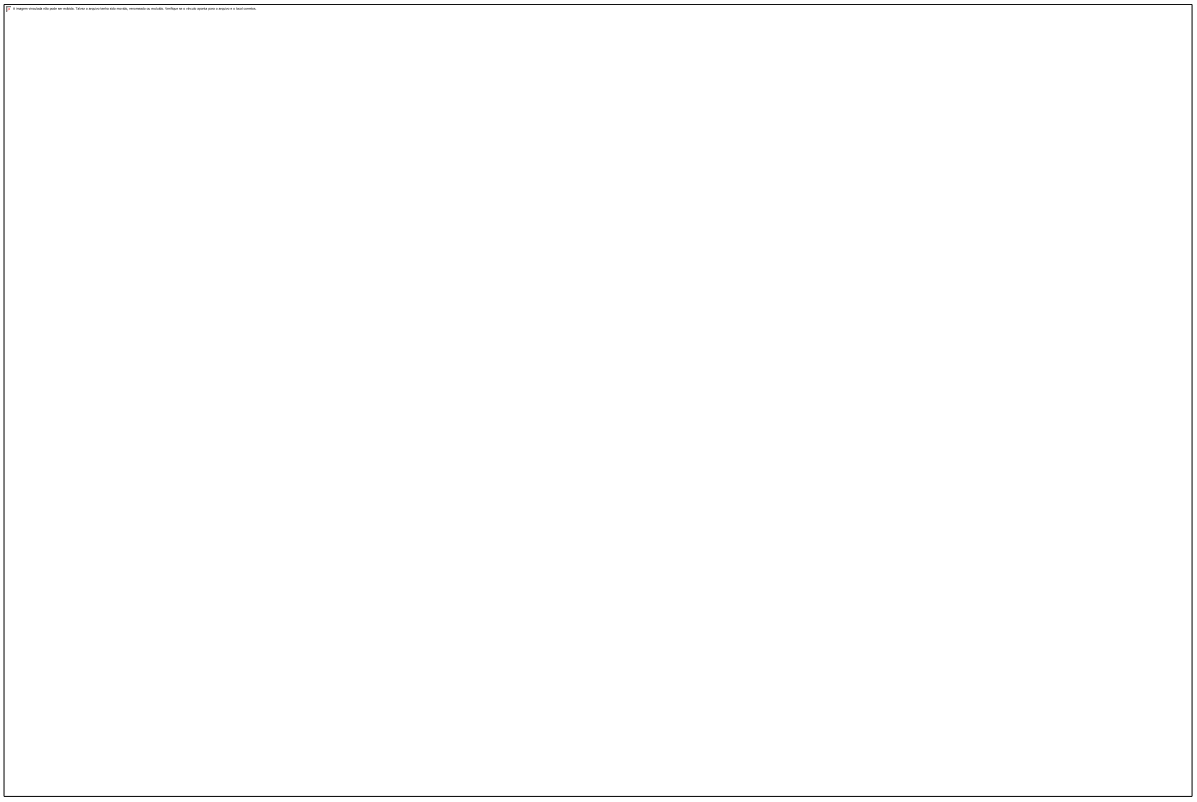


Figura 6. Número de categorias obtido para diferentes valores de k e estratégias de treinamento no conjunto de dados ORHD, quando r é somado à entrada do canal de características.

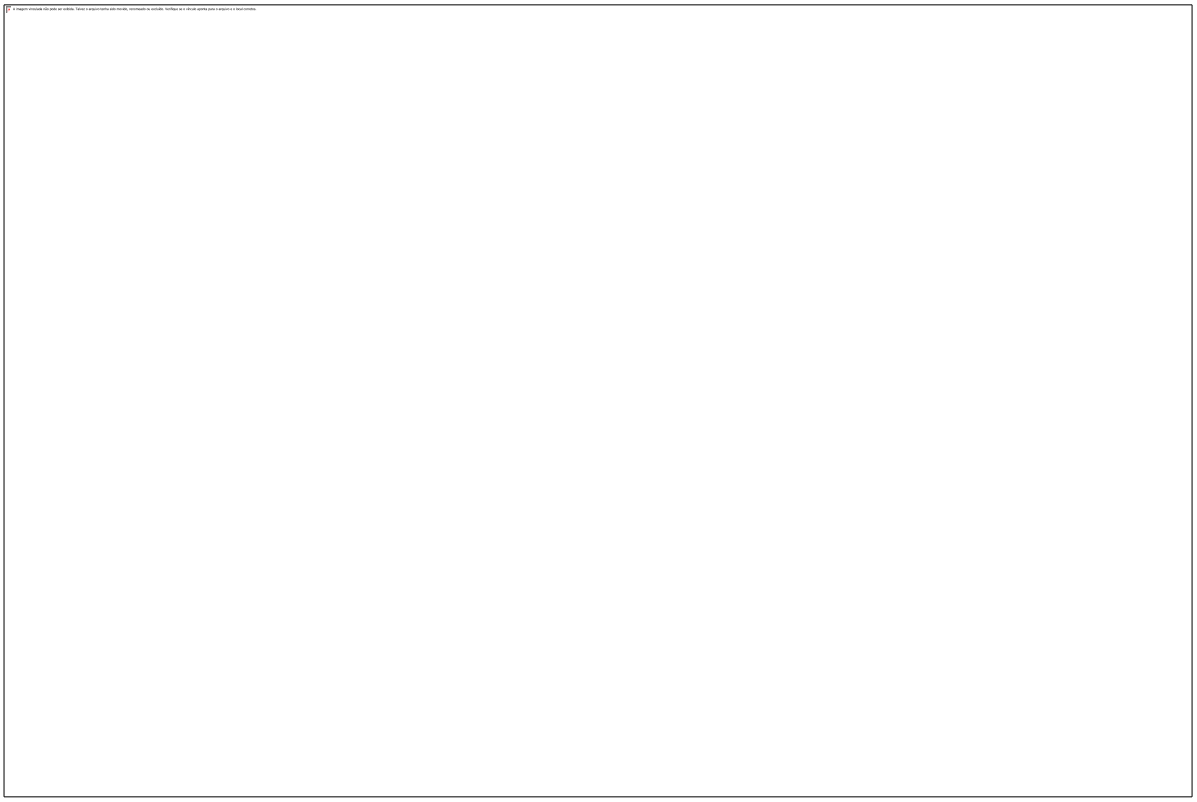


Figura 7. Tempo de treinamento para diferentes valores de k e estratégias de treinamento no conjunto de dados ORHD, quando r é somado à entrada do canal de características.

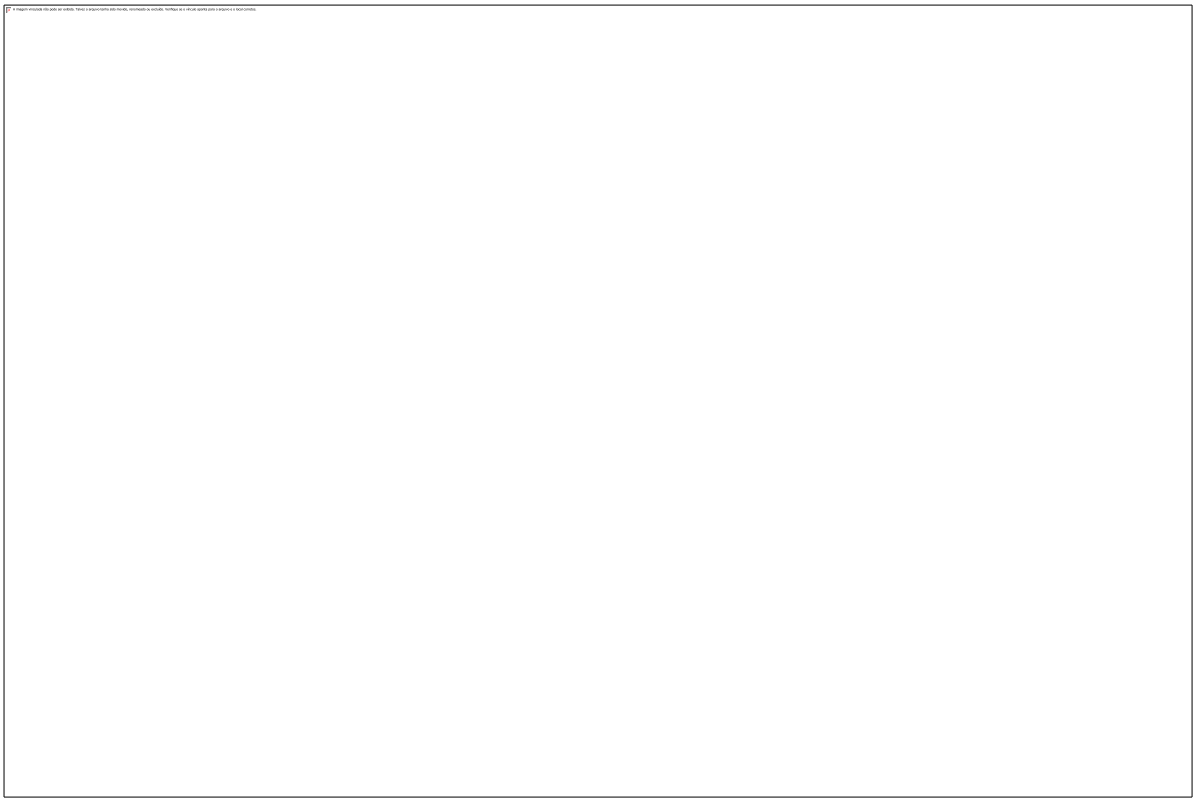


Figura 8. Comparação da acurácia de teste entre diferentes valores de k e diferentes estratégias de treinamento para o conjunto de dados USPS, quando r é somado à entrada do canal de características.

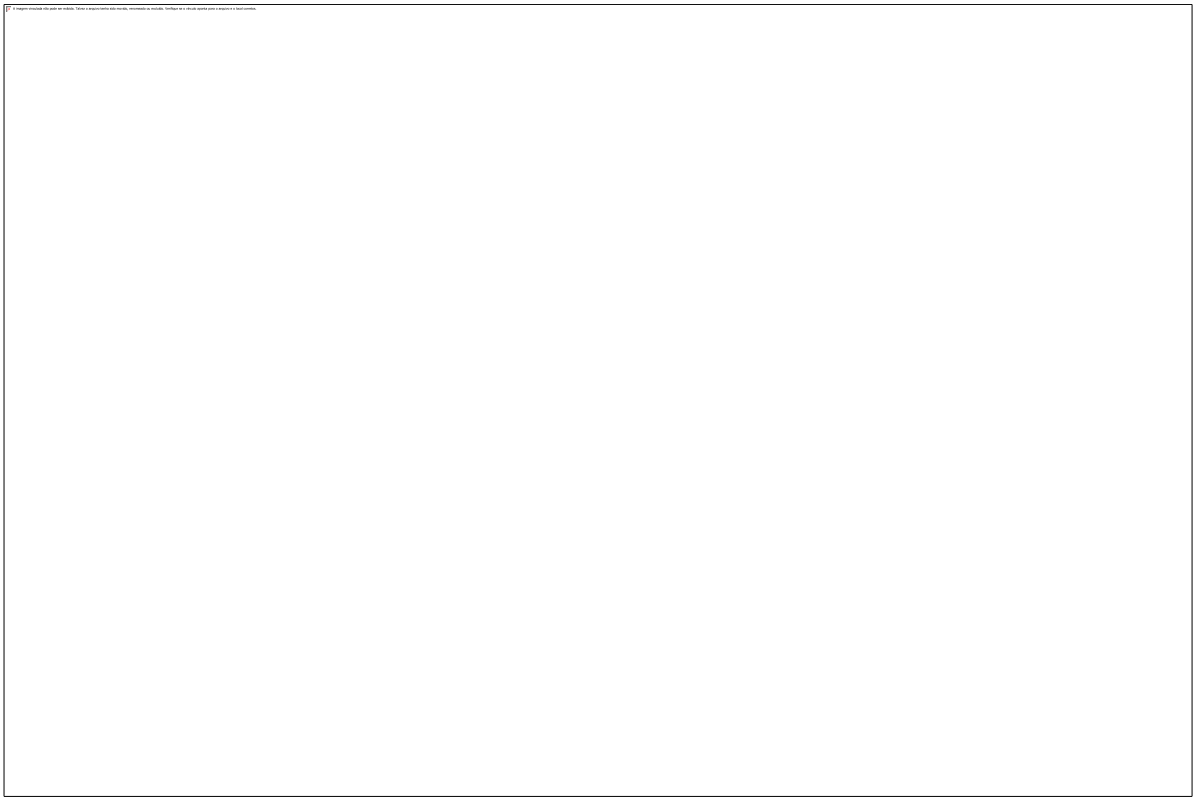


Figura 9. Número de categorias obtido para diferentes valores de k e estratégias de treinamento no conjunto de dados USPS, quando r é somado à entrada do canal de características.

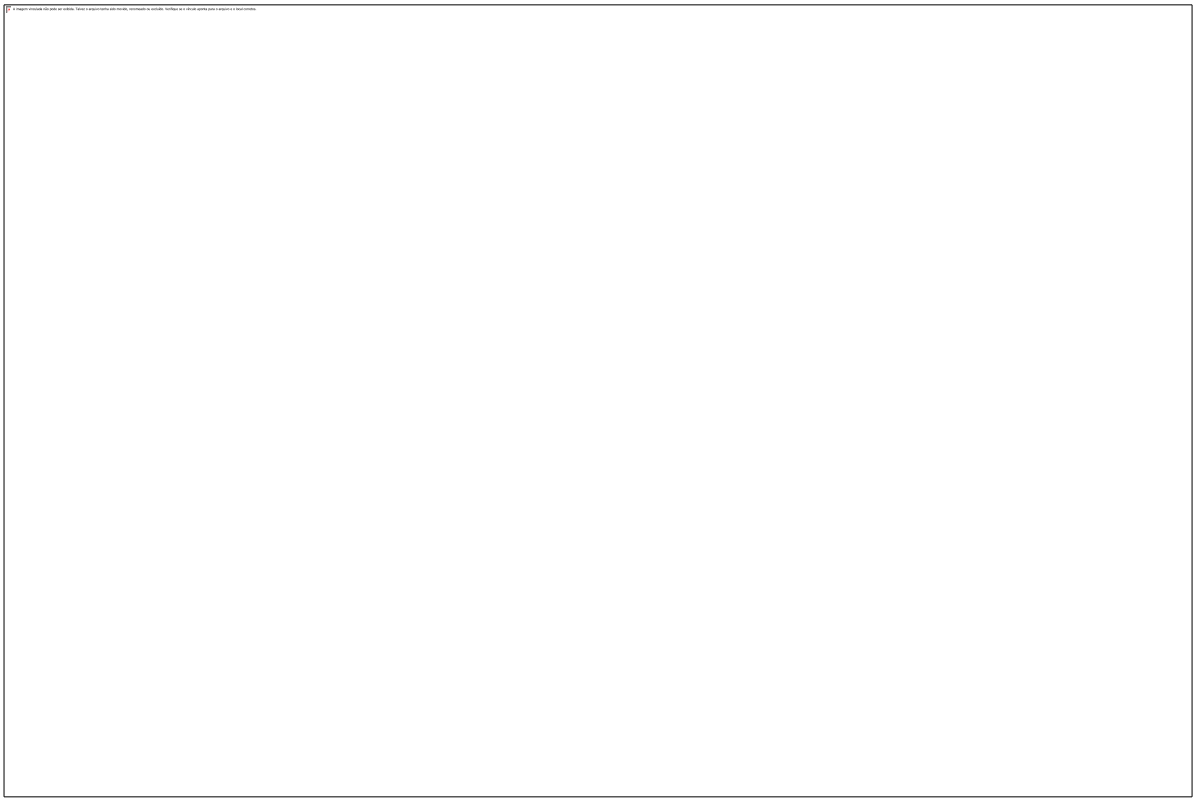


Figura 10. Tempo de treinamento para diferentes valores de k e estratégias de treinamento no conjunto de dados USPS, quando r é somado à entrada do canal de características.

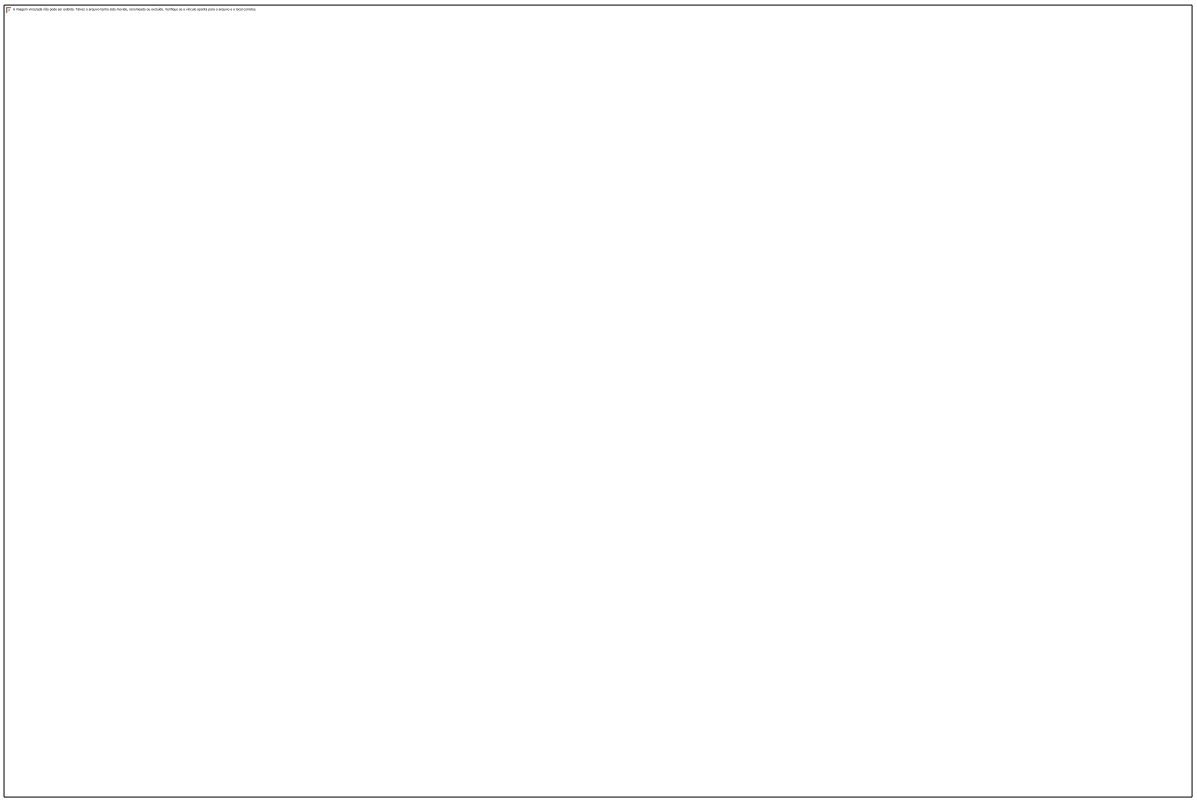


Figura 11. Comparação da acurácia de teste entre diferentes valores de k e diferentes estratégias de treinamento para o conjunto de dados ORHD, utilizando a codificação posicional de Máscara Binária.

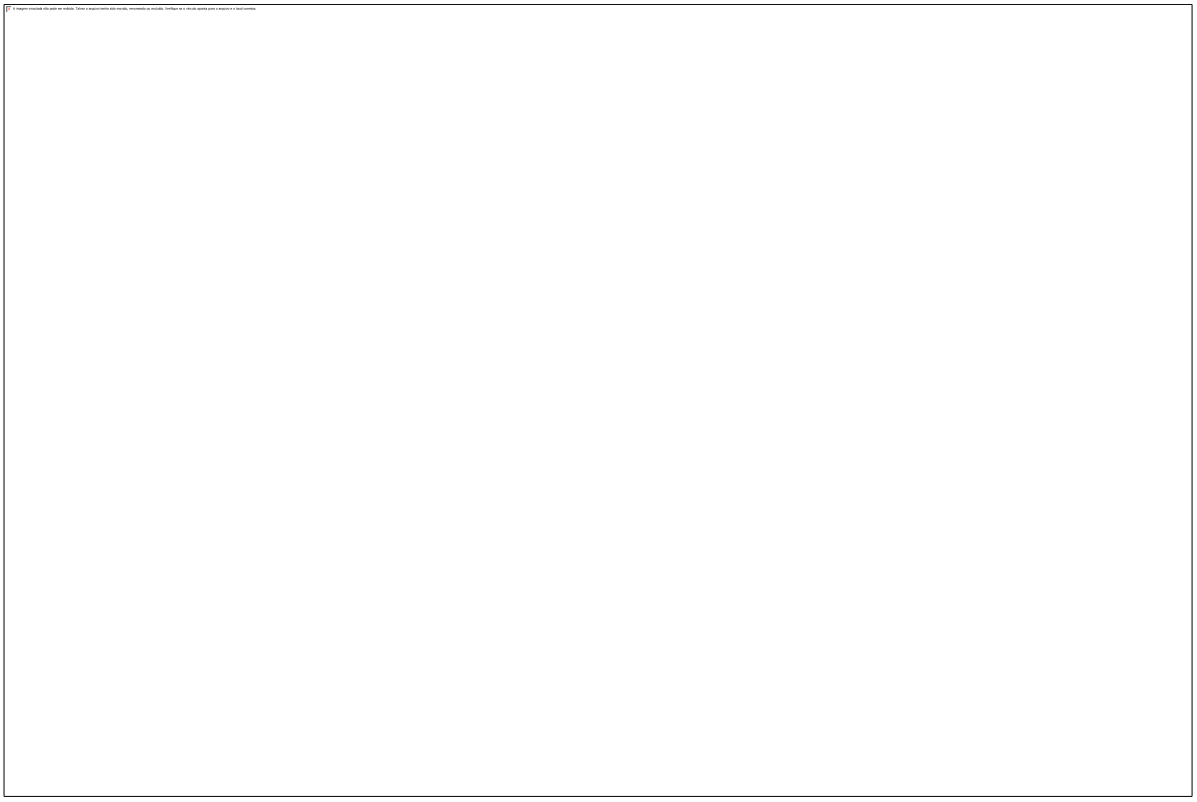


Figura 12. Categorias em função do número de camadas para diferentes valores de k e estratégias de treinamento no conjunto de dados ORHD, utilizando a codificação posicional de Máscara Binária.



Figura 13. Tempo de treinamento em função do número de camadas para diferentes valores de k e estratégias de treinamento no conjunto de dados ORHD, utilizando a codificação posicional de Máscara Binária.

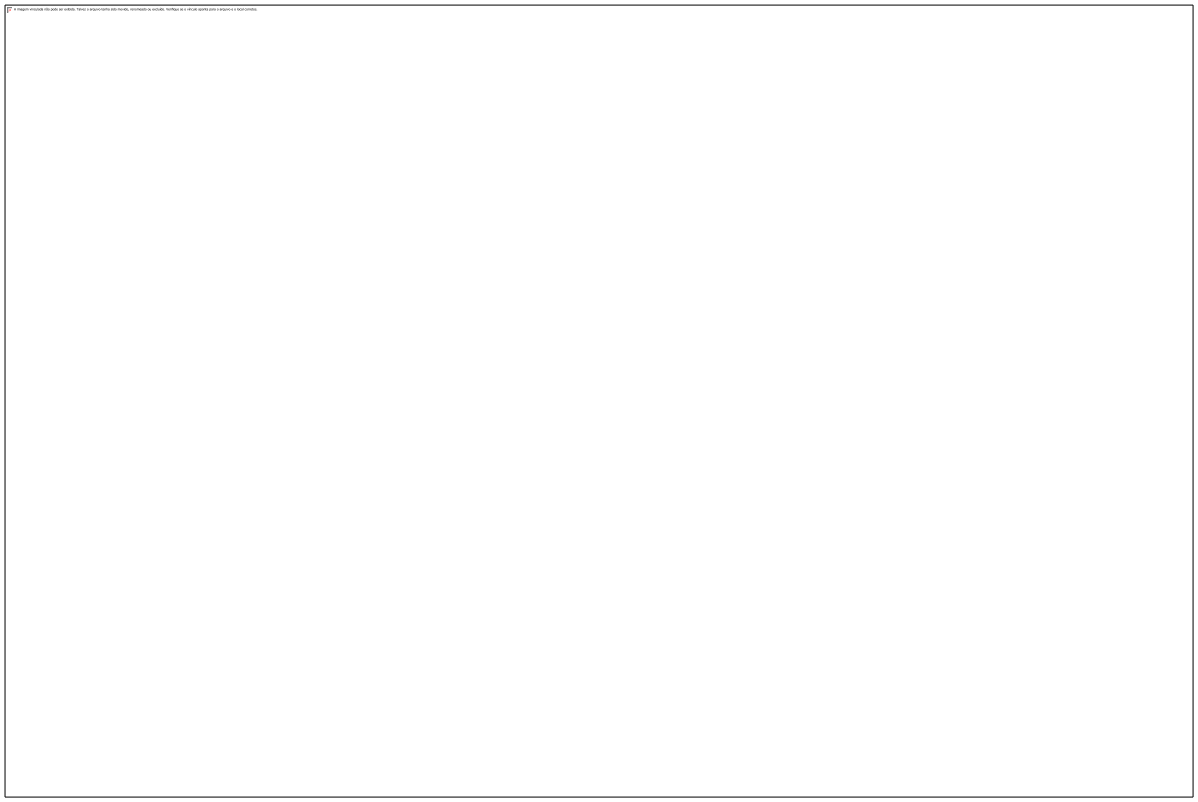


Figura 14. Comparação da acurácia de teste entre diferentes valores de k e diferentes estratégias de treinamento para o conjunto de dados USPS, utilizando a codificação posicional de Máscara Binária.

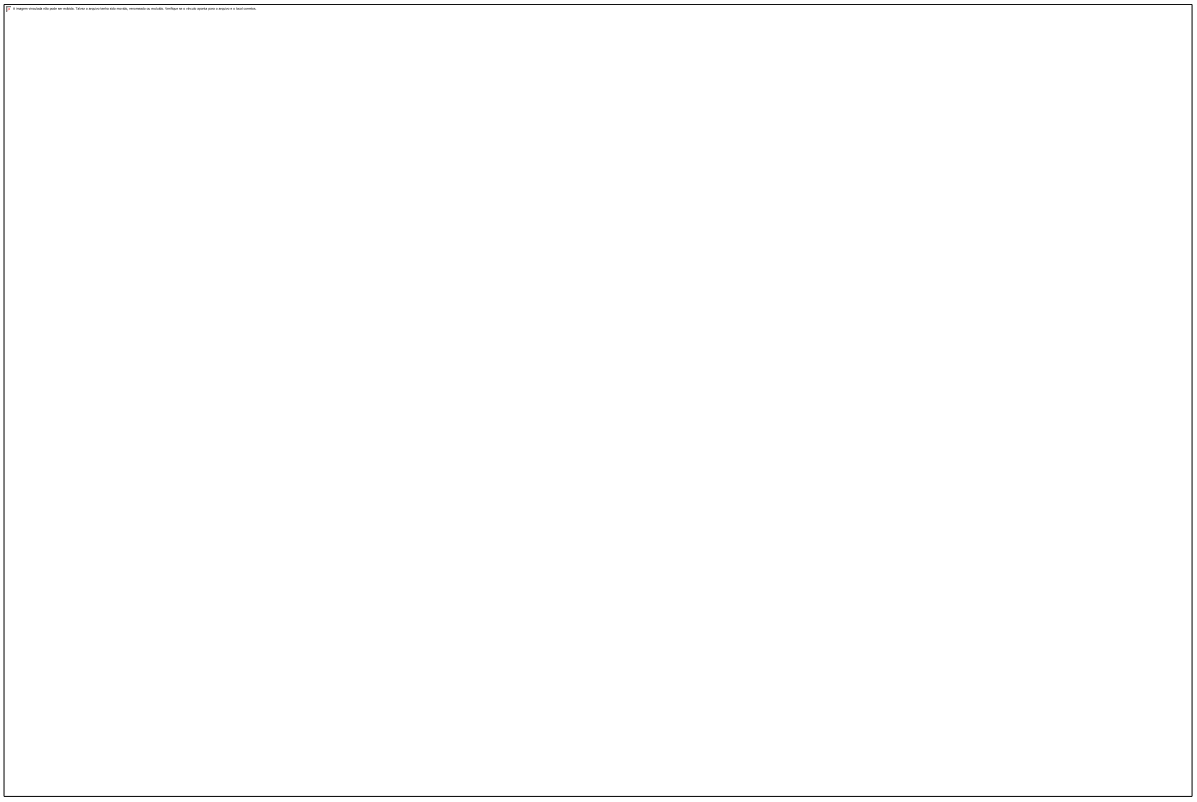


Figura 15. Categorias em função do número de camadas para diferentes valores de k e estratégias de treinamento no conjunto de dados USPS, utilizando a codificação posicional de Máscara Binária.



Figura 16. Tempo de treinamento em função do número de camadas para diferentes valores de k e estratégias de treinamento no conjunto de dados USPS, utilizando a codificação posicional de Máscara Binária.

CÓDIGO: ET1174

AUTOR: ALDO QUEIROZ CABRAL DA FONSECA TINOCO

COAUTOR: MARINA MEDEIROS DE ARAÚJO LEITE

COAUTOR: JOSÉ IVO SCHWADE ARAÚJO

ORIENTADOR: LEONARDO ENZO BRITO DA SILVA

TÍTULO: Um Estudo Sobre Estratégias De Codificação Posicional Para Modelos Fusion ART Multicamadas Em Tarefas De Classificação: Índices de Posição Normalizados e Codificação Senoidal

Resumo

As redes neurais baseadas na Teoria da Ressonância Adaptativa (ART) oferecem uma estrutura estável para o aprendizado incremental e baseado em correspondência match-based, mas seu custo computacional aumenta significativamente quando estendidas para arquiteturas multicamadas. Neste trabalho, investigamos estratégias para reduzir o custo computacional de modelos Fusion ART multicamadas supervisionados, preservando seu desempenho de classificação. Nossa abordagem baseia-se em restringir as entradas de canais de características entre camadas aos k maiores valores de ativação, aproveitando o mecanismo winner-takes-all intrínseco aos modelos ART. Demonstramos que a aplicação ingênua dessa estratégia leva a um overfitting severo e a uma generalização deficiente, especialmente em arquiteturas mais profundas. Em seguida, demonstramos como aplicar a codificação posicional a esse conjunto de dados truncado para solucionar tais problemas de overfitting, explorando uma arquitetura de três canais: um para processar as características (X), um para processar os rótulos (Y) e um para processar o vetor posicional (R). Duas estratégias de processamento diferentes foram aplicadas à arquitetura com canal adicional no que diz respeito ao vetor posicional: normalização, pré-processamento por função senoidal.

Palavras-chave: Ressonância Adaptativa.Aprendizado Multicamadas.Codificação Posicional

TITLE: A Study on Positional Encoding Strategies for Multilayer Fusion ART Models in Classification Tasks: Normalized Position Indices and Sinusoidal Encodings

Abstract

Adaptative Resonance Theory (ART) neural networks provide a stable framework for incremental and match-based learning, but their computational cost increases significantly when extended to multi-layer architectures. In this work, we investigate strategies to reduce the computational cost of supervised multi-layer Fusion ART models while preserving their classification performance. Our approach is based on restricting inter-layer feature-channel inputs to the top-k activation values, leveraging the winner-takes-all mechanism intrinsic to ART models. We demonstrate that naively applying this strategy leads to severe overfitting and poor generalization, especially in deeper architectures. We then demonstrate how we can apply positional encoding to this truncated dataset to solve those overfitting problems, exploring a three-channel architecture: on channel processing features (X), one channel processing labels (Y) and the last channel processing the positional vector (R). Two different processing strategies were applied to the extra-channel architecture, with regard to the positional vector: normalization, sinusoidal function preprocessing.

Keywords: Adaptive Resonance Theory.Multilayer Learning.Positional Encoding

Introdução

A Teoria da Ressonância Adaptativa (ART) é uma abordagem sobre como o cérebro humano percebe informações novas e já vivenciadas **(GROSSBERG, 1976a, 1976b)**, descrevendo o aprendizado de novos padrões como um processo baseado em match. Essa teoria culminou na idealização e no desenvolvimento de uma série de modelos de redes neurais **(BRITO DA SILVA; ELNABARAWY; WUNSCH II, 2019)**, como, por exemplo, ART **(GROSSBERG, 1976a)** ,ARTMAP e Fuzzy ART **(CARPENTER; GROSSBERG, 1987; CARPENTER; GROSSBERG; REYNOLDS, 1991; CARPENTER; GROSSBERG; ROSEN, 1991)**. Um dos modelos da família de redes ART é a Fusion ART **(TAN; CARPENTER; GROSSBERG, 2007)**, que utiliza uma arquitetura multicanal para processar diferentes tipos de dados e trocar informações entre eles, visando capturar relações sutis entre fontes de dados distintas. A entrada de um modelo Fusion ART pode ser definida por Ick = (Ick1, Ick2, Ick3...Ickn), onde Icki ∈ [0, 1] denota a entrada i para o canal ck. É importante notar que a entrada não precisa estar presente em todos os canais; para os canais que não recebem entrada, os vetores de entrada são inicializados com valores iguais a 1. Os canais de um modelo Fusion ART podem ser compostos por qualquer modelo ART não supervisionado, e um modelo Fusion ART supervisionado pode ser obtido utilizando-se um canal para processar características e outro para processar rótulos, de forma semelhante à arquitetura ARTMAP. No entanto, esses modelos Fusion ART supervisionados podem ser organizados sequencialmente em camadas para formar um modelo de N camadas, onde a saída das funções de ativação do canal de características do modelo na camada n, para cada amostra, é fornecida como entrada para o canal de características do modelo na camada n + 1, mantendo-se os mesmos rótulos. Essa abordagem proporciona um aumento na acurácia, embora esse ganho venha acompanhado de um aumento no custo computacional (Figura 1).

Metodologia

O alto custo computacional decorre do fato de que o tamanho do vetor de valores de ativação é diretamente proporcional ao número total de nós do modelo, crescendo a cada camada, já que esses valores são calculados para cada vetor de pesos e para cada amostra.

No entanto, como o modelo Fusion ART seleciona a categoria mais provável aplicando uma estratégia winner-takes-all, ao reduzirmos as ativações na entrada do canal de características apenas aos k valores mais altos preservamos o mesmo vencedor com uma ocupação de memória muito menor.

Contudo, como essa abordagem extrai apenas os k maiores valores independentemente de qual categoria é a vencedora, e cada valor corresponde a uma característica da amostra, ela acaba trocando incorretamente características entre posições do conjunto de dados truncado. A ordem de apresentação das características passa a variar a cada camada e, com isso, perde-se a correspondência entre a posição que um valor ocupa no vetor truncado e a característica que o originou. Codificação Posicional. Recuperar essa correspondência exige transmitir ao modelo, junto de cada valor de ativação, a posição que ele ocupava antes do truncamento. Para isso,aplicamos uma estratégia de codificação posicional **(VASWANI et al., 2017)** utilizando o vetor r, que contém a informação posicional de cada valor de ativação. Observamos os resultados de duas abordagens distintas para representar esses dados posicionais:

1. **Índices Normalizados.** Para um valor de ativação arbitrário Tn em um vetor de ativação T de tamanho N, o elemento correspondente rn do vetor de índices r é definido por rn = n / (N − 1).
2. **Função Senoidal.** A codificação posicional senoidal, originalmente proposta por Vaswani et al. (2017) para Transformers, codificava a posição de um token em uma sequência; neste trabalho, ela foi adaptada para codificar o índice das categorias mais ativadas de uma camada anterior da rede Fusion ART:

rnj = sin(n / L^(2i/D)), se j = 2i

rnj = cos(n / L^(2i/D)), se j = 2i + 1

onde L e D são hiperparâmetros. Os valores resultantes, originalmente no intervalo [-1,1], são normalizados para [0,1] e, em seguida, submetidos à codificação por complemento, mantendo-os compatíveis com a codificação já utilizada pelo Fuzzy ART.

Arquitetura.

Aproveitando o aspecto de múltiplas fontes de dados dos modelos Fusion ART, criamos um canal dedicado à informação posicional, separando x e r em dois canais diferentes (Figura 2). Ao lidar com os Índices Normalizados, simplesmente passamos ao canal de Codificação Posicional uma versão pré-processada dos índices dos elementos do vetor T.

Em seguida, aplicamos entre cada camada do modelo a função softmax com um parâmetro de temperatura T **(HINTON; VINYALS; DEAN, 2015)**, para melhor ponderar a importância dos valores de ativação mais altos:

$\sigma(z)_i = \exp(z_i/T) / \sum_{j=1}^{K} \exp(z_j/T)$

Como utilizamos um Fuzzy ART em cada canal do Fusion ART supervisionado, aplicamos as técnicas de pré-processamento exigidas pelos modelos ART baseados em lógica fuzzy: normalização das características para o intervalo [0,1] e codificação por complemento das amostras, aumentando a entrada Ick com um vetor complementar tal que $Ick_i = 1 - Ick_i$.

CONFIGURAÇÃO EXPERIMENTAL

Hardware e Reprodutibilidade

Todos os experimentos foram conduzidos exclusivamente na infraestrutura de computação de alto desempenho do Núcleo de Processamento de Alto Desempenho (NPAD) da Universidade Federal do Rio Grande do Norte (UFRN). A partição utilizada apresenta sete nós computacionais com 2x **AMD Genoa 9654 DP/UP 96C/192T, 3 TB de RAM DDR5, 8x NVMe Micron 7450 PRO 3.8TB em RAID0 e Infiniband Connect X6 100Gb/s.**

Conjuntos de Dados

A avaliação utilizou dois conjuntos de dados de imagens de dígitos manuscritos, amplamente empregados em tarefas de aprendizado de máquina: o ORHD **(ALPAYDIN; KAYNAK, 1998)**, com 5.620 instâncias descritas por 64 características, e o USPS **(LECUN et al., 1989)**, com maior variabilidade, totalizando 9.298 amostras e 256 características por imagem. Ambos foram pré-processados conforme os requisitos dos modelos adotados, garantindo compatibilidade e comparação justa entre os experimentos.

Parâmetros e Configuração

Para nosso Fusion ART supervisionado, utilizamos um modelo Fuzzy ART em cada canal, com os seguintes parâmetros:

Canal de Rótulos: Gamma (γ) = 0,0 | Rho (ρ) = 1,0 | Alpha (α) = 10^{-3} | Beta (β) = 1,0
Canais de Características e de Codificação Posicional: Gamma (γ) = 0,5 | Rho (ρ) = 0,0 | Alpha (α) = 10^{-3} | Beta (β) = 1,0

Realizamos o experimento variando k de k = 1 a k = 5, além de testar utilizando todos os valores de ativação, e utilizando 6 camadas em cada iteração.

Baseline

Como dito anteriormente, o modelo Fusion ART seleciona a categoria mais provável para uma amostra aplicando uma estratégia winner-takes-all. Assim, se reduzirmos as ativações na entrada do canal de características apenas aos k valores mais altos, preservamos o mesmo vencedor com uma ocupação de memória muito menor.

No entanto, isso leva a um problema: como o modelo multicamadas extrai apenas os k valores de ativação mais altos independentemente da posição, especialmente ao lidar com dados fortemente dependentes da ordem de apresentação de suas características, podemos acabar com valores muito semelhantes para amostras completamente diferentes, confundindo consequentemente o modelo.

O impacto dessa abordagem ingênua — de simplesmente restringir os valores de entrada aos k valores de ativação mais altos — em diferentes profundidades de rede, para os conjuntos de dados USPS e ORHD, é ilustrado na Figura 3 Os resultados dos testes mostram uma queda acentuada na acurácia à medida que o número de camadas aumenta.

Quando todos os valores são utilizados (k = 0), o modelo alcança as acurácias de teste mais altas e mais estáveis para todos os números de camadas. Em ambos os conjuntos de dados, arquiteturas mais profundas inicialmente melhoram o desempenho, atingindo um pico com duas camadas, após o qual a acurácia de teste degrada gradualmente. Esse comportamento sugere que, embora o empilhamento de camadas Fusion ART aumente o poder representacional, a profundidade excessiva introduz uma superespecialização.

Em contrapartida, em todas as configurações truncadas, isto é, com k entre 1 e 5, a acurácia de treinamento permanece alta, frequentemente ultrapassando 95% mesmo em modelos mais profundos. Contudo, a acurácia de teste sofre um colapso acentuado após a primeira camada. Essa divergência entre treino e teste indica um claro caso de overfitting.

Esse fenômeno pode ser atribuído à perda de informação estrutural causada pela remoção das ativações. Embora o mecanismo winner-takes-all garanta que a categoria mais ativada seja preservada, descartar as ativações de menor ranking remove pistas contextuais necessárias para um bom desempenho.

De modo geral, esses resultados mostram que a seleção ingênua de top-k ativações compromete severamente a generalização.

Para cada uma das abordagens específicas de tratamento dos dados posicionais no novo canal, realizamos uma comparação considerando duas condições: o uso do softmax entre camadas e o uso da codificação posicional. Assim, obtivemos quatro configurações para cada uma das duas abordagens. Além disso, analisamos o cenário utilizando todos os valores de ativação, para 5 valores diferentes de k, e para 4 camadas do Fusion ART. As abordagens foram comparadas levando em consideração a presença e a ausência de codificação posicional.

Abordagem de Normalização:

Observamos que a maior acurácia foi alcançada pela segunda camada do modelo que utiliza todos os valores de ativação, sem codificação posicional e sem a aplicação da função softmax. Este modelo será doravante referido como Modelo base.

As Figuras 4 e 7 mostram que, considerando tanto o conjunto de dados ORHD quanto o USPS, nenhuma das abordagens conseguiu superar a acurácia do modelo base, exceto quando todos os valores de ativação foram utilizados na segunda camada do conjunto de dados ORHD. Nesse caso, ambas as abordagens superaram o modelo base. No entanto, quando o modelo multicamadas está restrito às top-k ativações, observamos que a presença da codificação posicional permite uma melhor preservação da sua performance; de modo inverso, quando ela está ausente percebe-se uma maior degradação da acurácia.

Também é notável que, para ambas as abordagens, à medida que o número de valores de ativação utilizados aumenta no intervalo de 1 a 5, a acurácia tende a se estabilizar entre 0,6 e 0,8.

As Figuras 6 e 9 mostram o mesmo padrão de tempo de treinamento para os conjuntos de dados USPS e ORHD. Em praticamente todas as configurações, o baseline é a estratégia mais custosa, escalando de forma acentuada com a profundidade e atingindo os maiores tempos de execução em $k \geq 3$, exceto em $k = 1$, onde ela colapsa para a configuração mais barata, de aproximadamente 1 segundo por camada, uma vez removido o softmax. A abordagem de normalização apresenta consistentemente um custo intermediário a alto, crescendo de forma constante com o número de camadas e convergindo para valores semelhantes com ou sem o uso de softmax.

Portanto, esses resultados indicam que, embora o uso da normalização não supere consideravelmente a acurácia do modelo base, a abordagem oferece vantagens em eficiência computacional.

Abordagem Senoidal:

Para essa abordagem, seguindo o mesmo protocolo das seções anteriores, três configurações foram testadas, variando a dimensão de embedding d e a constante L: $d = 8$ com $L = 100$; $d = 16$ com $L = 100$; e $d = 16$ com $L = 1000$.

Em relação aos resultados de acurácia (Figuras 10 e 13), nenhuma configuração de modelo com a estratégia senoidal conseguiu superar o modelo base em nenhum dos dois conjuntos de dados. A melhor acurácia obtida no conjunto de dados ORHD foi 0,84, alcançada com $k = 4$, duas camadas, $d = 16$, $L = 100$, e com a função softmax aplicada. No conjunto de dados USPS, a melhor acurácia obtida foi 0,81, alcançada com $k = 2$, quatro camadas, $d = 16$, $L = 100$, e com a função softmax aplicada. Contudo, a estratégia de codificação senoidal exibe uma estabilidade considerável na acurácia ao longo da profundidade quando comparada ao modelo sem codificação posicional e restrito às top-k ativações.

A influência da função softmax é mais fraca para essa abordagem do que para as anteriormente discutidas. Nas três configurações, seu efeito é, no geral, pouco benéfico, mostrando mudanças notáveis na acurácia apenas para valores menores de k, entre 1 e 3. Isso reproduz, em escala menor, o padrão relatado para a estratégia de normalização: o softmax só proporciona uma sustentação na acurácia quando poucos valores de ativação são utilizados.

Os resultados indicam que os modelos são robustos à escolha dos hiperparâmetros D e L. Em média, considerando todas as camadas e ambas as configurações de softmax, as três combinações diferem por menos de um ponto percentual de acurácia em ambos os conjuntos de dados, e o ranking entre elas não é constante entre os conjuntos de dados. Portanto, não há evidências de que a codificação senoidal precise de um ajuste fino minucioso.

Essa abordagem apresenta um tempo de treinamento consideravelmente mais rápido em comparação com o baseline (Figuras 12 e 15). Esse ganho decorre diretamente de um número menor de categorias em comparação com a abordagem anterior, mostrando uma diferença crescente em camadas mais profundas, chegando a quase um quarto do número de categorias por volta das camadas 5 e 6.

Em termos de memória, a estratégia senoidal de codificação posicional é, na maioria dos casos, mais custosa do que o baseline sem codificação posicional. Como o canal posicional cresce tanto com k quanto com o tamanho do embedding, o modelo geralmente acaba com mais parâmetros do que o baseline sob o mesmo número de valores de ativação, mesmo utilizando menos categorias, como mostram as Figuras 11 e 14. Também observamos que o tamanho do embedding afeta a memória, mas afeta a acurácia em escala bem menor.

Conclusão

No presente trabalho, investigamos estratégias para mitigar o custo computacional associado ao uso de modelos Fusion ART multicamadas em tarefas de classificação. Este estudo se mostra necessário porque, como evidenciado na Figura 1, o uso de múltiplas camadas pode aumentar a acurácia de teste, ao mesmo tempo em que eleva o tempo de treinamento e o uso de memória. Motivados por esse compromisso, propusemos uma nova arquitetura para o modelo: introduzir um canal posicional dedicado com normalização ou codificação senoidal.

Nenhuma das duas estratégias conseguiu superar consideravelmente a acurácia do modelo base. No entanto, cada uma delas ofereceu um equilíbrio distinto entre tempo de treinamento, uso de memória e resultados de acurácia. A abordagem de Normalização apresentou um comportamento de acurácia mais elevado, porém com um custo computacional menos favorável, permanecendo em um nível intermediário ao longo das profundidades testadas. A codificação Senoidal, por sua vez, alcançou os tempos de treinamento mais rápidos e a acurácia mais estável ao longo da profundidade entre todas

as abordagens, ao custo de um maior uso de memória, já que seu canal posicional escala tanto com k quanto com a dimensão do embedding.

De modo geral, os resultados indicam que nenhuma estratégia se sobressai em todos os critérios avaliados. Portanto, a escolha da abordagem deve ser guiada pelas necessidades e restrições específicas da aplicação-alvo.

Referências

1. ALPAYDIN, E.; KAYNAK, C. Optical Recognition of Handwritten Digits [Dataset]. UCI Machine Learning Repository, 1998.
2. BRITO DA SILVA, L. E.; ELNABARAWY, I.; WUNSCH II, D. C. A survey of adaptive resonance theory neural network models for engineering applications. Neural Networks, v. 120, p. 167–203, 2019.
3. CARPENTER, G. A.; GROSSBERG, S. A massively parallel architecture for a self-organizing neural pattern recognition machine. Computer Vision, Graphics, and Image Processing, v. 37, n. 1, p. 54–115, 1987.
4. CARPENTER, G. A.; GROSSBERG, S.; REYNOLDS, J. H. ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network. Neural Networks, v. 4, n. 5, p. 565–588, 1991.
5. CARPENTER, G. A.; GROSSBERG, S.; ROSEN, D. B. Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system. Neural Networks, v. 4, n. 6, p. 759–771, 1991.
6. GROSSBERG, S. Adaptive pattern classification and universal recoding: I. Parallel development and coding of neural feature detectors. Biological Cybernetics, v. 23, p. 121–134, 1976a.
7. GROSSBERG, S. Adaptive pattern classification and universal recoding: II. Feedback, expectation, olfaction, illusions. Biological Cybernetics, v. 23, p. 187–202, 1976b.
8. HINTON, G.; VINYALS, O.; DEAN, J. Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531, 2015.
9. LECUN, Y. et al. Backpropagation applied to handwritten zip code recognition. Neural Computation, v. 1, n. 4, p. 541–551, 1989.
10. TAN, A.-H.; CARPENTER, G. A.; GROSSBERG, S. Intelligence through interaction: Towards a unified theory for learning. In: Advances in Neural Networks — ISNN 2007. Berlin: Springer, 2007. p. 1094–1103.
11. VASWANI, A. et al. Attention is all you need. In: Advances in Neural Information Processing Systems (NeurIPS 2017), 2017. p. 5998–6008.

Anexos

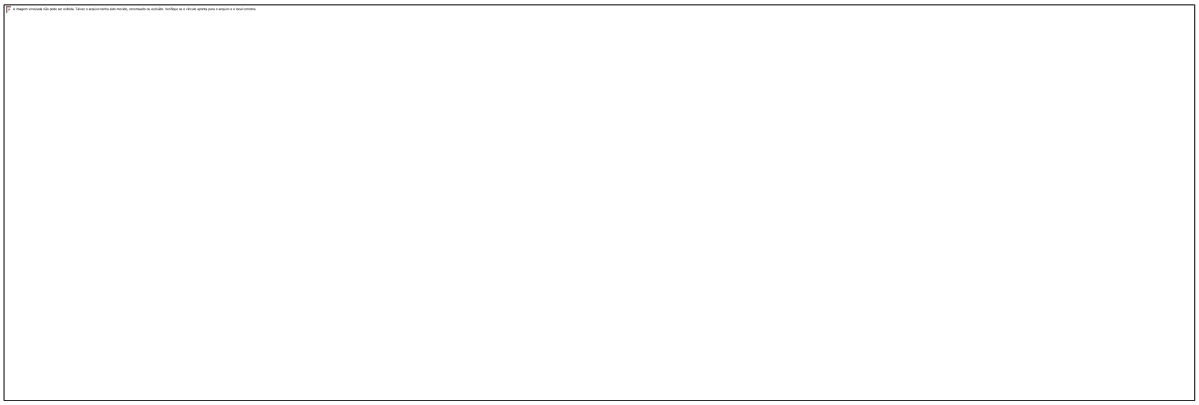


Figura 1. Comparação de desempenho e acurácia entre testes com 1 e 2 camadas, respectivamente, no conjunto de dados Optical Recognition of Handwritten Digits (ALPAYDIN; KAYNAK, 1998), com 5.620 amostras.

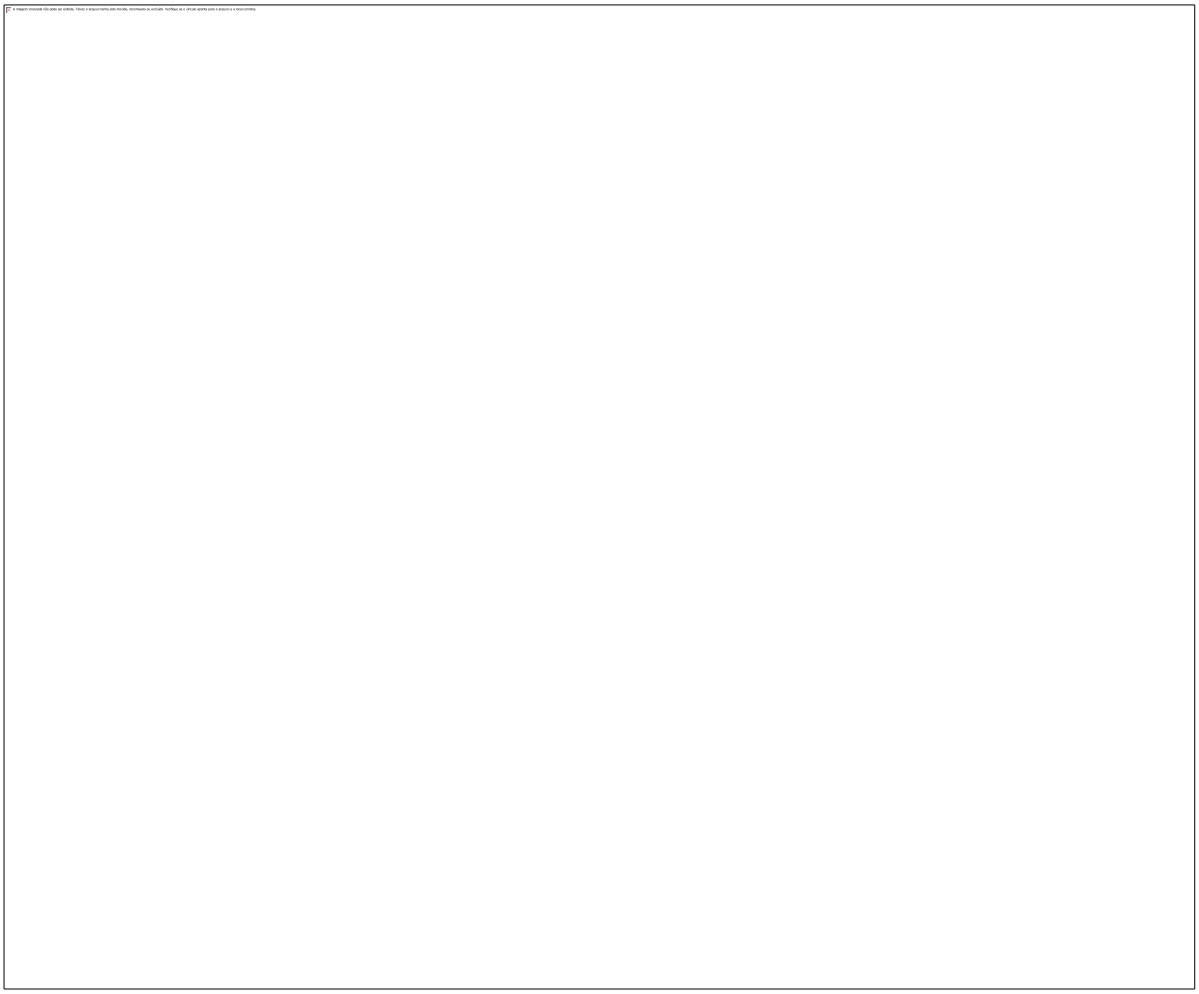


Figura 2 Diagrama da arquitetura que possui um canal dedicado à informação de posição.

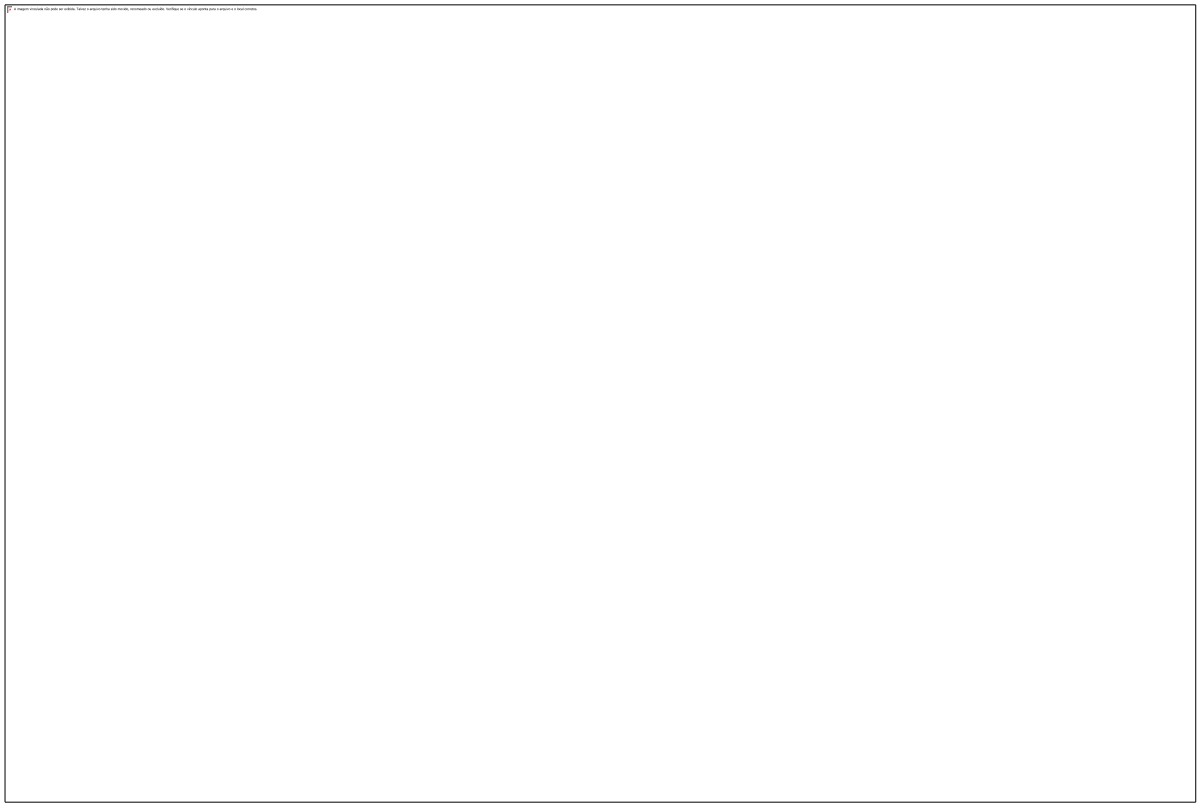


Figura 3. Acurácia em função do número de camadas para múltiplos valores de k nos conjuntos de dados USPS e ORHD. O caso k = 0 corresponde ao uso de todos os valores de ativação.

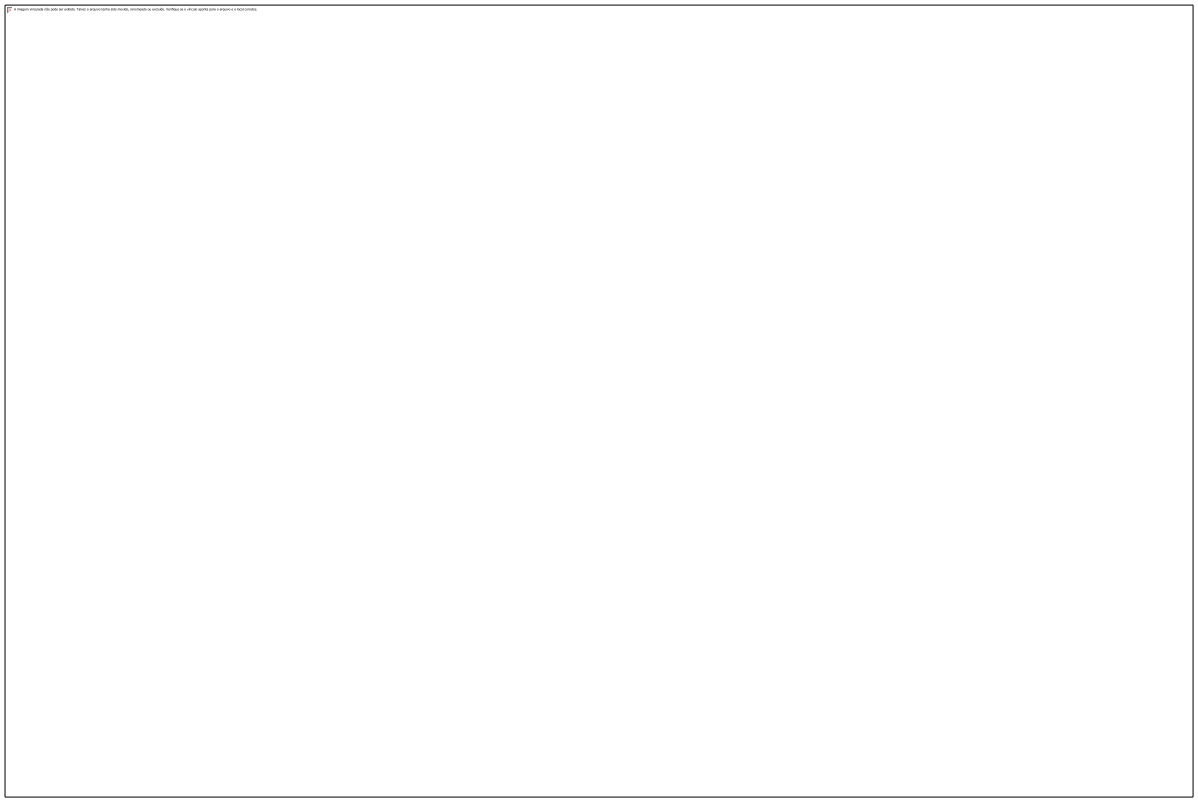


Figura 4. Comparação da acurácia de teste entre diferentes valores de k e diferentes estratégias de treinamento para o conjunto de dados ORHD, para a condição posicional de Índices Normalizados

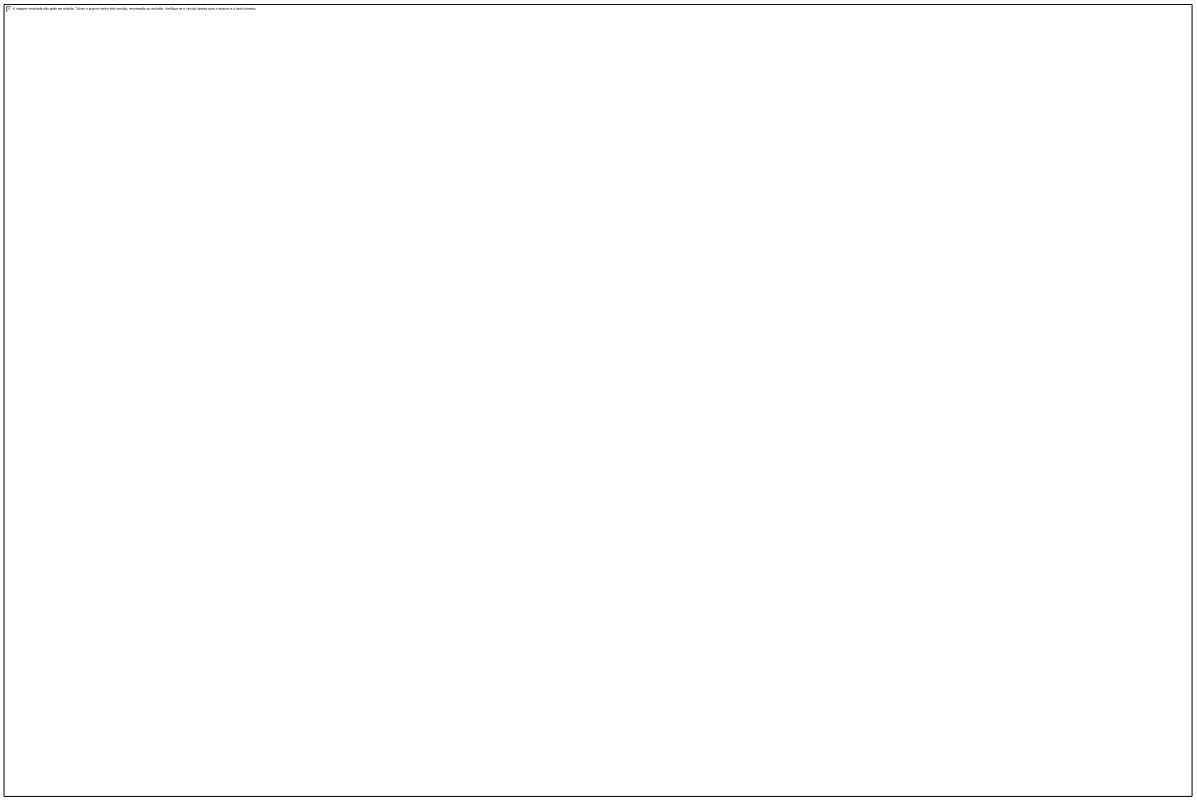


Figura 5. Categorias em função do número de camadas para diferentes valores de k e estratégias de treinamento no conjunto de dados ORHD, para a condição posicional de Índices Normalizados



Figura 6. Tempo de treinamento em função do número de camadas para diferentes valores de k e estratégias de treinamento no conjunto de dados ORHD, para a condição posicional de Índices Normalizados

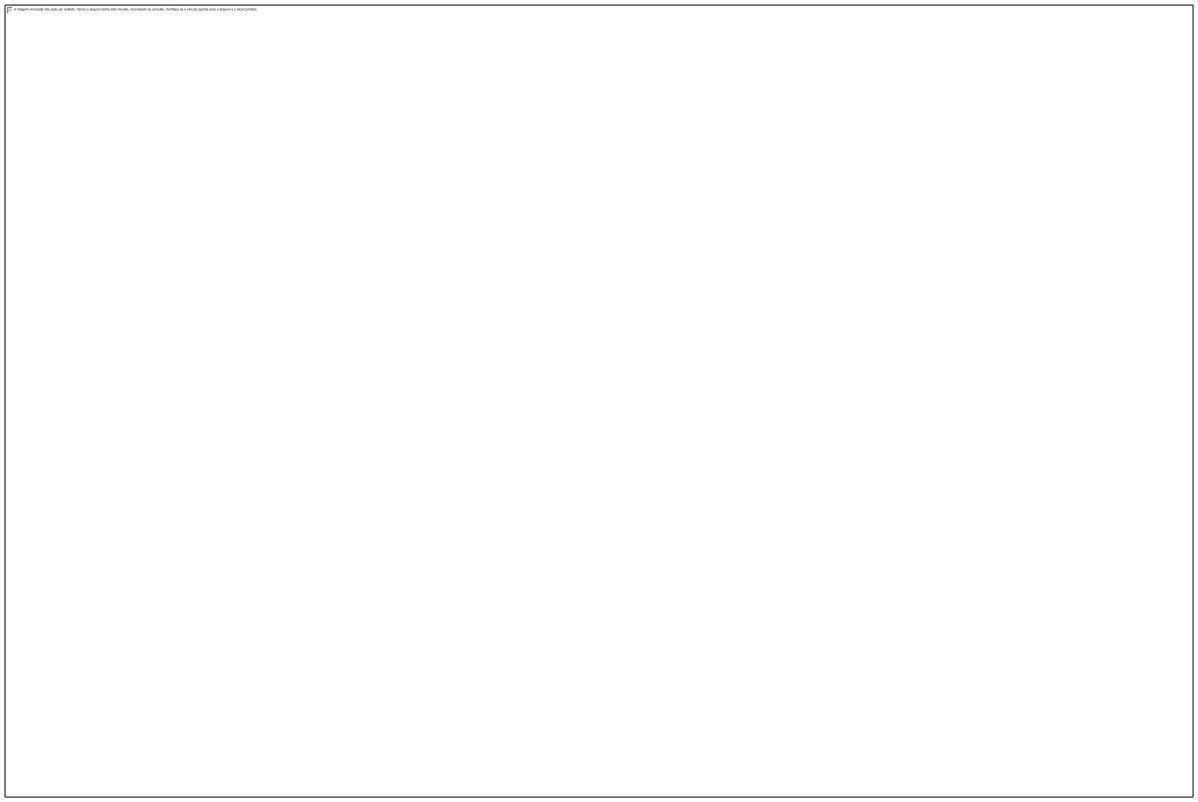


Figura 7. Comparação da acurácia de teste entre diferentes valores de k e diferentes estratégias de treinamento para o conjunto de dados USPS, para a condição posicional de Índices Normalizados

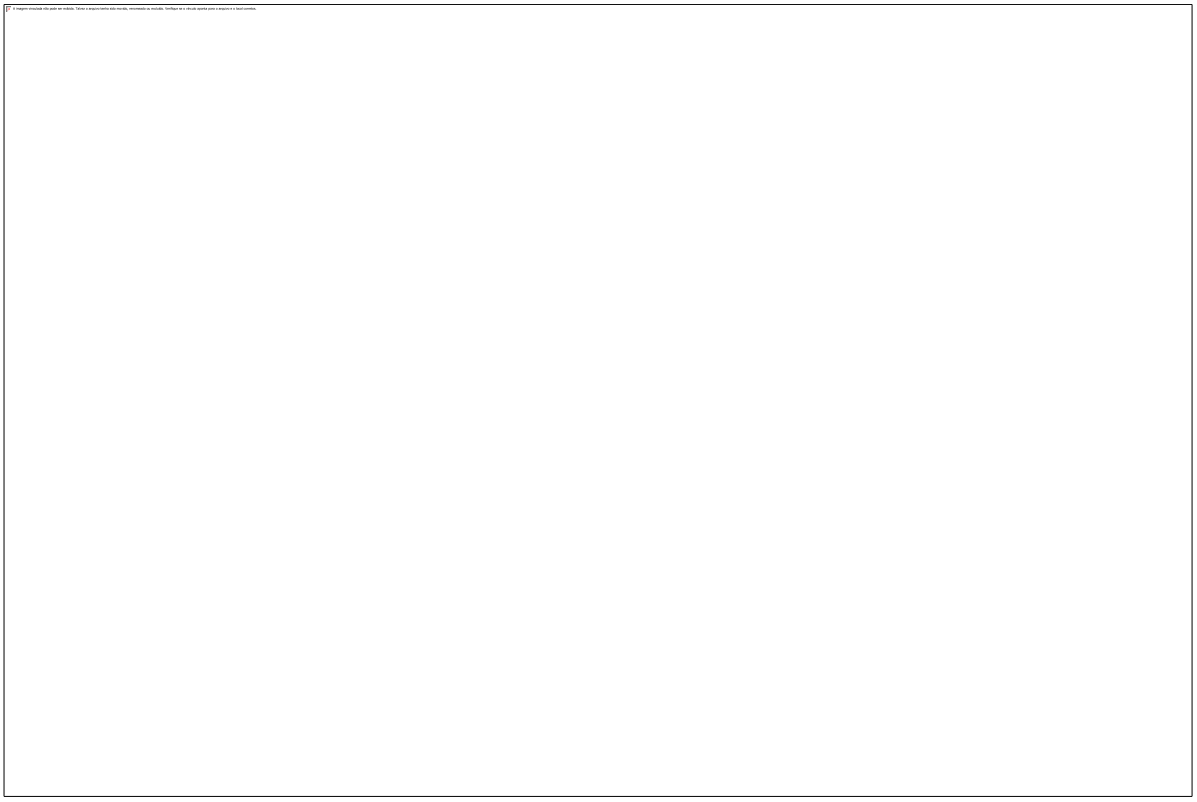


Figura 8. Categorias em função do número de camadas para diferentes valores de k e estratégias de treinamento no conjunto de dados USPS, para a condição posicional de Índices Normalizados



Figura 9. Tempo de treinamento em função do número de camadas para diferentes valores de k e estratégias de treinamento no conjunto de dados USPS, para a condição posicional de Índices Normalizados

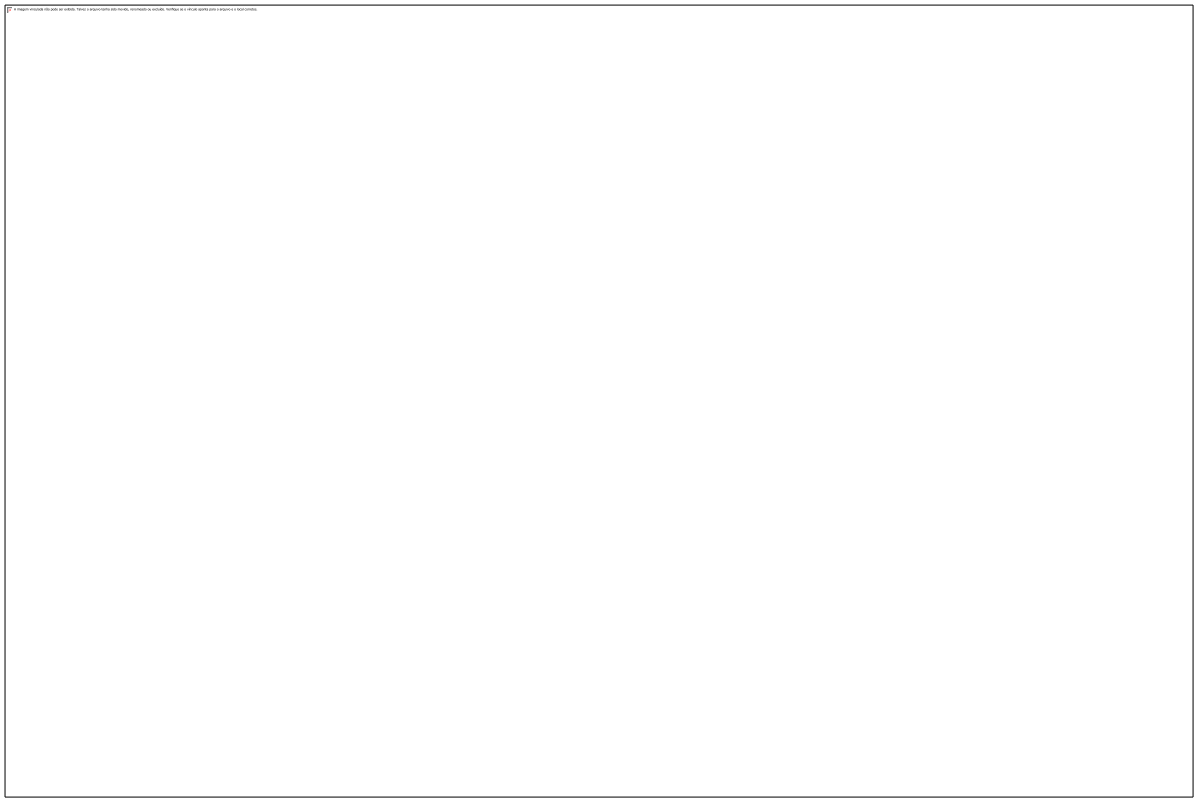


Figura 10. Comparação da acurácia de teste entre diferentes valores de k e diferentes estratégias de treinamento para o conjunto de dados ORHD, utilizando codificação posicional senoidal.

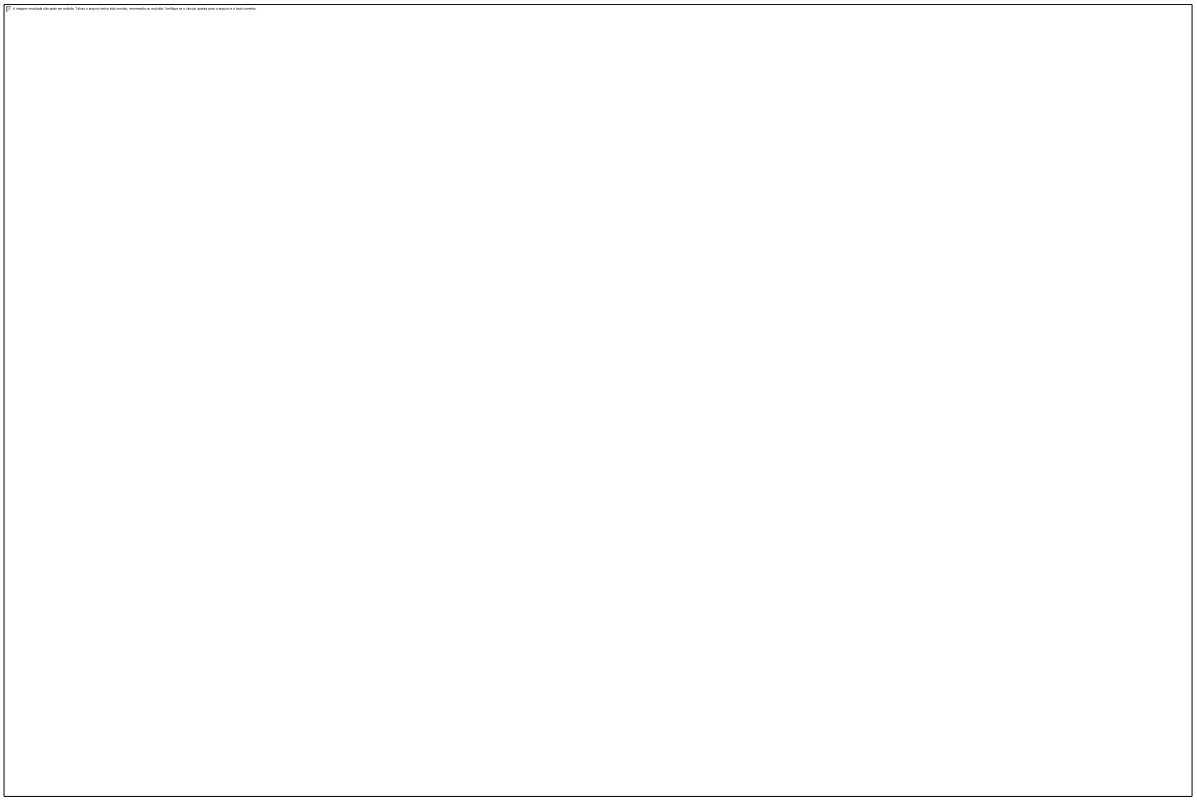


Figura 11. Categorias em função do número de camadas para diferentes valores de k e estratégias de treinamento no conjunto de dados ORHD, utilizando codificação posicional senoidal.



Figura 12. Tempo de treinamento em função do número de camadas para diferentes valores de k e estratégias de treinamento no conjunto de dados ORHD, utilizando codificação posicional senoidal.

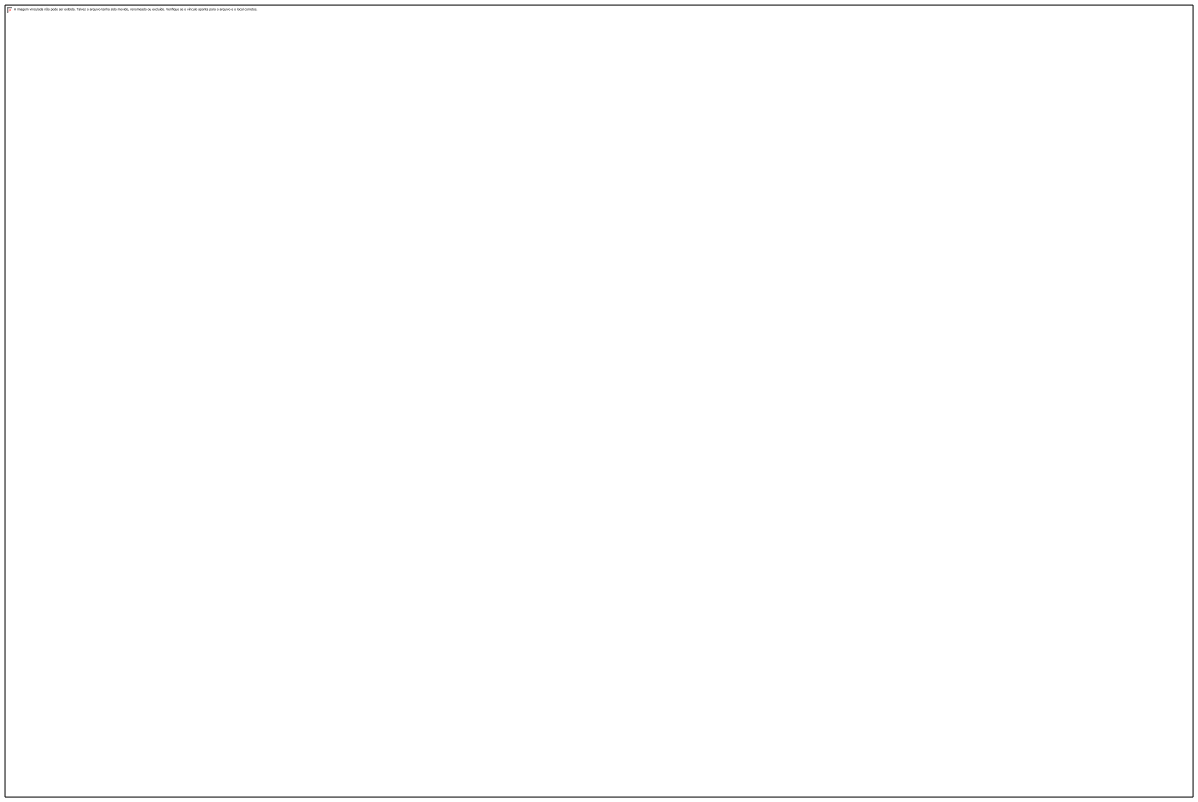


Figura 13. Comparação da acurácia de teste entre diferentes valores de k e diferentes estratégias de treinamento para o conjunto de dados USPS, utilizando codificação posicional senoidal.

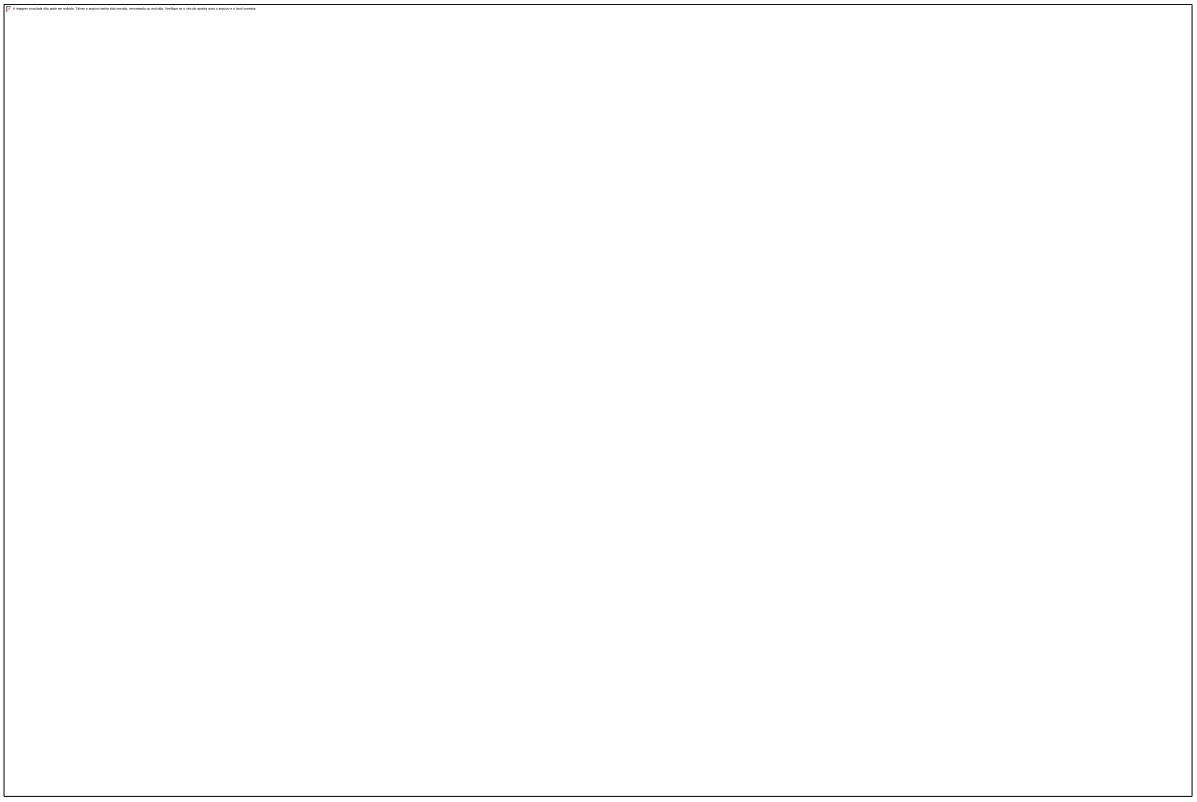


Figura 14. Categorias em função do número de camadas para diferentes valores de k e estratégias de treinamento no conjunto de dados USPS, utilizando codificação posicional senoidal.



Figura 15. Tempo de treinamento em função do número de camadas para diferentes valores de k e estratégias de treinamento no conjunto de dados USPS, utilizando codificação posicional senoidal.

CÓDIGO: ET1220

AUTOR: SAULO CARLOS PEREIRA DA ROCHA

ORIENTADOR: ROGER KREUTZ IMMICH

TÍTULO: Análise Comparativa de Desempenho entre os Padrões ERC-721 e ERC-1155 para Digitalização de Dispositivos IoT em Blockchain.

Resumo

A crescente adoção de dispositivos na Internet das Coisas (IoT) amplia a necessidade de mecanismos seguros e eficientes para sua representação e gerenciamento em arquiteturas descentralizadas. Embora a tokenização desses ativos em blockchain constitua uma alternativa para o estabelecimento de identidades digitais, os custos associados às operações de registro podem limitar sua aplicação em cenários de maior escala. Nesse contexto, este trabalho apresenta uma análise comparativa dos padrões ERC-721 e ERC-1155 para a digitalização de dispositivos IoT, avaliando o consumo de gas e os impactos sobre os recursos computacionais do nó blockchain. Foram implementados contratos inteligentes funcionalmente equivalentes em Solidity, os quais foram submetidos a cargas progressivas de 5 a 5.000 registros utilizando Hardhat e K6. Os resultados demonstraram um ponto de equilíbrio em aproximadamente 22 dispositivos, a partir do qual o ERC-1155 apresentou menor consumo acumulado de gas, alcançando uma vantagem de cerca de 6,6% na maior carga avaliada, além de menor consumo médio por token. Em contrapartida, não foi observada vantagem persistente entre os padrões quanto ao uso de CPU e memória RAM. Com isso, observou-se que o ERC-1155 apresenta maior eficiência em termos de gas com o aumento da escala de digitalização, mostrando-se uma alternativa promissora para a digitalização de dispositivos IoT em arquiteturas descentralizadas de registro e autenticação baseadas em blockchain.

Palavras-chave: Internet das Coisas, Blockchain, Tokenização, NFTs, Dispositivos

TITLE: Comparative Performance Analysis of ERC-721 and ERC-1155 Standards for Digitizing IoT Devices on Blockchain.

Abstract

The growing adoption of Internet of Things (IoT) devices increases the need for secure and efficient mechanisms for their representation and management within decentralized architectures. Although tokenizing these assets on a blockchain offers a way to establish digital identities, the costs associated with registration operations can limit their application in large-scale scenarios. In this context, this study presents a comparative analysis of the ERC-721 and ERC-1155 standards for digitizing IoT devices, evaluating gas consumption and the impact on blockchain node computational resources. Functionally equivalent smart contracts were implemented in Solidity and subjected to progressive loads ranging from 5 to 5,000 registrations using Hardhat and K6. The results revealed a break-even point at approximately 22 devices; beyond this point, ERC-1155 demonstrated lower cumulative gas consumption—achieving an advantage of about 6.6% at the highest load tested—as well as lower average consumption per token. Conversely, no consistent advantage was observed between the standards regarding CPU and RAM usage. Thus, ERC-1155 proved to be more gas-efficient as the scale of digitization increased, making it a promising alternative for digitizing IoT devices within blockchain-based decentralized registration and authentication architectures.

Keywords: Internet of Things, Blockchain, Tokenization, NFTs, Devices

Introdução

A expansão da Internet das Coisas (IoT) redefiniu a infraestrutura digital ao integrar dispositivos heterogêneos em ecossistemas críticos como saúde, indústria e cidades inteligentes [Majd e Sharko 2023, Garba et al. 2023]. No entanto, a dependência tradicional de arquiteturas centralizadas para o gerenciamento de identidade e autenticação dos dispositivos expõe esses sistemas a pontos únicos de falha, além de graves vulnerabilidades relacionadas à privacidade e à integridade dos dados [Hameed et al. 2021, Qiu et al. 2024]. Para superar a fragilidade desses mecanismos convencionais e prevenir operações não autorizadas [Liou e Lin 2021, Khashan e Khafajah 2023], a tecnologia blockchain surge como uma alternativa promissora ao prover descentralização, imutabilidade e rastreabilidade por meio de criptografia e consenso distribuído [Goswami e Choudhury 2022, Liu et al. 2024].

Nesse âmbito, a tokenização de dispositivos IoT a partir de contratos inteligentes (*smart contracts*) tem sido explorada para estabelecer identidades digitais singulares e apoiar mecanismos descentralizados de autenticação. Os Non-Fungible Tokens (NFTs) permitem associar ativos físicos a representações digitais registradas na blockchain, assegurando unicidade, rastreabilidade e integridade. Majd e Sharko (2023), por exemplo, vinculam dispositivos IoT a NFTs a partir de identificadores derivados de *Physical Unclonable Functions* (PUFs). Já Qiu et al. (2024) utilizam NFTs para associar individualmente dispositivos médicos vestíveis (IoMT) às identidades de seus usuários. De maneira semelhante, Khalil et al. (2024) empregam NFTs como representações digitais que incorporam atributos específicos para autenticação de nodes

em cidades inteligentes, destacando, em sua obra, o potencial da digitalização de ativos como mecanismo de credenciamento.

Sob a perspectiva de custo operacional, o padrão de *token* adotado desponta como um dos fatores determinantes no desempenho das arquiteturas. Nas soluções para IoT, o ERC-721 é o protocolo comum utilizado para a digitalização desses ativos. Majd e Sharko (2023), entretanto, evidenciam limitações operacionais ao reportarem a baixa eficiência econômica do protocolo na criação massiva de NFTs e exploram uma otimização como alternativa para mitigar o consumo de *gas*. A relevância desse custo operacional também é observada por Khalil et al. (2024), que adotam o *gas* gasto no *minting* como métrica central de eficiência para a arquitetura IoT proposta. Diante dessas restrições, Qiu et al. (2024) associam o ERC-721 aos gargalos no registro de múltiplos dispositivos na solução e sugerem a adoção do padrão *multi-token* ERC-1155 como uma alternativa promissora. Tais evidências reforçam a necessidade de avaliar o impacto da escolha do padrão de contrato inteligente sobre a viabilidade econômica e computacional da rede.

Diante desse cenário, a utilização de diferentes protocolos para realizar a mesma estratégia de digitalização permite investigar em que medida a escolha da base tecnológica influencia os custos associados à criação das representações digitais dos dispositivos. Essa perspectiva possibilita avaliar o comportamento dessas soluções à medida que cresce a quantidade de ativos a serem registrados, aspecto relevante em ambientes IoT altamente escaláveis. Nesse sentido, este trabalho tem como objetivo analisar comparativamente os padrões ERC-721 e ERC-1155, avaliando o consumo de gas associado à mintagem das representações digitais de dispositivos IoT e seus efeitos sobre os recursos computacionais do nó blockchain em diferentes escalas de operação. Especificamente, busca-se:

- Implantar contratos ERC-721 e ERC-1155 parametrizados para a digitalização e registro de dispositivos IoT;
- Simular rotinas de minting sob variações de escala a partir do gerador de carga K6;
- Mensurar o consumo de gas (individual e acumulado);
- Monitorar o estresse do nó blockchain (uso de CPU e RAM);
- Comparar o comportamento e a eficiência operacional dos padrões.

Metodologia

A análise da literatura correlata evidencia que, a despeito da diversidade de arquiteturas e fluxos operacionais propostos para a gestão de dispositivos na Internet das Coisas, a operação de mintagem (*minting*) (o processo de criação e registro formal de um ativo digital na blockchain) constitui a etapa central no processo de digitalização. A execução dessa operação consome uma determinada quantidade de gás (*gas*), que é a taxa que mensura a quantidade de esforço computacional exigido da rede para processar e armazenar as transações. Nesse contexto, para garantir uma avaliação experimental equitativa e minimizar a influência de variáveis de confusão, como custos decorrentes

de regras de negócio secundárias, optou-se pela implementação de contratos inteligentes de complexidade mínima. Essa abordagem abstrai funcionalidades não essenciais e concentra a análise na mecânica pura de registro dos dispositivos, possibilitando a comparação direta do consumo de gas entre os protocolos ERC-721 e ERC-1155 sob condições equivalentes de carga.

Considerando esses aspectos, foram desenvolvidos dois contratos inteligentes de escopo mínimo, baseados em padrões distintos e implementados com o auxílio da biblioteca *OpenZeppelin*. Para garantir condições equivalentes de comparação, ambos os *smart contracts* compartilham a mesma lógica de registro e governança, restringindo a digitalização dos dispositivos ao administrador da rede (*onlyAdmin*). Nesse processo, a implementação ERC-721 utiliza a função *_safeMint* para atribuir a cada dispositivo um identificador único (*tokenId*) e uma URI direcionada aos metadados do ativo. Seguindo a mesma lógica de representação individual, no ERC-1155 cada dispositivo é associado a um *tokenId* distinto, sendo emitida exatamente uma unidade desse token (*supply* = 1) por meio da função *_mint*. Dessa forma, busca-se manter a equivalência funcional entre as implementações, de modo que a principal variável experimental seja o protocolo de tokenização empregado.

Para ilustrar a implementação delineada, os códigos dos contratos *IoTdeviceERC721* e *IoTdeviceERC1155* são apresentados nas Listagens 1 e 2, respectivamente.

A fase experimental foi conduzida no ecossistema Ethereum, uma blockchain pública e de código aberto que oferece suporte à implementação de contratos inteligentes baseados nos padrões de tokenização analisados neste estudo. Nesse contexto, os *smart contracts* foram desenvolvidos em Solidity e inicialmente prototipados na IDE Remix, etapa destinada à verificação da lógica implementada e à depuração das funções de mintagem. Após essa validação inicial, a arquitetura foi migrada para o *framework* Hardhat, utilizado na implantação dos contratos e na execução dos experimentos em uma rede local simulada. Para hospedar esse ambiente, utilizou-se uma máquina virtual configurada no VirtualBox com o sistema operacional Ubuntu 20.10 (64 bits), 4 vCPUs e 3 600 MB de memória RAM. Por fim, para a aplicação das cargas de trabalho, foi empregada a ferramenta de teste de carga K6, por meio de scripts responsáveis por automatizar as requisições de registro dos dispositivos durante os testes de estresse.

Os testes foram conduzidos sob cargas progressivas de trabalho (N), contemplando os cenários N = {5, 10, 20, 50, 100, 200, 500, 1000, 2000, 5000} dispositivos. Em cada cenário, o K6 foi utilizado para submeter as operações de registro à rede, permitindo avaliar o comportamento das implementações à medida em que a carga escalava. Durante cada bateria de testes, foram monitoradas duas dimensões de desempenho: o custo computacional na perspectiva da rede blockchain, mensurado pelo consumo acumulado de gas e pelo consumo médio por token; e o impacto sobre a infraestrutura de hospedagem do nó, avaliado pelo uso médio de CPU (%) e pelo uso médio de memória RAM (MB).

Resultados e Discussões

A partir dos experimentos realizados, os resultados obtidos foram consolidados e analisados comparativamente entre os padrões ERC-721 e ERC-1155, considerando a evolução das métricas em função do aumento progressivo da carga de dispositivos registrados. Para facilitar a interpretação desse comportamento, os dados foram organizados em representações gráficas que permitem observar tanto as variações no consumo de *gas* quanto os efeitos da carga sobre os recursos computacionais da infraestrutura. A análise apresentada a seguir busca, portanto, evidenciar as diferenças observadas entre as duas implementações e como essas se manifestam à medida que o número de dispositivos digitalizados aumenta.

A Figura 1 apresenta a evolução do consumo acumulado de *gas* para ambos os contratos em função das cargas de trabalho aplicadas.

É possível observar que a variação do consumo acumulado de gas em escala *log-log* revela comportamentos distintos entre os protocolos no experimento. Para cargas inferiores a aproximadamente 22 dispositivos, o contrato baseado em ERC-721 apresenta menor consumo, comportamento associado ao seu menor custo de deploy em relação ao ERC-1155, cujo impacto é mais representativo nas cargas iniciais. Entretanto, identifica-se um ponto de cruzamento (*break-even*) em N≈22, no qual ambos os padrões apresentam consumo equivalente de aproximadamente 2,47×106 unidades de *gas*. A partir desse limiar, o ERC-1155 passa a apresentar menor consumo acumulado, indicando que, com o aumento da quantidade de dispositivos, o custo inicial de implantação perde representatividade frente às diferenças de consumo associadas à emissão dos *tokens*. Esse comportamento se torna mais evidente nos cenários de média e larga escala (N≥50), indicando uma vantagem do ERC-1155 em termos de consumo de *gas* nas condições avaliadas.

Para evidenciar a diferença relativa entre os padrões, a Figura 2 apresenta a variação percentual do consumo acumulado de *gas* mensurado na representação gráfica anterior.

A análise percentual reforça o comportamento inicialmente observado. Nas menores cargas, os valores negativos indicam maior consumo acumulado do contrato baseado no ERC-1155, com uma diferença de aproximadamente 5,1% para N=5, que diminui progressivamente com o aumento da quantidade de ativos registrados. Próximo a N=22, essa diferença converge para zero, caracterizando o ponto de *break-even* entre os padrões. A partir desse ponto, os valores positivos passam a representar menor consumo acumulado do ERC-1155 em relação ao ERC-721, com a diferença aumentando progressivamente até atingir aproximadamente 6,6% em N=5000. Observa-se, ainda, que esse crescimento se torna menos acentuado nas cargas de trabalho mais elevadas, indicando uma tendência de estabilização da vantagem relativa do ERC-1155 em termos de consumo acumulado de gas.

Complementando esta avaliação, a Figura 3 apresenta o consumo médio de *gas* por *token* para os padrões ERC-721 e ERC-1155 em função da carga de dispositivos, desconsiderando o custo de implantação dos contratos.

O gráfico evidencia que o consumo médio de *gas* por *token* no contrato baseado em ERC-1155 permanece praticamente constante em aproximadamente 57,4 mil unidades de gas, independentemente da carga aplicada. Por outro lado, o contrato baseado em

ERC-721 apresenta maior consumo médio nas cargas iniciais, partindo de aproximadamente 65 mil unidades para N=5 e diminuindo gradualmente até se estabilizar próximo a 61,6 mil unidades nas cargas mais elevadas. Essa redução decorre da maior representatividade da primeira operação de mintagem do ERC-721, cujo custo superior é progressivamente diluído no cálculo da média à medida que novas operações são realizadas. Ainda assim, o consumo médio do ERC-721 permanece superior ao observado para o ERC-1155, corroborando o comportamento do consumo acumulado apresentado nas figuras anteriores.

Além das métricas relacionadas ao consumo de *gas*, analisou-se o comportamento dos recursos computacionais do nó blockchain durante o processamento das transações em ambos os contratos. Para essa avaliação, foram considerados o uso médio de CPU e o consumo médio de memória RAM ao longo das diferentes cargas de trabalho, permitindo observar os efeitos de cada implementação sobre a infraestrutura de execução. As figuras subsequentes apresentam a evolução dessas métricas e possibilitam a comparação entre os padrões sob as condições avaliadas.

A Figura 4 apresenta a utilização média de CPU e sua respectiva variabilidade para os padrões ERC-721 e ERC-1155 em função do aumento da carga de dispositivos digitalizados.

Observa-se, para ambas as implementações, um aumento progressivo da utilização média do processador à medida que aumenta a quantidade de simulações de registro de dispositivos, seguido por uma tendência de estabilização nas cargas mais elevadas. Para $N \geq 1000$, os valores médios permanecem próximos, situando-se aproximadamente entre 68% e 75%, enquanto, em N=2000 e N=5000, ambos os padrões convergem para valores próximos de 73% a 75%. Esse comportamento indica que, apesar das diferenças observadas nas cargas intermediárias, não se verifica uma diferença expressiva e persistente entre os padrões quanto à utilização média de CPU nos maiores níveis de carga avaliados.

Em relação à variabilidade, os desvios padrão apresentam valores mais elevados nas cargas intermediárias, alcançando aproximadamente 40% em determinados cenários, e diminuem progressivamente com o aumento de N. Nas maiores cargas, essa redução torna-se mais evidente, embora os valores ainda sejam superiores a 15% para o ERC-1155 em N=1000, tornando-se inferiores a 10% para ambas as implementações apenas em N=5000. Essa reação evidencia maior estabilidade na utilização da CPU durante os experimentos de maior duração, em contraste com as cargas menores e intermediárias, nas quais a alternância entre períodos de maior e menor atividade (sob tempo curto de monitoramento) exerce maior influência sobre a dispersão das amostras.

Por fim, a Figura 5 apresenta o consumo médio de memória RAM e sua respectiva variabilidade para ambas as implementações em função do aumento da carga de dispositivos digitalizados.

Nota-se que o consumo médio de memória RAM apresenta variações distintas entre as implementações nas cargas iniciais, com maior utilização durante a atividade de requisições ao *smart contract* baseado em ERC-721, entre N=5 e N=20. A partir das cargas intermediárias, entretanto, os valores tornam-se mais próximos, permanecendo

em torno de 200 a 230 MB até N=1000. Nas cargas mais elevadas, verifica-se um crescimento mais acentuado do consumo para ambos os contratos, alcançando aproximadamente 280 MB para o serviço baseado em ERC-1155 e 290 MB para o ERC-721 em N=5000. Apesar das diferenças pontuais ao longo dos experimentos, as curvas apresentam comportamento semelhante com o aumento da carga, não evidenciando uma vantagem persistente de um dos padrões quanto ao consumo médio de memória RAM.

Quanto à variabilidade, os desvios padrão permanecem reduzidos nas cargas iniciais e intermediárias, mas aumentam de forma mais pronunciada a partir de N=1000, atingindo aproximadamente 47 MB para o ERC-1155 e 48 MB para o ERC-721 em N=5000. Além disso, a proximidade entre ambos os valores de desvio padrão nas maiores cargas de estresse indica um comportamento semelhante quanto à dispersão do consumo de memória. Dessa forma, diferentemente do observado para a utilização de CPU, cuja variabilidade diminuiu nas cargas mais elevadas, a memória RAM apresenta maior dispersão conforme o crescimento da carga de trabalho.

Conclusão

A digitalização de dispositivos IoT por meio de NFTs constitui uma alternativa para o estabelecimento de identidades digitais em arquiteturas descentralizadas, porém sua aplicação em cenários de maior escala torna relevante a avaliação dos custos associados às operações de registro desses ativos. Partindo dessa perspectiva, este trabalho analisou comparativamente o comportamento dos padrões ERC-721 e ERC-1155 na representação individual de dispositivos IoT, submetendo implementações funcionalmente equivalentes a diferentes cargas de trabalho. A análise considerou tanto o consumo de *gas* decorrente das operações de mintagem quanto os efeitos do processamento das transações sobre os recursos computacionais do nó blockchain, permitindo avaliar o impacto da escolha do padrão sob as mesmas condições experimentais.

Os resultados relacionados ao consumo de gas demonstraram que a escolha do padrão de tokenização influencia o custo da digitalização à medida que a quantidade de dispositivos registrados aumenta. Embora o ERC-721 tenha apresentado menor consumo acumulado nas cargas iniciais, essa vantagem foi revertida após o ponto de *break-even*, identificado em aproximadamente 22 dispositivos. A partir desse limiar, a diferença relativa tornou-se progressivamente favorável ao ERC-1155, alcançando cerca de 6,6% na maior carga avaliada. Esse comportamento é corroborado pelo consumo médio por *token*, que se manteve inferior no ERC-1155 e, com o aumento da carga, passou a exercer maior influência sobre o consumo acumulado em relação aos custos iniciais de implantação. Com isso, pôde-se identificar que, à medida que a escala de digitalização aumenta, o menor custo recorrente de emissão do ERC-1155 se reflete em uma redução progressiva do consumo acumulado de gas em relação ao ERC-721.

Em contrapartida, a análise dos recursos computacionais não revelou uma vantagem persistente entre as implementações. Apesar das variações observadas ao longo das

cargas, CPU e memória RAM apresentaram comportamentos próximos nos cenários mais elevados. Assim, a redução de *gas* observada para o ERC-1155 não se refletiu em uma diminuição proporcional da demanda computacional do nó, indicando que o principal impacto observado na transição entre os padrões esteve associado ao custo das operações executadas na rede blockchain.

Assim, como continuidade deste estudo, pretende-se desenvolver uma arquitetura descentralizada para o registro e a autenticação de dispositivos IoT baseada na tokenização dos ativos em blockchain, adotando o ERC-1155 como padrão para a emissão de NFTs. Essa ampliação permitirá avaliar seu comportamento em um fluxo funcional mais abrangente, incorporando outras etapas do gerenciamento e da autenticação dos dispositivos. Dessa forma, busca-se expandir os parâmetros de comparação com as arquiteturas propostas na literatura que empregam o ERC-721 como base para a representação digital dos dispositivos na Internet das Coisas.

Referências

Garba, A., Khoury, D. J., Balian, P., Haddad, S., Sayah, J., Chen, Z., Guan, Z., Hamdan, H., Charafeddine, J., and Mutib, K. A. (2023). Lightcert4iots: Blockchain-based lightweight certificates authentication for IoT applications. IEEE Access, 11:28370–28383.

Goswami, B. and Choudhury, H. (2022). A blockchain-based authentication scheme for 5g-enabled IoT. Journal of Network and Systems Management, 30(4):61–85.

Hameed, K., Garg, S., Amin, M. B., and Kang, B. (2021). A formally verified blockchain-based decentralised authentication scheme for the internet of things. The Journal ofSupercomputing, 77(7):14461–14501.

Khalil, U., Malik, O.A., Hong, O.W. et al. Leveraging a novel NFT-enabled blockchain architecture for the authentication of IoT assets in smart cities. Sci Rep 13, 19785 (2023). https://doi.org/10.1038/s41598-023-45212-1.

Khashan, O. A. and Khafajah, N. M. (2023). Efficient hybrid centralized and blockchain-based authentication architecture for heterogeneous IoT systems. Journal of King Saud University — Computer and Information Sciences, 35(2):726–739.

Liou, W.-C. and Lin, T. (2021). T-auth: A novel authentication mechanism for the IoT based on smart contracts and pufs. In 2021 IEEE International Conference on Communications Workshops (ICC Workshops), pages 1–6.

Liu, Y., Liu, A., Xia, Y., Hu, B., Liu, J., Wu, Q., and Tiwari, P. (2024). A blockchain-based cross-domain authentication management system for iot devices. IEEE Transactions on Network Science and Engineering, 11(1):115–127.

Majd, N. E. and Sharko, M. (2023). Secure and cost effective IoT authentication and data storage framework using blockchain nft. In Proceedings of the 32nd International Conference on Computer Communications and Networks(ICCCN), pages 1–7.

Qiu, S., Li, J., Di, X., Li, X., Wu, Y., and Ibrahim, M. (2024). Lightweight mutual authentication scheme based on blockchain for the internet of medical things. IEEE Internet of Things Journal, 12(7):8848–8861.

Anexos



Listagem 1 – Contrato inteligente IoTdeviceERC721 baseado no padrão ERC-721.



Listagem 2 – Contrato inteligente IoTdeviceERC1155 baseado no padrão ERC-1155.

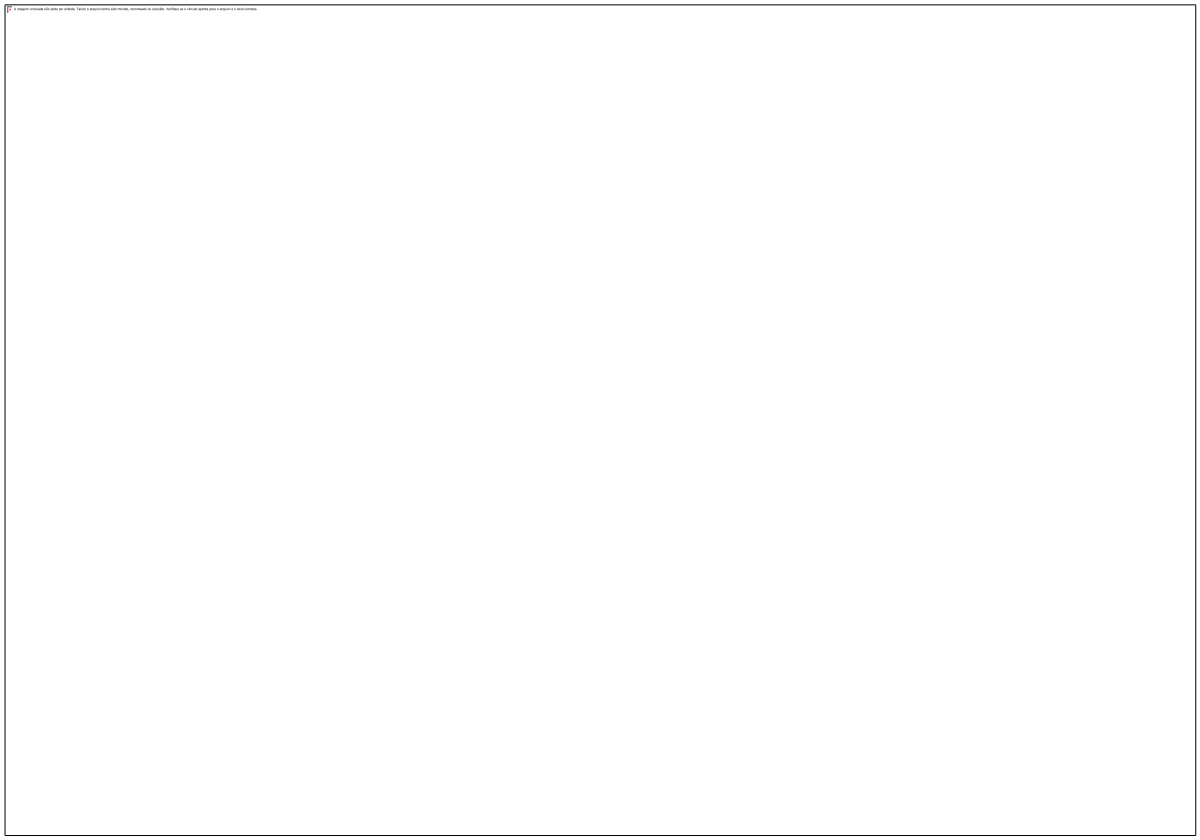


Figura 1 – Evolução do consumo acumulado de gas dos padrões ERC-721 e ERC-1155 sob diferentes cargas.



Figura 2 – Diferença percentual no consumo acumulado de gas entre os padrões ERC-1155 e ERC-721.

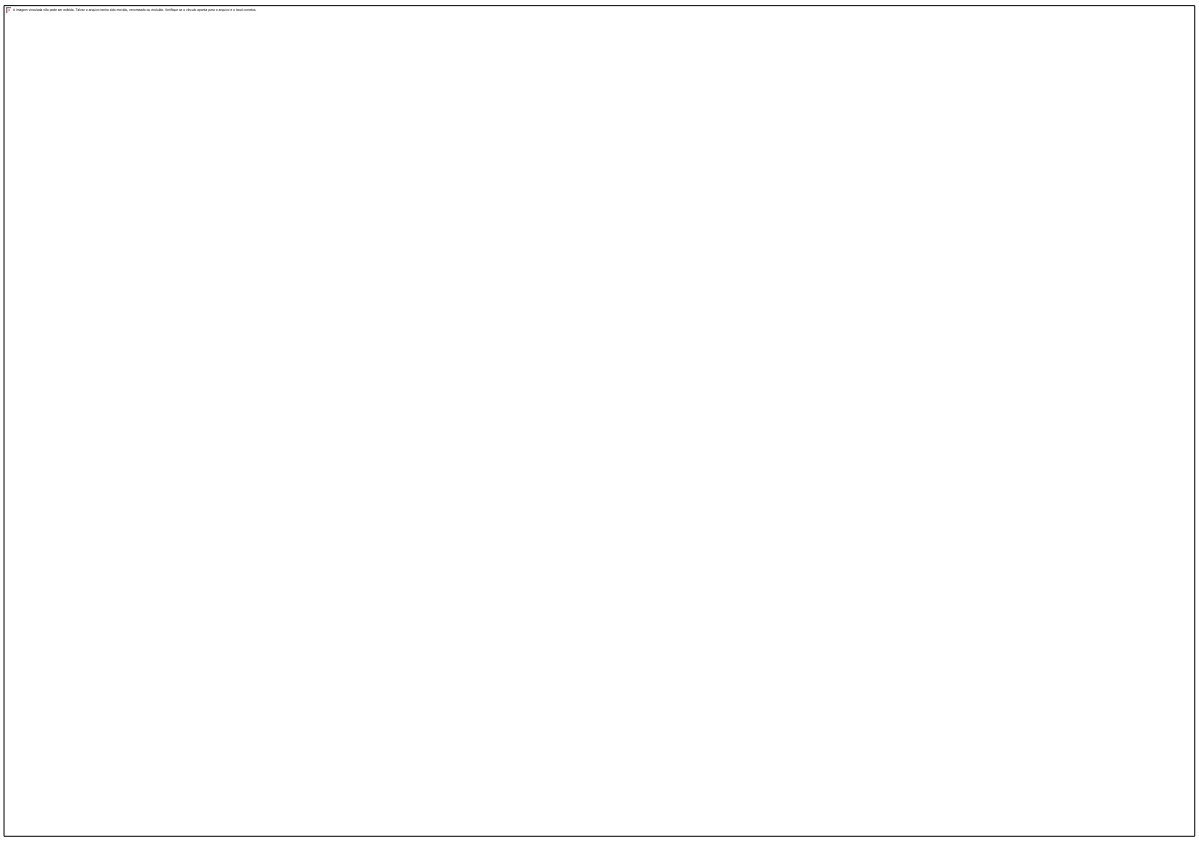


Figura 3 – Consumo médio de gas por token dos padrões ERC-721 e ERC-1155 em função da carga de dispositivos.

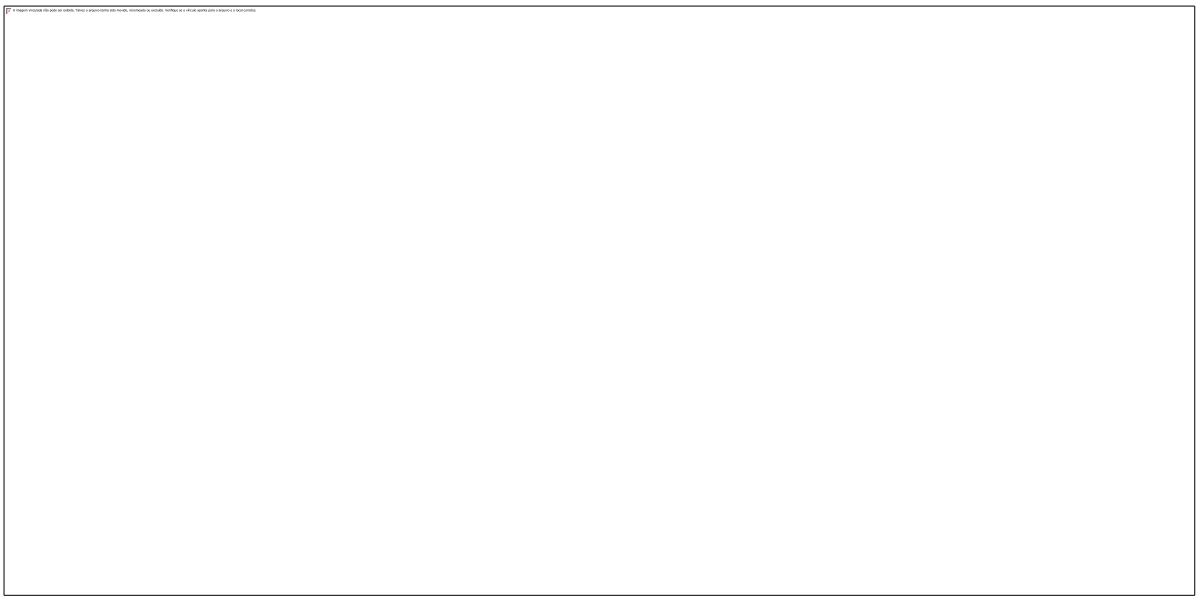


Figura 4 – Uso médio de CPU e respectivo desvio padrão para as implementações ERC-721 e ERC-1155 sob diferentes cargas de dispositivos registrados.



Figura 5 – Uso médio de memória RAM e respectivo desvio padrão para as implementações ERC-721 e ERC-1155 sob diferentes cargas de dispositivos.

CÓDIGO: ET1222

AUTOR: DAVI WHAGNER DE LIMA HOFFMANN

COAUTOR: STEFANE DE ASSIS ORICHUELA

COAUTOR: DAVI AUGUSTUS FIGUEIREDO LIMA PINHEIRO

COAUTOR: JONATAS SILVA SOARES

COAUTOR: SAMUEL SOARES PEREIRA COSTA

COAUTOR: GUILHERME SEVERO DE SOUZA

ORIENTADOR: ALYSON MATHEUS DE CARVALHO SOUZA

TÍTULO: Análise Multidimensional de Copresença e Imersão em Concerto de Djembe: Estudo Comparativo entre Realidade Virtual HMD, Projeção CAVE e Janela Virtual

Resumo

Este trabalho investiga a percepção de presença e imersão do público em uma performance musical virtual por meio de um estudo de caso comparativo do sistema multiusuário Djembefolar. Foram analisadas três formas de apresentação de um concerto de Djembe: Realidade Virtual (VR) imersiva por meio de dispositivos HMD, ambiente de projeção CAVE, composto por seis superfícies de projeção, e uma Janela Virtual, utilizada como ponte audiovisual entre os contextos físico e virtual. A avaliação envolveu 32 participantes, distribuídos entre três grupos experimentais: Grupo A, com 10 participantes em VR; Grupo B, com 16 participantes na Janela Virtual; e Grupo C, com 6 participantes na CAVE. A experiência foi analisada a partir de cinco dimensões relacionadas à presença: riqueza sensorial, qualidade audiovisual, interatividade, realismo social e familiaridade com a tecnologia. Os resultados indicaram maior riqueza sensorial e maior sensação de proximidade com o músico no Grupo A, desempenho sonoro comparável entre VR e CAVE e maior percepção de presença social no Grupo C. A Janela Virtual apresentou menor imersão, mas reduziu a necessidade de conhecimento técnico relacionado ao uso de dispositivos HMD, configurando-se como alternativa acessível para públicos menos familiarizados com realidade virtual. Conclui-se que a escolha da tecnologia deve considerar o equilíbrio entre imersão, acessibilidade e interação social.

Palavras-chave: Realidade Virtual. CAVE. Janela Virtual. Copresença. Performance.

TITLE: Multidimensional Analysis of Co-presence and Immersion in a Djembe Concert: A Comparative Study of HMD Virtual Reality, CAVE Projection, and Virtual Window

Abstract

This study investigates audience presence and immersion in a virtual musical performance through a comparative case study of the multi-user Djembefolar system. Three presentation configurations were analyzed during a Djembe concert: immersive Virtual Reality (VR) using head-mounted displays, a CAVE projection environment composed of six projection surfaces, and a Virtual Window used as an audiovisual bridge between physical and virtual contexts. The study involved 32 participants distributed across three experimental groups: Group A, with 10 participants in VR; Group B, with 16 participants using the Virtual Window; and Group C, with 6 participants in the CAVE. The experience was examined through five dimensions related to presence: sensory richness, audiovisual quality, interactivity, social realism, and familiarity with the technology. Results indicated greater sensory richness and a stronger sense of closeness to the musician in Group A, comparable auditory evaluations between VR and CAVE, and stronger social presence in Group C. The Virtual Window provided lower immersion but reduced the need for technical knowledge related to HMD use, making it an accessible alternative for audiences less familiar with virtual reality. The findings suggest that technology selection should consider the balance among immersion, accessibility, and social interaction.

Keywords: Virtual Reality. CAVE. Virtual Window. Co-presence. Performance.

Introdução

A integração entre arte e tecnologia tem ampliado as formas de produção e fruição de apresentações musicais, incorporando recursos audiovisuais e ambientes digitais às experiências do público [1]. No campo das performances imersivas, a Realidade Virtual permite construir espaços que deslocam o espectador do contexto físico imediato e oferecem novas perspectivas para a narrativa e para a relação com a obra [2]. Estudos sobre concertos em ambientes virtuais também apontam possibilidades de ampliar a participação do público e modificar sua percepção da apresentação musical [3].

Para compreender essas experiências, é necessário distinguir imersão e presença. A imersão corresponde às características técnicas do sistema que determinam o quanto ele substitui ou amplia os estímulos provenientes do ambiente físico, enquanto a presença está relacionada à resposta subjetiva do participante e à sensação de estar localizado no ambiente virtual [4]. Entre os fatores técnicos associados à imersão estão a quantidade de modalidades sensoriais reproduzidas, o campo de visão e a qualidade da apresentação

visual. Entretanto, a sensação de presença não depende exclusivamente do realismo visual, podendo também ser influenciada por rastreamento, taxa de quadros, som e formas de interação [4].

Além dos dispositivos HMD, outras configurações permitem explorar diferentes relações entre público, espaço e performance. A CAVE constitui um ambiente de projeção que envolve o participante por múltiplas superfícies, enquanto a chamada Janela Virtual estabelece uma relação visual entre um espaço real e um ambiente virtual, funcionando como uma interface entre os dois contextos [5,6]. Essas modalidades podem produzir experiências distintas quanto à percepção espacial, à interação e à presença social. Trabalhos sobre concertos virtuais, colaboração remota e presença social reforçam que a experiência do público não é determinada apenas pelo grau de imersão técnica, mas também pelas condições de interação e compartilhamento da experiência [3,7,8].

Apesar do avanço dessas pesquisas, permanece relevante comparar diferentes modalidades de exibição dentro de um mesmo contexto artístico e com critérios de avaliação comuns. Essa comparação pode contribuir para compreender quais características de cada configuração favorecem a riqueza sensorial, a qualidade audiovisual, a interatividade, o realismo social e a adaptação dos usuários. A partir dessa perspectiva, este trabalho utiliza o sistema multiusuário Djembefolar como estudo de caso e compara três configurações de apresentação.

São formuladas cinco questões de pesquisa: (RQ1) Como o tipo de plataforma influencia a percepção de riqueza sensorial, qualidade audiovisual, interatividade e realismo social? (RQ2) Qual plataforma proporciona maior sensação de imersão? (RQ3) Qual plataforma apresenta melhor percepção da apresentação sonora? (RQ4) Qual configuração favorece maior interatividade percebida entre público e músico? (RQ5) Em qual ambiente os participantes percebem maior interação com os demais espectadores?

O objetivo é identificar diferenças observadas entre as configurações e discutir as relações entre as características técnicas de cada plataforma e a experiência relatada pelos participantes. Dessa forma, o estudo busca contribuir para o desenvolvimento e a avaliação de experiências musicais imersivas e híbridas, considerando não apenas a intensidade da imersão, mas também acessibilidade e dimensão social.

Metodologia

2.1 Ferramentas e infraestrutura

O sistema de performance multiusuário Djembefolar foi desenvolvido utilizando o motor Unity, versão 6. Os modelos tridimensionais e as texturas utilizados na construção dos ambientes foram produzidos no Blender, versão 4.

O sistema foi estruturado para permitir a execução da mesma performance em diferentes configurações de apresentação. A arquitetura multiusuário possibilita que o músico e os

espectadores compartilhem o mesmo ambiente virtual, enquanto parâmetros da experiência são sincronizados entre os participantes.

2.2 Estrutura da performance

A apresentação foi organizada em três atos. No primeiro, a cena ocorre em um auditório virtual de uma escola de música associada a uma instituição de ensino superior. O músico inicia a apresentação tocando o Djembe e cada batida aciona efeitos visuais baseados em um algoritmo de comportamento de bando (boids), produzindo círculos de luz que se deslocam a partir do instrumento em direção ao público. Elementos vocais previamente gravados são incorporados durante o ato.

O segundo ato conduz a experiência do auditório para uma representação virtual de uma savana africana inspirada na região da Guiné. O ambiente apresenta vegetação, terreno aberto e uma fogueira posicionada próxima ao músico. As partículas luminosas da fogueira e os elementos boids respondem aos batimentos do Djembe, estabelecendo uma relação entre os estímulos musicais e visuais. Ao final do ato, os elementos luminosos formam um anel que envolve o cenário e conduz o retorno ao auditório virtual.

No terceiro ato, a experiência retorna ao auditório. Os elementos visuais associados aos boids convergem novamente para o Djembe e, ao final da apresentação, desaparecem, dando lugar aos créditos.

2.3 Rede multiusuário

Nas configurações de Realidade Virtual e Janela Virtual, foi utilizada a biblioteca de rede disponível no Unity para estabelecer a comunicação entre os participantes. A arquitetura emprega uma conexão centralizada associada ao Unity Cloud. O administrador cria a sala de conexão e os demais participantes ingressam como clientes, sendo posicionados no anfiteatro virtual.

Entre os parâmetros sincronizados estão sceneAct, que identifica o ato atual da experiência; totalBoids, que representa a quantidade de elementos boids ativos; e virtualWindows, atributo booleano utilizado para identificar a configuração de participação associada à Janela Virtual.

2.4 Janela Virtual

A Janela Virtual foi concebida como uma ponte entre os contextos físico e virtual. Enquanto o músico e parte da audiência participam da experiência em Realidade Virtual, outro grupo acompanha a apresentação por meio de uma projeção em um plano vertical.

Para implementar essa configuração, foi desenvolvida uma API local responsável por capturar a imagem de uma webcam e transmiti-la para a aplicação. A captura opera a 30 quadros por segundo e é convertida em uma sequência de bytes. Na aplicação Unity, os dados são reconstruídos como uma textura dinâmica aplicada a um plano no ambiente virtual. A sequência é compartilhada com os demais clientes conectados, permitindo que a transmissão seja visualizada em sincronia.

2.5 Participantes e procedimentos de teste

O estudo contou com 32 participantes, distribuídos entre três grupos experimentais. O Grupo A foi composto por 10 participantes na configuração de Realidade Virtual imersiva; o Grupo B reuniu 16 participantes na configuração de Janela Virtual; e o Grupo C contou com 6 participantes na configuração CAVE.

O Grupo A utilizou dispositivos Oculus Quest 2 ou 3 para acompanhar a apresentação em ambiente imersivo. O músico encontrava-se em uma sala separada e sua representação era transmitida em tempo real para os espectadores por meio da conexão de rede. O Grupo B permaneceu em uma sala dedicada e acompanhou a apresentação por uma projeção vertical correspondente à Janela Virtual. O Grupo C participou da experiência dentro de uma sala CAVE composta por seis superfícies de projeção gerenciadas pelo software MadMapper, utilizando uma gravação em vídeo 360 graus da performance do músico.

Também foi planejado um Grupo D, correspondente a uma apresentação com o músico fisicamente presente e uma projeção virtual ao fundo. Esse grupo não foi executado devido às limitações de hardware e, portanto, não foi utilizado nas comparações empíricas.

2.6 Instrumento de avaliação

Após as sessões experimentais, os participantes responderam a um questionário em escala Likert de 1 a 7, variando de discordo totalmente a concordo totalmente. As questões avaliaram seis aspectos da experiência: satisfação visual; satisfação sonora; confusão durante a mudança para a savana; entretenimento associado à interação do músico com a música; sensação de presença espacial; e tensão provocada pelos efeitos visuais.

Esses itens foram utilizados como indicadores para discutir as cinco dimensões gerais consideradas no estudo: riqueza sensorial, qualidade audiovisual, interatividade, realismo social e familiaridade com a tecnologia. Como os dados disponíveis são apresentados de forma agregada e descritiva no estudo de origem, a análise deste artigo mantém o caráter comparativo e descritivo, sem atribuir significância estatística a diferenças para as quais não foi apresentado teste estatístico.

Resultados e Discussões

Os resultados permitem observar diferenças entre as três configurações avaliadas. A interpretação foi organizada segundo as dimensões de presença utilizadas no estudo e considerando as respostas obtidas no questionário. Como o conjunto de dados apresentado não contém testes estatísticos ou valores individuais completos, os resultados são discutidos em termos de diferenças observadas entre os grupos, sem afirmar significância estatística.

3.1 Riqueza sensorial e qualidade audiovisual

O Grupo A, correspondente à Realidade Virtual imersiva, apresentou as avaliações mais altas para a experiência visual e sonora. O isolamento dos estímulos externos e a utilização de áudio espacializado favoreceram uma experiência sensorial mais concentrada e uma maior sensação de proximidade com o músico. Esse resultado é compatível com estudos que associam características de imersão técnica e presença à forma como o usuário percebe e interage com o ambiente virtual [4,9].

O Grupo C, por sua vez, apresentou avaliação sonora próxima à observada na Realidade Virtual. A configuração CAVE permitiu uma apresentação audiovisual em escala ampla, mas a projeção em superfícies físicas impôs limitações espaciais que afetaram a percepção de profundidade. Essa característica é coerente com a própria natureza dos ambientes CAVE, nos quais a imersão é obtida por meio de múltiplas superfícies de projeção [5].

O Grupo B apresentou os menores níveis de riqueza sensorial entre as três configurações. Ainda assim, a Janela Virtual permitiu que os participantes acompanhassem a performance e os elementos visuais do espetáculo. O resultado evidencia um compromisso entre a intensidade da imersão e a facilidade de acesso: a ausência de um HMD reduz a imersão espacial, mas também simplifica a participação.

3.2 Qualidade visual e transições de cena

As representações tridimensionais do auditório e da savana foram avaliadas positivamente nos grupos A e C, com destaque para a clareza dos elementos gráficos e para a consistência visual do cenário. A avaliação reforça a importância da apresentação audiovisual em experiências musicais imersivas [3,9].

Uma dificuldade observada ocorreu durante a transição entre o auditório e a savana, especialmente no Grupo C. Participantes relataram confusão e desorientação durante a mudança de cena. A dimensão física da projeção pode ter contribuído para essa percepção, uma vez que mudanças amplas de referência visual podem alterar a orientação espacial do participante. Em ambientes imersivos, questões relacionadas a movimento e cinetose constituem fatores relevantes para a experiência [10].

No Grupo B, a ausência de uma imersão espacial completa reduziu a ocorrência desse tipo de desorientação. Entretanto, a própria moldura da Janela Virtual estabeleceu uma distância perceptiva em relação ao ambiente virtual, reduzindo o envolvimento espacial quando comparado às configurações imersivas.

3.3 Interatividade e proximidade com o músico

A percepção de interatividade foi mais destacada no Grupo A, no qual a experiência imersiva proporcionou maior sensação de proximidade com a representação virtual do músico. A relação entre participação, controle e presença é discutida na literatura de ambientes virtuais e presença social [7,8].

No Grupo C, a interatividade assumiu caráter predominantemente coletivo. A ausência de HMDs permitiu que os participantes permanecessem visualmente conscientes da presença dos demais espectadores, favorecendo o compartilhamento da experiência. Esse

aspecto se relaciona ao conceito de copresença, associado à percepção de compartilhar um ambiente ou experiência com outras pessoas [8,11].

Na Janela Virtual, a interação foi percebida como mais limitada. A separação entre o público físico e o ambiente virtual estabeleceu uma barreira espacial e reduziu a possibilidade de participação direta. Ainda assim, a configuração permitiu uma forma de conexão entre os dois contextos [6].

3.4 Realismo social

O Grupo C apresentou a maior percepção de realismo social. A possibilidade de permanecer no mesmo espaço físico que outros participantes, enquanto todos acompanhavam as projeções, favoreceu a sensação de assistir à apresentação coletivamente. Esse resultado está relacionado ao papel da presença social na construção de experiências compartilhadas em ambientes virtuais [8,11].

O Grupo A também apresentou percepção positiva de realismo, mas a experiência foi mais orientada para a relação entre espectador e músico virtual. A utilização dos HMDs reduz a percepção direta do espaço físico e modifica a maneira como a audiência percebe os demais participantes.

No Grupo B, o realismo social foi menor. A Janela Virtual permitiu a visualização da performance e estabeleceu uma ligação entre os contextos, mas não reproduziu integralmente as condições de um ambiente físico compartilhado. Estudos sobre concertos virtuais indicam que a sensação de conexão social constitui um componente relevante da experiência musical [12].

3.5 Familiaridade tecnológica e acessibilidade

A familiaridade prévia com realidade virtual mostrou-se relevante principalmente nas configurações que exigiam o uso de dispositivos HMD. Participantes com maior experiência apresentaram maior facilidade de adaptação e maior percepção de imersão, resultado coerente com a influência de fatores individuais sobre a experiência de presença [4].

Na Janela Virtual, a necessidade de conhecimento técnico específico para operar um HMD foi reduzida. Isso tornou a participação mais direta para usuários menos familiarizados com realidade virtual. O resultado não significa que a Janela Virtual seja mais imersiva, mas evidencia uma vantagem relacionada à acessibilidade de uso.

Assim, as três configurações apresentam vantagens diferentes. A Realidade Virtual favorece a concentração sensorial e a proximidade com o músico; a CAVE favorece a experiência coletiva; e a Janela Virtual reduz barreiras técnicas de entrada. Essa complementaridade reforça a importância de selecionar a plataforma de acordo com o objetivo da experiência artística.

3.6 Discussão das questões de pesquisa

RQ1. O tipo de plataforma influenciou as percepções dos participantes. Os grupos A e C apresentaram resultados superiores ao Grupo B em diferentes aspectos de riqueza sensorial, interatividade e realismo social, embora cada configuração tenha produzido um perfil de experiência distinto.

RQ2. O Grupo A apresentou a maior percepção de imersão e de presença espacial, seguido pelo Grupo C. A experiência por HMD concentrou os estímulos no ambiente virtual e favoreceu a sensação de proximidade com o músico.

RQ3. O Grupo C apresentou percepção sonora comparável à da Realidade Virtual. Dessa forma, a CAVE mostrou-se capaz de proporcionar uma experiência auditiva de alta qualidade mesmo sem o uso de HMD.

RQ4. Não foi possível realizar uma comparação completa da interatividade entre público e músico, pois o Grupo D, que representaria a condição presencial com projeção virtual, não foi executado. Portanto, essa questão permanece parcialmente respondida pelos dados disponíveis.

RQ5. O Grupo C apresentou a maior percepção de interação e presença social entre os participantes. A possibilidade de compartilhar fisicamente o mesmo espaço contribuiu para a sensação de assistir à apresentação em conjunto.

Conclusão

Este estudo comparou três configurações tecnológicas para a apresentação de uma mesma experiência musical: Realidade Virtual imersiva, CAVE e Janela Virtual. Os resultados indicam que não há uma configuração única que maximize simultaneamente todas as dimensões analisadas.

A Realidade Virtual apresentou o maior nível de riqueza sensorial e maior sensação de proximidade com o músico, favorecendo experiências centradas na imersão individual. A CAVE destacou-se pela dimensão coletiva e pela percepção de presença social, além de apresentar avaliação sonora próxima à da VR. A Janela Virtual apresentou menor imersão, mas reduziu a necessidade de familiaridade técnica com dispositivos HMD, constituindo uma alternativa de acesso mais simples para públicos menos familiarizados com realidade virtual.

Entre as limitações do estudo está o tamanho da amostra, que restringe a generalização dos resultados. Além disso, o Grupo D não pôde ser executado, impossibilitando uma comparação empírica direta com uma apresentação presencial acompanhada de projeção virtual. Outra limitação é que os dados disponíveis foram analisados de forma descritiva, sem testes estatísticos que permitissem afirmar significância estatística entre os grupos.

Como trabalhos futuros, recomenda-se ampliar a amostra e aplicar procedimentos estatísticos adequados para aumentar a robustez das comparações; executar o protocolo

do Grupo D; e aprofundar o desenvolvimento da Janela Virtual como espaço de interação social entre participantes físicos e virtuais, incluindo a investigação da comunicação por voz entre os diferentes contextos.

Referências

[1] BYSTROM, K.-E.; BARFIELD, W.; HENDRIX, C. A conceptual model of the sense of presence in virtual environments. Presence: Teleoperators and Virtual Environments , v. 8, n. 2, p. 241-244, 1999. doi: 10.1162/105474699566107.

[2] ONDERDIJK, K. E. et al. Concert experiences in virtual reality environments. Virtual Reality , v. 27, n. 3, p. 2383-2396, 2023. doi: 10.1007/s10055-023-00814-y.

[3] SOUZA, A. M. D. C.; CAMPOS, C. D. S.; RENNÓ-COSTA, C. Opus imperfectum - a mixed reality concert about the origin of life. In: Symposium on Virtual and Augmented Reality , p. 294-298. ACM, 2023. doi: 10.1145/3625008.3625052.

[4] SANCHEZ-VIVES, M. V.; SLATER, M. From presence to consciousness through virtual reality. Nature Reviews Neuroscience , v. 6, n. 4, p. 332-339, 2005. doi: 10.1038/nrn1651.

[5] CRUZ-NEIRA, C. et al. The cave: audio visual experience automatic virtual environment. Communications of the ACM , v. 35, n. 6, p. 64-72, 1992. doi: 10.1145/129888.129892.

[6] LIN, B. Keep it real: a window to real reality in virtual reality. arXiv preprint arXiv:2004.10313 , 2020. doi: 10.48550/arXiv.2004.10313.

[7] LO, S.-Y.; LAI, C.-Y. Investigating how immersive virtual reality and active navigation mediate the experience of virtual concerts. Scientific Reports , v. 13, n. 1, p. 8507, 2023. doi: 10.1038/s41598-023-35369-0.

[8] SON, G.; TIEMANN, A.; RUBO, M. I am here with you: an examination of factors relating to social presence in social VR. Frontiers in Virtual Reality , v. 6, p. 1558233, 2025. doi: 10.3389/frvir.2025.1558233.

[9] PETROVIC, N. Augmented and virtual reality web applications for music stage performance. In: 2020 55th International Scientific Conference on Information, Communication and Energy Systems and Technologies (ICEST) , p. 33-36. IEEE, 2020. doi: 10.1109/ICEST49890.2020.9232713.

[10] GASSELSEDER, H.-P.; KALLIONPÄÄ, M. Beyond the audience seat: The recording and production of immersive opera and interactive concerto programmes for VR experiences. In: Proceedings of EVA London 2019 , 2019. doi: 10.14236/ewic/EVA2019.75.

[11] GOEBBELS, G.; LALIOTI, V. Co-presence and co-working in distributed collaborative virtual environments. In: Proceedings of the 1st International Conference on Computer Graphics, Virtual Reality and Visualisation , p. 109-114. ACM, 2001. doi: 10.1145/513867.513891.

[12] SWARBRICK, D. et al. Corona concerts: The effect of virtual concert characteristics on social connection and kama muta. Frontiers in Psychology , v. 12, p. 648448, 2021. doi: 10.3389/fpsyg.2021.648448.

[13] SILVA, I. M. C. Cinema VR: concepções teóricas e técnicas sobre a fruição do espectador em narrativas cinematográficas em 360 graus . Dissertação (Mestrado em Comunicação) - Universidade Federal da Paraíba, João Pessoa, 2021.

Anexos

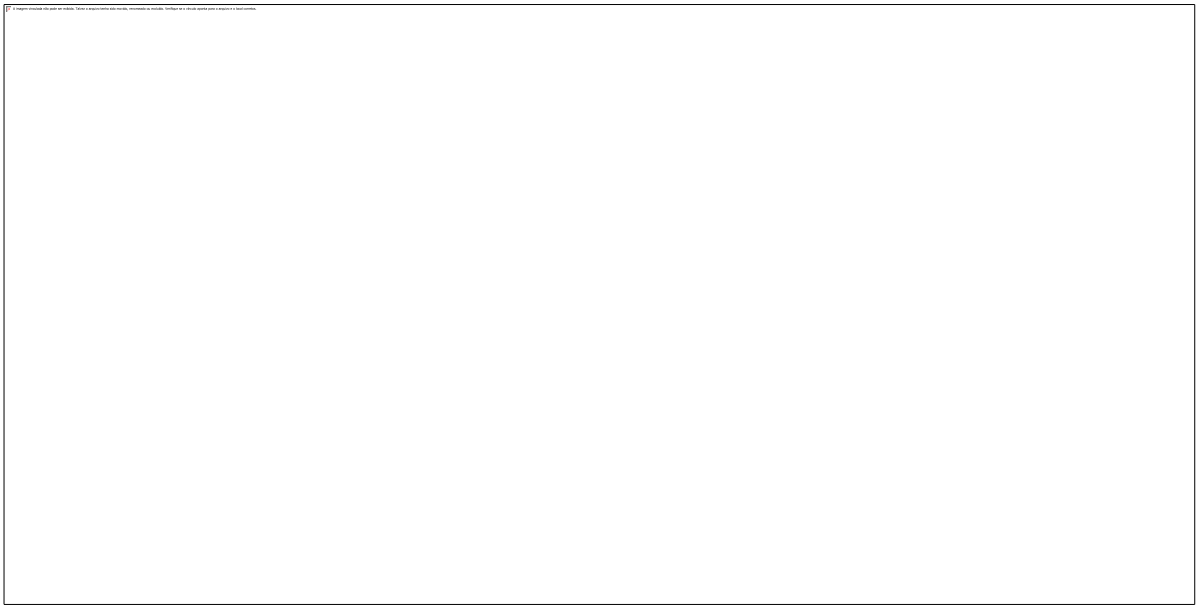


Figura 1 - Detalhe cênico do Ato 2 do sistema Djembefolar em ambiente tridimensional de savana africana.

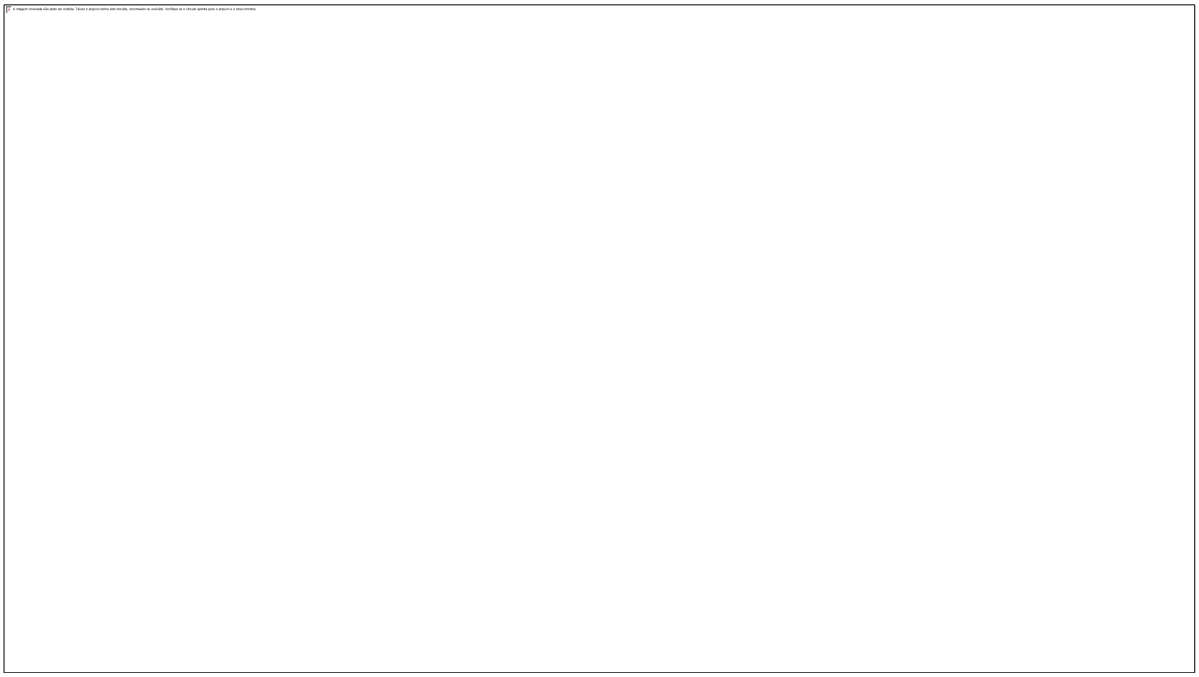


Figura 2 - Espectadores acompanhando a apresentação musical em escala real dentro da sala CAVE.

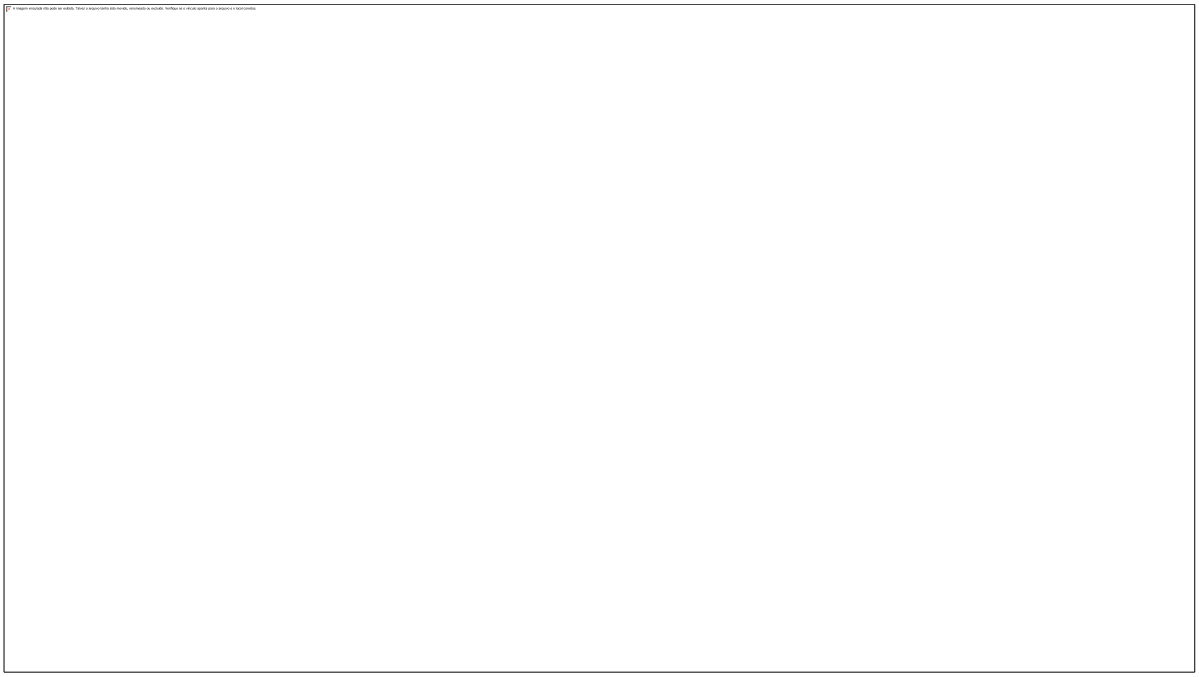


Figura 3 - Projeção de um momento do Ato 1 nas paredes internas da CAVE, com participantes observando os efeitosvisuais.

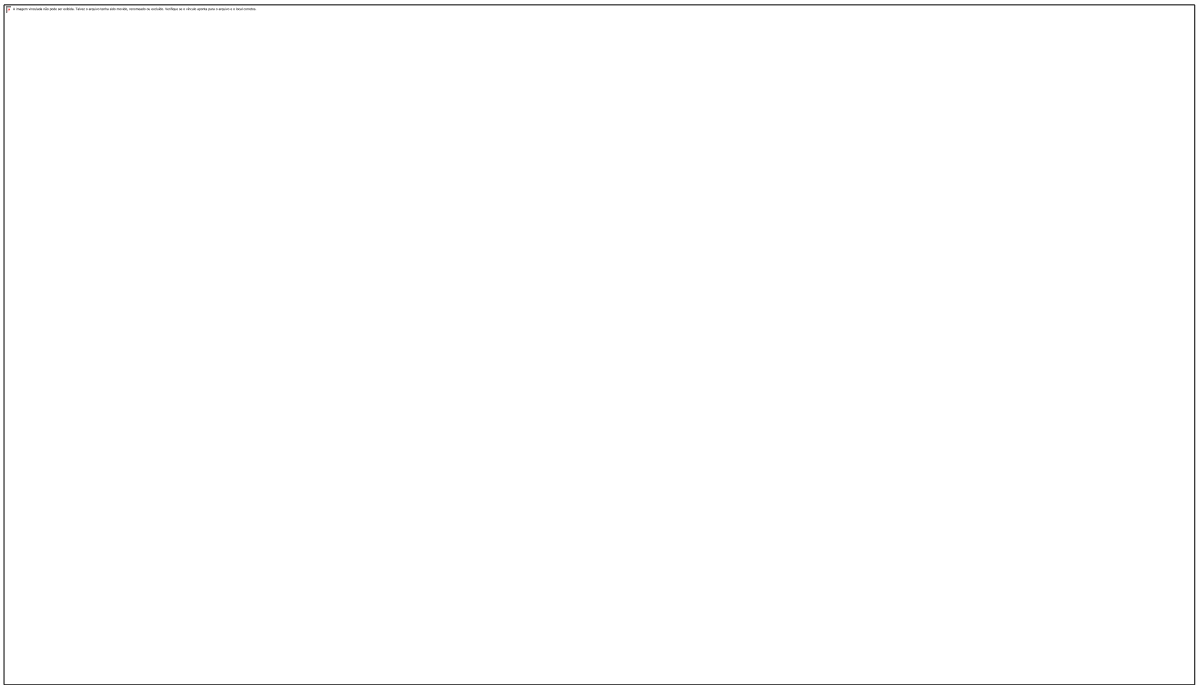


Figura 4 - Sala de testes destinada à audiência imersiva de realidade virtual, com participantes utilizando dispositivos HMD.



Figura 5 - Cena da apresentação observada por meio da Janela Virtual, com o Djembe, o músico e os elementos visuais daperformance.

CÓDIGO: ET1375

AUTOR: GABRIELLE DE VASCONCELOS BORJA

ORIENTADOR: TIBERIO AZEVEDO PEREIRA

TÍTULO: Uso de Modelos tipo BERT Aplicados à Análise de RNA-Seq

Resumo

O sequenciamento de RNA permite a caracterização de padrões de expressão gênica sob diferentes condições biológicas, mas sua alta dimensionalidade e as relações complexas entre genes criam desafios para as abordagens estatísticas e de aprendizado de máquina convencionais. Nesse contexto, modelos baseados na arquitetura Transformer, particularmente os modelos derivados do BERT, têm surgido como alternativas promissoras para representar e analisar dados biológicos. Esta pesquisa em andamento tem como objetivo investigar a aplicação de modelos baseados em BERT à análise de RNA-Seq e a predições em bioinformática. O fluxo de trabalho proposto inclui o pré-processamento, a normalização e o estudo de estratégias de tokenização biológica para conjuntos de dados de RNA-Seq em bulk; a adaptação ou ajuste fino (fine-tuning) de um modelo Transformer pré-treinado para uma tarefa específica, como a classificação de amostras. Estratégias de explicabilidade e interpretabilidade do modelo também serão exploradas para investigar se as representações aprendidas e os padrões de atenção fornecem insights biologicamente relevantes. Em seu estágio atual, o projeto está focado na revisão de arquiteturas adequadas, na seleção de conjuntos de dados e na avaliação do pipeline. A contribuição esperada é avaliar o potencial e as limitações dos modelos baseados em BERT para a análise de dados transcriptômicos, desenvolvendo um arcabouço interdisciplinar conectando bioinformática, biologia molecular e IA.

Palavras-chave: RNA-Seq, Transformers, BERT, Inteligência Artificial, Expressão Gênica

TITLE: Use of BERT models in RNA-Seq analysis

Abstract

RNA sequencing enables the characterization of gene expression patterns under different biological conditions, but its high dimensionality and the complex relationships among genes creates challenges for conventional statistical and machine-learning approaches. In this context, models based on the Transformer architecture, particularly BERT-derived models, have emerged as promising alternatives for representing and analyzing biological data. This ongoing research aims to investigate the application of BERT-based models to RNA-Seq analysis and bioinformatics

predictions. The proposed workflow includes the preprocessing, normalization and study of biological tokenization strategies for bulk RNA-Seq datasets; the adaptation or fine-tuning of a pretrained Transformer model for a specific task, such as sample classification. Model explicability and interpretability strategies will also be explored to investigate whether the learned representations and attention patterns provide biologically meaningful insights. At its current stage, the project focuses on reviewing suitable architectures, selecting datasets, and evaluating the pipeline. The expected contribution is to assess the potential and limitations of BERT-based models for transcriptomic data analysis while developing an interdisciplinary framework connecting bioinformatics, molecular biology, and deep learning.

Keywords: RNA-Seq, Transformers, BERT, Artificial Intelligence, Gene Expression

Introdução

A Inteligência Artificial se tornou uma ferramenta essencial em uma ampla gama de áreas, incluindo a biologia molecular, onde apoia a análise em larga escala de dados biológicos complexos. A crescente disponibilidade de conjuntos de dados genômicos tem ampliado as possibilidades de identificação de padrões de expressão, caracterização de estados celulares e investigação de mecanismos moleculares associados a doenças. No entanto, os dados de RNA-Seq são altamente dimensionais e envolvem relações complexas, não lineares e dependentes de contexto entre genes, que podem não ser totalmente capturadas por abordagens estatísticas e de aprendizado de máquina convencionais.

Nesse contexto, modelos baseados em Transformers têm surgido como alternativas promissoras para a representação e análise de dados biológicos. Seu mecanismo de autoatenção (self-attention) permite que o modelo capture relações entre múltiplos elementos de entrada e gere representações dependentes de contexto. Modelos biológicos baseados em BERT adaptam conceitos originalmente desenvolvidos para o processamento de linguagem natural, tratando genes, fragmentos de nucleotídeos ou perfis de expressão ordenados (ranked) como tokens de uma linguagem biológica. Modelos como o scBERT e o Geneformer utilizam perfis transcriptômicos para aprender representações de estados celulares, enquanto o DNABERT é voltado principalmente para a análise de sequências de nucleotídeos.

Esses modelos pré-treinados podem ser adaptados a tarefas biológicas específicas por meio de transfer learning e fine-tuning, incluindo anotação de tipos celulares, classificação de doenças, predição de resposta a tratamento e identificação de biomarcadores. Além do desempenho preditivo, métodos de interpretabilidade podem ajudar a identificar genes e vias biológicas associados às decisões do modelo, embora esses achados precisem ser validados por meio de análises de enriquecimento funcional e de evidências biológicas já existentes.

Metodologia

O fluxo de trabalho proposto investiga a adaptação de modelos pré-treinados baseados em Transformers a dados de RNA-Seq de tecido total (bulk), integrando pré-processamento, tokenização biológica, fine-tuning, avaliação preditiva e interpretabilidade do modelo. Como os dados biológicos não possuem uma estrutura natural de tokens, a etapa de tokenização exige a avaliação de diferentes estratégias já exploradas na literatura. O pipeline completo foi projetado para garantir reprodutibilidade e apoiar a análise de padrões biológicos extraídos dos dados de expressão gênica.

A arquitetura Transformer processa embeddings de entrada por meio de camadas repetidas de codificadores (encoders), compostas por autoatenção multi-cabeça (multi-head self-attention), redes feed-forward e camadas de normalização. A autoatenção permite que o modelo capture relações contextuais entre os tokens de entrada, enquanto os codificadores empilhados refinam progressivamente essas representações. O BERT se baseia nessa arquitetura por meio de aprendizado contextual bidirecional ao longo da pilha de codificadores. Neste projeto, um modelo biológico pré-treinado baseado em BERT será adaptado a dados de RNA-Seq tokenizados e ajustado (fine-tuned) com uma camada de classificação para a tarefa de predição selecionada.

As amostras de RNA-Seq bulk serão transformadas por meio de estratégias de codificação biológica inspiradas em modelos como o scBERT e o Geneformer. Métodos de seleção de características e redução de contexto serão explorados para adaptar os dados de expressão gênica às restrições de entrada do Transformer, incluindo filtragem de genes altamente variáveis (highly variable genes — HVG), esquemas de ranqueamento de genes, discretização/binning dos valores de expressão e técnicas de redução de dimensionalidade para reduzir o comprimento da sequência preservando o sinal biologicamente relevante. A análise de interpretabilidade, que rastreia as predições até os genes relevantes e os padrões aprendidos, é uma etapa central desse pipeline, pois pode revelar informações biológicas que abordagens estatísticas convencionais não capturam totalmente.

Resultados e Discussões

A revisão de literatura e o desenho do pipeline conduzidos até o momento, embora anteriores à fase experimental, já produziram insights metodológicos relevantes que orientam os próximos passos do projeto:

- A comparação entre as estratégias de tokenização biológica existentes revelou um trade-off entre a codificação baseada em ranqueamento (rank-based encoding, ex.: Geneformer), que é mais robusta a efeitos de lote (batch effects) e a variações de profundidade de sequenciamento, mas descarta a magnitude absoluta da expressão, e a discretização/binning (ex.: scBERT), que preserva melhor a escala de expressão, mas é mais sensível à variabilidade técnica. Esse

trade-off informará diretamente a estratégia de tokenização adotada para os dados de RNA-Seq bulk.

- Trabalhos relacionados que aplicam arquiteturas baseadas em BERT a RNA-Seq bulk, como o BulkRNABert, concentraram-se principalmente no desempenho preditivo, deixando a interpretabilidade em grande parte inexplorada — uma lacuna explicitamente reconhecida como trabalho futuro pelos próprios autores originais. Isso reforça a relevância da abordagem orientada à interpretabilidade proposta neste projeto.

- Como os pesos de atenção não têm garantia inerente de refletir explicações fiéis ou causais, a estratégia de avaliação foi desenhada para tratar os mapas de atenção como uma ferramenta geradora de hipóteses, e não como evidência definitiva, exigindo validação frente a conhecimento biológico independente por meio de análise de enriquecimento funcional (GO/KEGG).

- A escala de modelos pré-treinados comparáveis (ex.: BulkRNABert, com aproximadamente 6 milhões de parâmetros) indica que o fine-tuning é computacionalmente viável dentro da infraestrutura de HPC disponível, o que apoia a adoção de transfer learning em vez do treinamento de um modelo do zero.

Esses achados, embora preliminares, já moldaram decisões metodológicas importantes, incluindo a priorização das estratégias de tokenização a serem testadas e o desenho do pipeline de validação da interpretabilidade, e serão diretamente aplicados na próxima fase experimental do projeto.

Conclusão

Até o momento, este projeto em andamento resultou na definição de um desenho metodológico para a adaptação de modelos Transformer baseados em BERT a dados de RNA-Seq bulk, apoiado em uma revisão crítica da literatura que apontou trade-offs relevantes entre estratégias de tokenização gênica e a necessidade de validar padrões de atenção frente a conhecimento biológico já estabelecido. As próximas etapas do projeto envolvem a implementação do fine-tuning, a avaliação comparativa com modelos de aprendizado de máquina convencionais e a análise de enriquecimento funcional (GO/KEGG) dos genes mais relevantes segundo o modelo, com o objetivo de contribuir com um arcabouço interdisciplinar que conecta aprendizado profundo, biologia molecular e bioinformática.

Referências

Conesa A, et al. **A survey of best practices for RNA-seq data analysis.** Genome Biol. 2016;17:13. doi:10.1186/s13059-016-0881-8.

Suárez-Fariñas M, et al. **RNA sequencing atopic dermatitis transcriptome profiling provides insights into novel disease mechanisms with potential therapeutic implications**.J Allergy Clin Immunol. 2015;135(5):1218--1227. doi:10.1016/j.jaci.2015.03.003.

Zhang S, et al. **Applications of transformer-based language models in bioinformatics: a survey.** Bioinformatics Advances. 2023;3(1):vbad001. doi:10.1093/bioadv/vbad001.

Yang F, et al. **scBERT as a large-scale pretrained deep language model for cell type annotation of single-cell RNA-seq data.** Nat Mach Intell. 2022;4(10):852--866. doi:10.1038/s42256-022-00534-z.

Anexos



Pipeline

CÓDIGO: ET1463

AUTOR: CAIO VICTOR ALVES JUSTINO

COAUTOR: ANDRE SOARES GOMES MOREIRA

ORIENTADOR: ALYSON MATHEUS DE CARVALHO SOUZA

TÍTULO: Uma Arquitetura Transparente de Classificação Multiclasse para Categorização de Variáveis de Crédito

Resumo

A integração de dados alternativos na pontuação de crédito expõe traços sensíveis dos consumidores e coloca a privacidade em risco. Para neutralizar essa vulnerabilidade, apresentamos uma ferramenta automatizada de triagem. Ela categoriza descrições textuais de variáveis em uma taxonomia de onze níveis. Para contornar a escassez de amostras rotuladas, geramos um conjunto de dados sintéticos usando o modelo Gemma4 31B. O cenário regulatório exige transparência. Por isso, a arquitetura apoia-se exclusivamente em modelos explicáveis de aprendizado de máquina. A validação empírica comprovou alta coesão nos agrupamentos. A Regressão Logística alcançou 73,02% de acurácia. Esse fluxo de trabalho garante a interpretabilidade necessária para uma governança financeira segura.

Palavras-chave: Ferramenta automatizada de triagem. Governança. Pontuação de Crédito.

TITLE: A Transparent Multiclass Classification Architecture for Credit Variable Categorization

Abstract

Integrating Alternative data in credit scoring risks revealing sensitive consumer traits. To prevent privacy violations, this study proposes an automated screening tool that categorizes textual variable descriptions into an eleven category taxonomy. To overcome labeled data scarcity, the methodology generates a synthetic dataset using the Gemma4 31B model. Because regulations demand transparency, the architecture relies exclusively on explainable machine learning models. Validation on empirical data showed high clustering cohesion, and Logistic Regression achieved 73.02% accuracy. This workflow guarantees the interpretability required for safe financial governance.

Keywords: Automated screening tool. Credit Scoring. Governance.

Introdução

1. Introdução

Populações jovens e de países em desenvolvimento frequentemente carecem de um histórico financeiro formal, o que limita a capacidade dos modelos clássicos de estimar escores de risco precisos (GUO et al., 2016; BJöRKEGREN; GRISSEN, 2018). Para contornar essa lacuna, instituições financeiras integram hoje dados alternativos aos seus conjuntos de treinamento para mapear o comportamento e a cultura de crédito local (MWESIGWA; JACKSON; CSIKSZENTMIHALYI, 2026a; TOUNSI; HASSOUNI; ANOUN, 2017). A combinação dessas métricas com os avanços do big data deflagrou uma transformação profunda na avaliação de risco e predição de inadimplência (HURLEY; ADEBAYO, 2017).

O uso irrestrito de dados alternativos, contudo, impõe graves desafios técnicos e jurídicos. Essas informações atuam como variáveis proxy, revelando indiretamente características sensíveis dos consumidores (SCHWARCZ; PRINCE). Leis como a LGPD no Brasil (BRASIL, 2018) e o GDPR na Europa (European Union, 2016) exigem tratamentos distintos para diferentes tipos de informação. Inserir uma variável no algoritmo sem identificar seu peso discriminatório e sua restrição legal configura violação de privacidade e compromete a validade jurídica de todo o modelo.

Para mitigar esse risco, propomos uma ferramenta de triagem que atua como um filtro preventivo de governança. O sistema opera como um classificador automatizado, identificando a natureza de cada nova variável antes de integrá-la aos modelos de crédito. Utilizamos a taxonomia de Çallı e Coşkun (2021), que divide as variáveis em dez categorias (Socioeconômicas, Demográficas, Valores e Comportamento, Institucionais e Financeiras, Personalidade, Situacionais, Educacionais, Macroeconômicas, Saúde e Alternativas). A esse modelo, introduzimos uma décima primeira categoria denominada Operacional, estruturada para organizar exclusivamente atributos de sistema, como chaves de identificação e registros de tempo (timestamps). Esse processo mapeia os dados, separa as informações de livre uso daquelas restritas e garante a auditoria automática do fluxo.

A construção dessa arquitetura enfrenta dois grandes obstáculos. O primeiro é a escassez de conjuntos de dados rotulados, já que exigências normativas obrigam os bancos a isolarem seus dicionários de dados (MWESIGWA; JACKSON; CSIKSZENTMIHALYI, 2026b). O segundo é a exigência de governança imposta pela complexidade da Inteligência Artificial no setor financeiro (KURSHAN; SHEN; CHEN, 2021). Adotar algoritmos com explicabilidade intrínseca é indispensável para assegurar a supervisão humana

e o cumprimento das diretrizes regulatórias, evitando os riscos associados a modelos opacos (ANOMAH, 2026).

Nossa metodologia propõe um fluxo em três etapas para contornar a escassez de dados e garantir total transparência. Primeiro, geramos um conjunto de treinamento sintético utilizando o modelo Gemma4 31B (TEAM et al., 2024), baseando a amostragem na literatura por meio da técnica de few-shot prompting (BROWN et al., 2020). Na segunda etapa, o modelo BAAI/bge-large-en-v1.5 (XIAO et al., 2024) vetoriza os textos para capturar o contexto semântico, calibrado pela função Batch Hard Triplet Loss (HERMANS; BEYER; LEIBE, 2017) para maximizar a separação entre as categorias. Por fim, classificamos as variáveis com três algoritmos clássicos: Regressão Logística, Random Forest e Máquinas de Vetores de Suporte (SVM) (FABIAN, 2011). Essa escolha arquitetônica evita redes neurais complexas na etapa final, garantindo a rastreabilidade matemática exigida pela legislação de proteção de dados.

Metodologia

2. Metodologia

A arquitetura proposta é um modelo transparente de classificação multiclasse. Ele categoriza descrições textuais em onze macrocategorias. Dez delas derivam da taxonomia de Çallı e Coşkun (2021): Socioeconômica, Demográfica, Valores Atitudes e Comportamento, Institucional e Financeira, Personalidade, Situacional, Educacional, Macroeconômica, Saúde e Alternativa. Nosso estudo expande esse arcabouço teórico. Introduzimos a classe Operacional para lidar diretamente com dados do mundo real.

A Figura 1 ilustra o pipeline do projeto. O fluxo possui três módulos principais: Treinamento com Dados Sintéticos, Consolidação dos Dados de Teste, e Inferência e Avaliação. Para garantir transparência e pesquisa reprodutível, o código-fonte completo e as instruções de execução estão públicos no GitHub: https://github.com/ThisAnonymousUser/feature-distribution.

2.1. Geração de Dados Sintéticos

O treinamento desta arquitetura esbarra na escassez de dados abertos, rotulados e balanceados. Para viabilizar a prova de conceito, criamos um conjunto de dados sintético. Essa etapa supre a falta de dados empíricos. Ela também garante o treinamento robusto do modelo preditivo.

Sintetizamos as variáveis com o modelo Gemma4 31B em quantização de 8 bits (TEAM et al., 2024)]. Fixamos a temperatura em 0,4. Isso estabiliza a semântica e reduz alucinações. Usamos few-shot prompting para o contexto (BROWN et

al., 2020). Isso ancora a amostragem diretamente nos conceitos teóricos de Çallı e Coşkun (2021).

As dez macrocategorias de Çallı e Coşkun (2021) cobrem apenas a dimensão teórica. Elas ignoram a estrutura técnica dos bancos de dados reais. Para classificar as variáveis em ambientes de produção, introduzimos uma décima primeira macrocategoria: a Operacional. Ela agrupa exclusivamente metadados de sistema. Isso inclui identificadores primários, hashes, timestamps e flags de processamento.

Submetemos os dados gerados a uma normalização de texto. Convertemos as sentenças para letras minúsculas e removemos os caracteres especiais. Isso gerou um conjunto de treinamento limpo e rotulado sinteticamente para o aprendizado supervisionado.

Avaliamos a qualidade do espaço vetorial resultante e a coesão entre as onze classes. Utilizamos quatro métricas de agrupamento: Silhouette Score, Índice Davies-Bouldin (DB), Índice Calinski-Harabasz (CH) e Similaridade do Cosseno Intraclasse. O Silhouette Score (ROUSSEEUW, 1987) quantificou a proximidade dos elementos de uma mesma classe em relação à vizinha mais próxima. O Índice DB (DAVIES; BOULDIN, 1979) calculou a razão entre a dispersão interna dos agrupamentos e a separação entre eles. O Índice CH (CALIŃSKI; HARABASZ, 1974) examinou a variância inter e intragrupos para medir a precisão da separação. A Similaridade do Cosseno Intraclasse (SALTON; WONG; YANG, 1975) atestou a consistência semântica dos vetores de texto de uma mesma categoria.

2.2. Modelo de Embedding

A arquitetura processa as descrições textuais das variáveis. Ela não depende de rótulos isolados. Isso garante a captura do contexto semântico necessário para uma classificação precisa. Por exemplo, se uma variável traz o rótulo "idade", o modelo recebe a descrição completa: "idade do indivíduo em anos".

O modelo fundacional BAAI/bge-large-en-v1.5 vetoriza essas entradas (XIAO et al., 2024). O ajuste fino (fine-tuning) alinha o espaço vetorial às onze macrocategorias. Essa etapa utiliza a função Batch Hard Triplet Loss (HERMANS; BEYER; LEIBE, 2017). A otimização penaliza as trincas mais difíceis do lote. A abordagem maximiza a separação interclasses e reforça a coesão intraclasse. O resultado é uma matriz de embeddings altamente calibrada.

2.3. Treinamento do Classificador

Na fase de modelagem supervisionada, treinamos os modelos no conjunto de dados sintéticos. Os dados foram previamente rotulados em onze macrocategorias. Utilizamos características preditivas derivadas da matriz de embeddings calibrada. Aplicamos três algoritmos clássicos de classificação: Regressão Logística, Random Forest e Máquinas de Vetores de Suporte (SVM)

(FABIAN, 2011). Escolhemos métodos consolidados em vez de redes complexas e opacas. Isso garante a explicabilidade exigida pela conformidade financeira.

2.4. Avaliação de Desempenho

A fase de testes avaliou a capacidade de generalização da arquitetura. Utilizamos três conjuntos de dados reais: Adult (BECKER; KOHAVI, 1996), Prosper Loan Dataset (Prosper Marketplace, 2014) e Mortgage Data (HMDA 2017) (Federal Financial Institutions Examination Council, 2025). A escolha dessas bases foi intencional. Elas fornecem os rótulos-alvo e descrições textuais detalhadas das variáveis.

As bases originais não trazem rótulos nativos para a taxonomia expandida. Por isso, aplicamos um protocolo de anotação híbrida. Primeiro, os modelos Gemma4 31B (TEAM et al., 2024) e Qwen 2.5 32B (YANG et al., 2024) classificaram as variáveis de forma autônoma. Depois, um comitê de três validadores humanos analisou as divergências. O rótulo de referência (ground-truth) final foi decidido por maioria de votos.

Resultados e Discussões

3. Resultados e Discussão

3.1. Avaliação dos Dados Sintéticos

A avaliação quantitativa do espaço de embeddings sintético revelou alta coesão interna. Os resultados também mostram uma nítida separação entre as macrocategorias analisadas. A Tabela 1 detalha as métricas de agrupamento. Elas confirmam a formação de grupos compactos e bem delimitados. A Similaridade de Cosseno Intraclasse elevada indica um alinhamento semântico robusto entre vetores da mesma classe. Isso valida a qualidade da geração dos dados.

3.2. Análise Visual da Projeção de Dados

Projetamos os vetores dos dados sintéticos de treinamento e empíricos de teste em um espaço bidimensional. O objetivo foi avaliar a capacidade de o modelo de embeddings alinhar textos de diferentes domínios. Utilizamos o algoritmo UMAP para a redução de dimensionalidade, como mostra a Figura 2.

As projeções de foco local e global revelam onze clusters bem definidos. Eles correspondem diretamente às macrocategorias teóricas. As amostras sintéticas (círculos) e as reais (losangos com contorno preto) ocupam exatamente as mesmas regiões em cada agrupamento. Essa forte sobreposição indica que os

textos de ambas as fontes possuem representações semânticas semelhantes no espaço vetorial. Nenhum subgrupo se isolou devido à origem dos dados.

3.3. Desempenho Preditivo dos Modelos

A Tabela 2 mostra a acurácia global dos classificadores nos conjuntos de dados avaliados.

A Regressão Logística obteve o melhor desempenho geral. Este modelo também apresentou a maior consistência preditiva nas avaliações individuais. Os três modelos registraram acurácia idêntica nas bases Adult e Prosper Loan. O conjunto HMDA 2017 impôs o cenário de generalização mais complexo. Esta base resultou na menor acurácia entre todos os algoritmos testados.

A Figura 3 detalha o comportamento do melhor modelo no conjunto de teste completo por categoria. O classificador alcançou resultados equilibrados na classe Macroeconômica, com precisão e recall de 0,83 em 6 instâncias. A classe Demográfica obteve precisão de 0,73 e recall de 0,89 em 9 instâncias. A categoria Operacional exibiu resultados mais moderados em suas 17 instâncias. O classificador registrou precisão de 0,67, recall de 0,59 e F1-score de 0,62.

Na classe majoritária, Institucional e Financeira (64 instâncias), o algoritmo alcançou revocação de 0,97. A precisão atingiu apenas 0,74. Isso indica uma concentração de falsos positivos. A categoria Socioeconômica obteve precisão de 0,78. São 18 instâncias com baixa revocação (0,39). Isso evidencia uma alta taxa de falsos negativos.

O classificador falhou nas classes Alternativa (10 instâncias) e Valores, Atitudes e Comportamento (2 instâncias). Ambas registraram precisão e revocação nulas. Geraram apenas falsos negativos. As categorias Situacional, Personalidade, Educacional e Saúde estavam ausentes do conjunto de teste. O modelo não fez previsões para as três últimas. Ele classificou um caso Institucional e Financeiro como Situacional. Isso gerou um falso positivo isolado.

3.4. Discussões

O modelo de embedding agrupou corretamente as onze macrocategorias. O Silhouette Score de 0,768 e o Índice Davies-Bouldin de 0,344 evidenciam classes bem separadas. A dispersão interna é baixa. O Índice Calinski-Harabasz de 1742,9 reforça a distância geométrica entre os clusters.

A Similaridade de Cosseno Intraclasse alcançou 0,946. O modelo Gemma4 31B gerou amostras com pouca variação de vocabulário. A ausência de ruído natural reduz a diversidade semântica no treinamento. A arquitetura perde capacidade de generalização em contextos específicos. Isso explica a queda de desempenho no conjunto HMDA. As acurácias variaram entre 0,5667 e 0,6333. O motivo é o rigoroso padrão governamental e regulatório dos dados. O cenário difere das descrições financeiras amplas do Prosper Loan Dataset. Esse

conjunto obteve o melhor desempenho da avaliação (0,7778). O conjunto Adult apresentou generalização satisfatória (0,6667).

A avaliação no conjunto de teste indicou um desempenho satisfatório da arquitetura. Os resultados evidenciaram um alinhamento técnico entre os classificadores. A Regressão Logística alcançou acurácia de 0,7302. Ela foi seguida de perto pelo SVM Linear (0,7222) e pelo Random Forest (0,7143). Essa equivalência reflete fronteiras de decisão estáveis na matriz de embeddings. Isso permite que algoritmos lineares simples atinjam a eficiência de modelos complexos.

A Regressão Logística obteve o melhor desempenho nesta etapa. A análise de erros foca em sua matriz de confusão (Figura 3). A avaliação revela desequilíbrio na separação das categorias. O modelo linear tende a alocar casos ambíguos nas classes dominantes. A classe Institucional e Financeira atraiu o maior volume de erros. Ela registrou 22 falsos positivos. Seu critério de classificação tornou-se amplo demais no espaço vetorial. Já a classe Socioeconômica apresentou uma fronteira excessivamente rígida. Ela teve 11 falsos negativos contra 7 predições corretas, com recall de 0,39. Esse desequilíbrio afeta a segmentação das categorias residuais. O impacto fica evidente pela total ausência de acertos na classe Alternativa (10 falsos negativos em 10 amostras). O erro se repete na classe Valores, Atitudes e Comportamental (2 falsos negativos em 2 instâncias).

Contudo, essa dificuldade com clusters residuais ou altamente abstratos não invalida a arquitetura. O modelo demonstrou forte generalização e sucesso na recuperação das demais categorias. As classes Macroeconômica e Demográfica apresentaram alto desempenho preditivo e fronteiras bem definidas. A primeira atingiu uma precisão de 0,83, classificando corretamente 5 das 6 instâncias disponíveis com apenas 1 omissão. Já a segunda foi recuperada com alta eficiência, registrando um recall de 0,89. Além desses resultados, o desempenho na recém-introduzida classe Operacional reforçou a viabilidade prática da solução, contabilizando 10 predições corretas e um F1-Score de 0,62.

Conclusão

4. Conclusões

Por fim, a avaliação global atesta a viabilidade do método. A principal vantagem da arquitetura proposta reside em sua estrutura modular. Ela permite substituir de forma isolada o gerador de dados sintéticos ou o modelo de embedding sem comprometer o restante do pipeline. Essa flexibilidade facilita experimentos futuros com novos encoders e a injeção de ruído direcionado para aumentar a diversidade nas categorias semânticas de nível superior, buscando uma acurácia acima de 0,9.

Além disso, ao delegar a predição final para algoritmos clássicos, o modelo treinado opera de maneira determinística. Diferentemente da classificação direta por modelos generativos, que carecem de interpretabilidade e exigem análises linguísticas e de gradiente post hoc complexas para explicar suas decisões (RIBEIRO et al., 2024), essa abordagem garante a rastreabilidade matemática das fronteiras de decisão. Isso entrega o exato nível de auditabilidade exigido para uma implantação segura em ambientes de produção. O método alinha-se diretamente aos preceitos éticos e legais necessários para uma governança de dados robusta, responsabilização algorítmica e sistemas confiáveis em setores altamente regulamentados (ALVARENGA et al., 2026).

Referências

ALVARENGA, L. F. et al. Legal and ethical considerations for translating federated learning into cross-border healthcare innovation. IEEE Journal of Biomedical and Health Informatics, n. 1, 2026.

ANOMAH, S. Bridging the interpretability gap: integrating explainable ai into public sector audit systems in emerging economies. Accounting Research Journal, Emerald Publishing Limited, p. 1–27, 2026.

BECKER, B.; KOHAVI, R. Adult. 1996. UCI Machine Learning Repository. DOI: https://doi.org/10.24432/C5XW20.

BJöRKEGREN, D.; GRISSEN, D. The Potential of Digital Credit to Bank the Poor. AEA Papers and Proceedings, v. 108, p. 68–71, maio 2018. ISSN 2574-0768, 2574-0776. Disponível em: <https://www.aeaweb.org/articles?id=10.1257/pandp.20181032>.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). 2018. Acesso em: Jun 21, 2026. Disponível em: <https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm>.

BROWN, T. et al. Language models are few-shot learners. Advances in neural information processing systems, v. 33, p. 1877–1901, 2020.

CALIŃSKI, T.; HARABASZ, J. A dendrite method for cluster analysis. Communications in Statistics-theory and Methods, Taylor & Francis, v. 3, n. 1, p. 1–27, 1974.

ÇALLI, B. A.; COŞKUN, E. A longitudinal systematic review of credit risk assessment and credit default predictors. Sage Open, SAGE Publications Sage CA: Los Angeles, CA, v. 11, n. 4, 2021.

DAVIES, D. L.; BOULDIN, D. W. A cluster separation measure. IEEE transactions on pattern analysis and machine intelligence, Ieee, n. 2, p. 224–227, 1979.

European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). 2016. Official Journal of the European Union. Disponível em: <https://eur-lex.europa.eu/eli/reg/2016/679/oj>.

FABIAN, P. Scikit-learn: Machine learning in python. Journal of machine learning research 12, p. 2825, 2011.

Federal Financial Institutions Examination Council. Home Mortgage Disclosure Act (HMDA) Data. 2025. FFIEC / CFPB. Disponível em: <https://ffiec.cfpb.gov/>.

GUO, G. et al. From Footprint to Evidence: An Exploratory Study of Mining Social Data for Credit Scoring. ACM Trans. Web, v. 10, n. 4, p. 22:1–22:38, 2016. ISSN 1559-1131. Disponível em: <https://dl.acm.org/doi/10.1145/2996465>.

HERMANS, A.; BEYER, L.; LEIBE, B. In defense of the triplet loss for person re-identification. arXiv preprint arXiv:1703.07737, 2017.

HURLEY, M.; ADEBAYO, J. Credit scoring in the era of big data. Yale Journal of Law and Technology, 2017.

KURSHAN, E.; SHEN, H.; CHEN, J. Towards self-regulating ai: challenges and opportunities of ai model governance in financial services. In: Proceedings of the First ACM International Conference on AI in Finance. New York, NY, USA: Association for Computing Machinery, 2021. (ICAIF '20). ISBN 9781450375849. Disponível em: <https://dl.acm.org/doi/10.1145/3383455.3422564>.

MWESIGWA, D.; JACKSON, S. J.; CSIKSZENTMIHALYI, C. Risk, Data, Alignment: Making Credit Scoring Work in Kenya. arXiv, 2026. Version Number: 1. Disponível em: <https://arxiv.org/abs/2602.01824>.

MWESIGWA, D.; JACKSON, S. J.; CSIKSZENTMIHALYI, C. Risk, data, alignment: Making credit scoring work in kenya. In: Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems. New York, NY, USA: Association for Computing Machinery, 2026. (CHI '26). ISBN 9798400722783. Disponível em: <https://dl.acm.org/doi/10.1145/3772318.3790924>.

Prosper Marketplace. Prosper Loan Data. 2014. Kaggle. Disponível em: <https://www.kaggle.com/datasets/yousuf28/prosper-loan>.

RIBEIRO, M. et al. A methodology for explainable large language models with integrated gradients and linguistic analysis in text classification. arXiv, n. arXiv:2410.00250, 2024. Disponível em: <https://arxiv.org/abs/2410.00250>.

ROUSSEEUW, P. J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of computational and applied mathematics, Elsevier, v. 20, p. 53–65, 1987.

SALTON, G.; WONG, A.; YANG, C.-S. A vector space model for automatic indexing. Communications of the ACM, ACM New York, NY, USA, v. 18, n. 11, p. 613–620, 1975.

SCHWARCZ, D.; PRINCE, A. E. R. Proxy Discrimination in the Age of Artificial Intelligence and Big Data.

TEAM, G. et al. Gemma: Open models based on gemini research and technology. arXiv preprint arXiv:2403.08295, 2024.

TOUNSI, Y.; HASSOUNI, L.; ANOUN, H. Credit scoring in the age of Big Data – A State-of-the-Art. v. 15, n. 7, 2017.

XIAO, S. et al. C-pack: Packed resources for general chinese embeddings. In: Proceedings of the 47th international ACM SIGIR conference on research and development in information retrieval. [S.l.: s.n.], 2024. p. 641–649.

YANG, A. et al. Qwen2.5 technical report. arXiv preprint arXiv:2412.15115, 2024.

Anexos



Figura 1 - Diagrama da arquitetura proposta para classificação multiclasse transparente. O esquema detalha o fluxo de treinamento com dados sintéticos, o processo de consolidação dos dados empíricos de teste e a etapa final de avaliação.



Figura 2 - Visualização bidimensional via UMAP do alinhamento entre os domínios sintético (treino) e empírico (teste).



Figura 3 - Mapa de calor com a distribuição absoluta de verdadeiros positivos, falsos negativos (omissões) e falsos positivos (alarmes falsos) em todo o conjunto de teste.

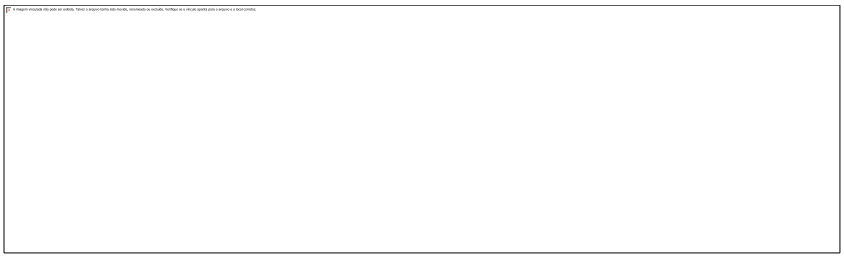


Tabela 1 - Métricas de avaliação do espaço de embeddings gerado a partir do conjunto de dados sintéticos.

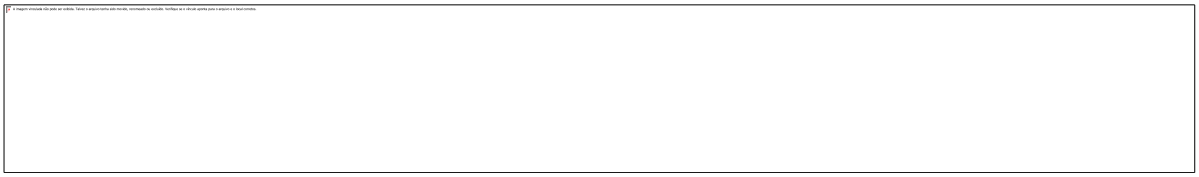


Tabela 2 - Acurácia global dos modelos de classificação nos conjuntos de teste.

CÓDIGO: ET1471

AUTOR: FELIPE SAMUEL BATISTA MONTEIRO

COAUTOR: IAN SILVA CRUZ DE MEDEIROS

ORIENTADOR: DANIEL LOPES MARTINS

TÍTULO: Monitoramento visual com ESP32-CAM via rede Wi-Fi: Uma introdução à visão computacional embarcada

Resumo

Este trabalho apresenta o desenvolvimento de um sistema de monitoramento visual utilizando o módulo ESP32-S3 CAM para captura e transmissão de imagens em tempo real por meio de uma rede Wi-Fi criada pelo próprio microcontrolador. O sistema foi desenvolvido como parte das atividades de iniciação científica, tendo como objetivo desenvolver uma infraestrutura de baixo custo para aquisição e transmissão de imagens que possa ser utilizada em aplicações de monitoramento e visão computacional. Para a implementação, utilizamos a biblioteca esp_camera para configuração e captura das imagens, a biblioteca WiFi para criação da rede sem fio e as bibliotecas ESPAsyncWebServer e AsyncTCP para implementação do servidor web assíncrono. Os quadros capturados são comprimidos em formato JPEG e transmitidos por meio de uma conexão persistente utilizando WebSocket. Durante os testes, verificamos que o sistema foi capaz de capturar e transmitir continuamente as imagens para um navegador conectado à rede criada pelo ESP32-S3 CAM. Como aplicação, o sistema foi relacionado ao monitoramento visual de controladores lógicos programáveis, no qual as imagens dos LEDs podem ser utilizadas como entrada para modelos de visão computacional. Os resultados demonstram a viabilidade do ESP32-S3 CAM como plataforma de baixo custo para aquisição e transmissão de imagens em aplicações de monitoramento.

Palavras-chave: ESP-32. Monitoramento visual. WebSocket. Visão computacional.

TITLE: Visual monitoring with ESP32-CAM via Wi-Fi network: An introduction to embedded computer vision.

Abstract

This work presents the development of a visual monitoring system using the ESP32-S3 CAM module to capture and transmit images in real time through a Wi-Fi network

created by the microcontroller itself. The system was developed as part of undergraduate research activities, aiming to establish a low-cost infrastructure for image acquisition and transmission that can be used in monitoring and computer vision applications. The implementation uses the esp_camera library for camera configuration and image capture, the WiFi library to create the wireless network, and the ESPAsyncWebServer and AsyncTCP libraries to implement the asynchronous web server. Captured frames are compressed in JPEG format and transmitted through a persistent WebSocket connection. During the tests, we verified that the system was able to continuously capture and transmit images to a browser connected to the network created by the ESP32-S3 CAM. As an application, the system was related to visual monitoring of programmable logic controllers, in which images of LEDs can be used as input for computer vision models. The results demonstrate the feasibility of the ESP32-S3 CAM as a low-cost platform for image acquisition and transmission in monitoring applications.

Keywords: ESP-32. Visual monitoring. WebSocket. Computer vision.

Introdução

A transmissão de imagens em tempo real usando microcontroladores abre diversas possibilidades de aplicação nas áreas de Internet das Coisas (IoT), automação, monitoramento remoto e visão computacional. Com a crescente evolução dos microcontroladores, tornou-se possível integrar, em pequenos dispositivos de baixo custo, recursos que anteriormente dependiam de computadores ou equipamentos dedicados, como Wi-Fi, bluetooth, captura de imagens e processamento de dados.

Nesse contexto, os módulos da família ESP32 aparecem como alternativas relevantes para o desenvolvimento de projetos embarcados devido à integração entre a capacidade de processamento, memória, rede sem fios e interfaces para periféricos. Entre esses dispositivos encontra-se o ESP32-S3 CAM, que foi utilizado neste projeto como dispositivo para captura e transmissão de imagens.

O objetivo deste projeto é desenvolver e avaliar um sistema de monitoramento visual capaz de capturar imagens utilizando uma câmera acoplada ao ESP32-S3 e transmiti-las ao vivo por meio de uma rede Wi-Fi. Para isso, foi implementado um servidor web embarcado, para poder disponibilizar o fluxo de imagens ao usuário conectado por meio de um navegador.

Esse projeto foi desenvolvido como parte das atividades de iniciação científica e também como uma etapa de preparação para aplicações posteriores de visão computacional embarcada. Nesse sentido, o sistema de transmissão constitui uma infraestrutura que pode ser utilizada por aplicações capazes de analisar as imagens capturadas, seja diretamente no microcontrolador ou em um dispositivo externo conectado à mesma rede.

Uma das aplicações desenvolvidas em conjunto com este trabalho consiste no monitoramento visual de controladores lógicos programáveis (CLPs). Nesse caso, um módulo ESP32-S3 CAM foi posicionado diante do painel do equipamento para capturar

os LEDs indicadores de seu estado. As imagens obtidas foram analisadas por um modelo de visão computacional, permitindo identificar automaticamente diferentes condições de funcionamento do CLP.

Isso demonstra que o sistema desenvolvido não se limita à transmissão de imagens para visualização humana, mas também pode atuar como uma camada de aquisição de dados para aplicações de inteligência artificial. Dessa forma, o sistema pode ser utilizado em diferentes áreas e casos de uso.

OBJETIVO GERAL

Desenvolver um sistema de monitoramento visual de baixo custo utilizando o ESP32-S3 CAM para captura e transmissão de imagens por meio de uma rede Wi-Fi, estabelecendo uma infraestrutura que possa ser utilizada em aplicações de visão computacional e inteligência artificial embarcada.

OBJETIVOS ESPECÍFICOS

- Estudar os principais conceitos relacionados a sistemas embarcados, Internet das Coisas, redes sem fio, aquisição de imagens e visão computacional;
- Configurar a câmera utilizada pelo módulo ESP32-S3 CAM;
- Implementar um servidor web embarcado para disponibilização das imagens;
- Estabelecer uma rede Wi-Fi por meio do próprio dispositivo para comunicação com o cliente;
- Implementar a transmissão contínua dos frames capturados pela câmera;
- Utilizar WebSocket para estabelecer comunicação persistente entre o dispositivo e os clientes conectados;
- Organizar o código de forma modular, facilitando sua manutenção e utilização em aplicações futuras;
- Avaliar o funcionamento e a estabilidade do sistema em uma rede Wi-Fi criada pelo próprio ESP;
- Investigar a utilização do sistema como infraestrutura de aquisição de imagens para aplicações de visão computacional;

Metodologia

O desenvolvimento do projeto foi realizado utilizando o ESP32-S3 CAM, uma rede Wi-Fi e um servidor web, também desenvolvido no projeto, responsável pela disponibilização do fluxo de imagens. O desenvolvimento foi dividido em etapas relacionadas à configuração do hardware, programação do sistema, comunicação de rede, transmissão dos frames e avaliação da aplicação.

ESP32-S3 CAM

O ESP32-S3 CAM foi utilizado como ferramenta principal do projeto por integrar recursos de processamento, memória, conectividade Wi-Fi e interface para câmera em um único dispositivo, reduzindo a necessidade de componentes adicionais e permitindo uma solução compacta para captura e transmissão de imagens (ESPRESSIF SYSTEMS, 2026a).

Embora a proposta inicial tenha considerado o uso de módulos ESP32-CAM, os experimentos apresentados foram realizados com o ESP32-S3 CAM. A escolha considerou principalmente a disponibilidade de memória. O módulo utilizado, ESP32-S3-N16R8, possui 16 MB de memória Flash e 8 MB de PSRAM, enquanto o ESP-WROOM-32D, considerado como alternativa, possui 4 MB de Flash e 520 KB de SRAM interna. A maior disponibilidade de memória é relevante para a aplicação, pois a captura e transmissão contínua de imagens requerem o armazenamento temporário dos quadros e o processamento dos dados antes de sua transmissão.

CAPTURA DAS IMAGENS

A captura das imagens foi realizada por meio do sensor de câmera conectado ao ESP32-S3 CAM. Para sua configuração e controle, foi utilizada a biblioteca esp_camera, conforme a documentação da Espressif Systems (2026b). Os principais parâmetros empregados são apresentados na Tabela 1.

O sistema foi estruturado segundo o modelo produtor-consumidor, utilizando duas tarefas independentes. A tarefa responsável pela câmera atua como produtora, realizando a captura contínua dos quadros e armazenando-os em buffer de memória. A tarefa associada à interface web atua como consumidora, retirando os quadros disponíveis do buffer para transmissão ao dispositivo conectado.

Devido à orientação do ESP32-S3 CAM no suporte utilizado, foi realizada a inversão horizontal das imagens. Após o processamento de cada quadro, foi utilizado um intervalo de aproximadamente 10 ms entre as iterações da tarefa de captura, permitindo a alternância da execução entre as tarefas do sistema.

COMUNICAÇÃO WI-FI

A comunicação foi realizada por meio de uma rede Wi-Fi gerada pelo próprio microcontrolador, configurada como Ponto de Acesso (Access Point). Para essa finalidade, foi utilizada a biblioteca WiFi, conforme a documentação da Espressif Systems (2026b).

Após a criação da rede, um dispositivo cliente poderia conectar-se ao ponto de acesso denominado meu-esp32 e acessar, por meio de um navegador, o servidor disponibilizado pelo ESP. Dessa forma, o fluxo de imagens poderia ser visualizado sem conexão física entre o ESP32-S3 CAM e o computador utilizado para visualização.

SERVIDOR WEB ASSÍNCRONO

Para exibir as imagens, foi implementado um servidor web utilizando as bibliotecas ESPAsyncWebServer e AsyncTCP, operando na porta 80. O servidor era iniciado assim que o ESP recebia energia e ativava o ponto de acesso Wi-Fi.

Foi disponibilizada uma rota denominada /stream, responsável por fornecer os dados das imagens capturadas ao cliente, podendo ser acessada e utilizada por outros programas, como apresentado na Figura 1. A rota não possui elementos gráficos, mantendo sua implementação mais leve.

A abordagem assíncrona foi utilizada para reduzir bloqueios durante a execução do microcontrolador, permitindo que as operações de comunicação fossem realizadas enquanto o sistema continuava executando a captura das imagens.

TRANSMISSÃO UTILIZANDO WEBSOCKET

Fette e Melnikov (2011) apresentam o WebSocket como um protocolo destinado à comunicação persistente entre cliente e servidor, permitindo a troca contínua de dados. Essa característica foi utilizada no projeto para realizar a transmissão contínua dos quadros capturados.

Diferentemente de requisições HTTP independentes, o WebSocket mantém uma conexão persistente entre cliente e servidor, permitindo o envio de novos quadros sem a necessidade de estabelecer uma nova requisição para cada imagem (FETTE; MELNIKOV, 2011).

Os quadros capturados pela câmera são transmitidos pela conexão WebSocket como mensagens binárias codificadas em JPEG. No navegador, os dados recebidos são decodificados para atualização da imagem apresentada ao usuário.

APLICAÇÃO NO MONITORAMENTO DE CLP

Como aplicação do sistema desenvolvido, foi considerada sua utilização no monitoramento visual de controladores lógicos programáveis. Nessa aplicação, o ESP32-S3 CAM, com auxílio de um suporte, pode ser instalado em frente ao painel do CLP para capturar imagens dos LEDs.

Howard et al. (2017) apresentam a arquitetura MobileNet, desenvolvida para aplicações de visão computacional com restrições de recursos. No trabalho desenvolvido em conjunto, foi utilizado um modelo baseado nessa arquitetura para classificar individualmente os LEDs do CLP como acesos ou apagados.

Para avaliar a aplicação, o ESP32-S3 CAM foi posicionado em frente ao painel do CLP, mantendo os LEDs no campo de visão da câmera. A transmissão foi acompanhada por meio do navegador de internet, verificando-se a capacidade do sistema de fornecer imagens suficientemente claras para a identificação visual dos estados dos LEDs.

Resultados e Discussões

Os testes realizados demonstraram que o ESP32-S3 CAM foi capaz de realizar a captura contínua de imagens e sua transmissão por meio de uma rede Wi-Fi para um navegador web. Durante os testes, o dispositivo conseguiu estabelecer a conexão com a rede, iniciar o servidor web e manter a comunicação com o cliente conectado.

A transmissão dos frames ocorreu utilizando imagens JPEG enviadas em formato binário por meio do WebSocket. A utilização de uma conexão WebSocket mostrou-se adequada para a aplicação, uma vez que permitiu o envio contínuo dos quadros sem a necessidade de realizar novas requisições HTTP para cada imagem. Como consequência, foi possível obter uma atualização contínua da imagem apresentada no navegador.

Durante os testes, a transmissão foi limitada a 25 quadros por segundo (FPS), com intervalo mínimo de 40 ms entre quadros. A taxa efetiva, no entanto, pode ser menor devido a atrasos na rede Wi-Fi, à sobrecarga de processamento para captura e compressão das imagens e à execução de outras tarefas no microcontrolador.

No exercício envolvendo a inferência do modelo de inteligência artificial, a taxa de transmissão foi reduzida para 10 FPS, utilizando-se um intervalo de 100 ms entre quadros consecutivos. Essa configuração teve como objetivo reduzir a competição por recursos computacionais entre as tarefas de captura e transmissão das imagens, a inferência dos modelos e a disponibilização, por meio do WebSocket, das imagens e dos resultados obtidos. Durante os testes, essa redução contribuiu para minimizar a ocorrência de perdas de quadros durante a execução simultânea dessas tarefas.

A utilização de JPEG contribuiu para reduzir o volume de dados transmitidos. Entretanto, existe uma relação entre a qualidade da imagem, a resolução, o tamanho dos quadros e o desempenho da transmissão. Resoluções maiores produzem imagens com maior quantidade de informações, mas também aumentam a necessidade de processamento, armazenamento temporário e transmissão.

A disponibilidade da memória PSRAM no ESP se mostrou importante para a aplicação. Essa memória adicional permitiu trabalhar de maneira mais confortável com o buffer necessário à captura e transmissão das imagens, contribuindo para uma maior estabilidade do sistema.

AVALIAÇÃO FUNCIONAL

Os testes validaram as principais funcionalidades do sistema: a captura de imagens pela câmera do ESP32 e sua exibição no navegador via servidor web, sem falhas de comunicação ou perda de quadros, conforme a captura ilustrado na Figura 1.

Os resultados indicam que o sistema atende ao principal objetivo de suprir a necessidade de uma solução embarcada em um ESP para captura e transmissão de imagens em tempo real dentro da rede Wi-Fi criada pelo próprio microcontrolador.

Entretanto, os testes realizados foram majoritariamente funcionais. Portanto, não foi realizada neste trabalho uma avaliação aprofundada de métricas como taxa de quadros por segundo, latência média, limite de clientes na rede, consumo energético ou desempenho sob diferentes condições de congestionamento da rede. Essas características podem ser investigadas em trabalhos futuros.

RELAÇÃO COM A APLICAÇÃO DE VISÃO COMPUTACIONAL

A aplicação no monitoramento de CLPs demonstra uma possibilidade concreta para utilização do sistema. Nesse cenário, a câmera não precisa ser utilizada apenas para visualização por um operador, podendo fornecer imagens para uma etapa posterior de análise automática.

O trabalho desenvolvido em conjunto mostrou que imagens dos LEDs de um CLP podem ser utilizadas para treinar um modelo de visão computacional capaz de classificá-los como acesos ou apagados. O estudo utilizou uma MobileNet reduzida, adequada à proposta de execução em dispositivos com recursos limitados.

Nesse contexto, o sistema de transmissão desenvolvido neste projeto constitui uma etapa complementar importante, pois fornece uma forma de adquirir e disponibilizar as imagens necessárias para esse tipo de aplicação.

Conclusão

O projeto demonstrou a viabilidade da utilização do ESP32-S3 CAM como uma plataforma de baixo custo para aquisição e transmissão de imagens em tempo real por meio de uma rede Wi-Fi.

A implementação realizada integrou a captura de imagens, conexão à rede, servidor web assíncrono e comunicação persistente por WebSocket. Os testes realizados no Wi-Fi do ESP demonstraram funcionamento de comunicação estável e ininterrupta do sistema, permitindo que os quadros capturados fossem transmitidos em formato JPEG e visualizados continuamente em um navegador web.

Além de atender ao objetivo inicial de desenvolver uma solução funcional de monitoramento visual, o projeto estabeleceu uma infraestrutura que pode ser utilizada em aplicações mais avançadas de visão computacional embarcada. A utilização do sistema no contexto de monitoramento de CLPs demonstra uma aplicação prática dessa infraestrutura, permitindo que imagens dos indicadores do equipamento sejam adquiridas remotamente e posteriormente analisadas por modelos de inteligência artificial.

Dessa forma, o trabalho contribui para a construção de uma plataforma experimental de baixo custo para estudos envolvendo sistemas embarcados, comunicação sem fio, aquisição de imagens e visão computacional. Os resultados também demonstram que

dispositivos da família ESP32 podem ser utilizados como elementos de aquisição e comunicação em sistemas de monitoramento, servindo como base para futuras pesquisas envolvendo inteligência artificial embarcada e edge computing.

Referências

- ESPRESSIF SYSTEMS. ESP32-S3 Series Datasheet. Disponível em: https://www.espressif.com/en/support/documents/technical-documents. Acesso em: 28 ago. 2026a.
- ESPRESSIF SYSTEMS. Arduino ESP32 Documentation. Disponível em: https://docs.espressif.com/projects/arduino-esp32/en/latest/. Acesso em: 28 ago. 2026b.
- FETTE, I.; MELNIKOV, A. The WebSocket Protocol. RFC 6455, Internet Engineering Task Force, 2011. Disponível em: https://www.rfc-editor.org/rfc/rfc6455. Acesso em: 28 ago. 2026.
- HOWARD, A. G. et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv preprint arXiv:1704.04861, 2017. Disponível em: https://arxiv.org/abs/1704.04861. Acesso em: 28 ago. 2026.

Anexos

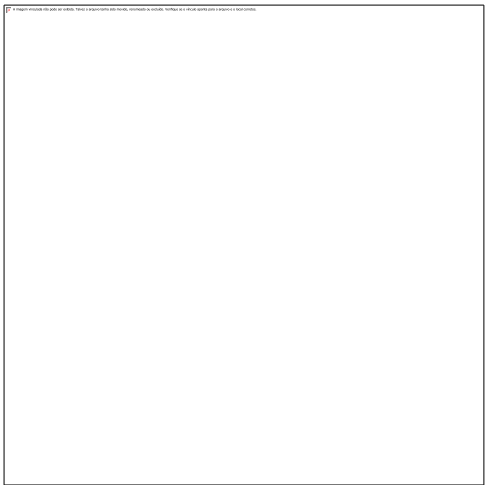


Figura 1

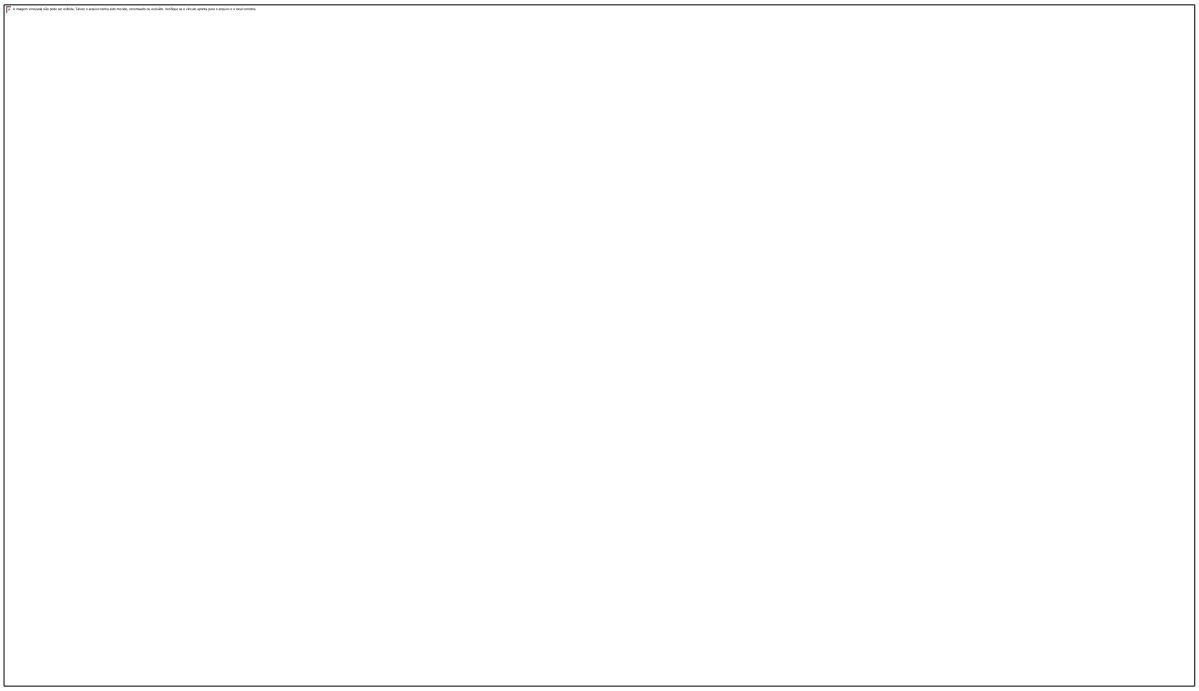


Tabela 1

CÓDIGO: ET1515

AUTOR: IAN SILVA CRUZ DE MEDEIROS

ORIENTADOR: DANIEL LOPES MARTINS

TÍTULO: Aplicação de visão computacional embarcada com ESP32-CAM para monitoramento de processos industriais

Resumo

O objetivo do projeto é o desenvolver uma aplicação de inteligência artificial (IA) embarcada em ESP32 para solução de um problema industrial, em particular, a identificação de alertas de falha em um Controlador Lógico Programável (CLP). Uma câmera acoplada ao ESP32 captura uma imagem que é passada ao modelo, ele identifica a configuração dos LEDs (Light Emitting Diode; Diodo Emissor de Luz) do painel do CLP, e usa essa informação para detectar funcionamento normal ou falha. A metodologia consiste em criar um conjunto de imagens do CLP, usá-las para treinar um modelo pelo Google Colab, convertê-lo no formato TensorFlow Lite Micro para embarque no ESP32, aplicando quantização para redução de tamanho, avaliar seu desempenho por F1-Score fora do dispositivo, colocá-lo no ESP32, medir o tempo de inferência embarcado e testar com uma imagem do CLP capturada no local. Como resultado, foram selecionados dois modelos quantizados em INT8: um baseado na MobileNetV1 e outro no YOLOv8n. Na avaliação sobre o conjunto de dados de teste, a MobileNetV1 obteve F1-score de 92% com tempo de inferência de 1 segundo no ESP32, enquanto o YOLOv8n alcançou F1-score de 100% com tempo de inferência de 16 segundos. Contudo, as detecções com captura de imagem local perdem precisão por variações de câmera e ambiente. Para superar essas limitações e viabilizar a aplicação em uma indústria real, propõe-se expandir a diversidade do conjunto de imagens, aumentando sua capacidade de generalização e robustez.

Palavras-chave: IA embarcada; Visão Computacional; Detecção de LED; Diagnóstico de CLP

TITLE: Embedded Computer Vision AI Models in ESP-32 for LED Identification and PLC Fault Detection

Abstract

The objective of the project is to develop an artificial intelligence (AI) application embedded in ESP32 to solve an industrial problem, in particular, the identification of failure alerts in a Programmable Logic Controller (PLC). A camera attached to the ESP32

captures an image that is passed to the model, which identifies the configuration of the LEDs (Light Emitting Diode) on the PLC panel, and uses this information to detect normal operation or failure. The methodology consists of creating a set of PLC images, using them to train a model via Google Colab, converting it into TensorFlow Lite Micro format for onboarding on the ESP32, applying quantization for size reduction, evaluating its performance by F1-Score outside the device, placing it on the ESP32, measuring the onboard inference time and testing with an image of the PLC captured on site. As a result, two models quantized in INT8 were selected: one based on MobileNetV1 and the other on YOLOv8n. In the evaluation on the test dataset, MobileNetV1 achieved an F1-score of 92% with an inference time of 1 second on ESP32, while YOLOv8n achieved an F1-score of 100% with an inference time of 16 seconds. However, detections with local image capture lose accuracy due to camera and environmental variations. To overcome these limitations and enable application in a real industry, we propose to expand the diversity of the image set, increasing its generalization capacity and robustness.

Keywords: Embedded AI; Computer Vision; LED Detection; PLC Diagnostics

Introdução

A Indústria 4.0 tem transformado as fábricas por meio da integração de sistemas físicos e digitais, com destaque para tecnologias como a IoT (Internet of Things; Internet das Coisas), a Inteligência Artificial (IA) e a Visão Computacional. A IoT permite a conexão de dispositivos capazes de perceber e atuar no ambiente, enquanto a IA possibilita a criação de modelos que aprendem a resolver problemas e a Visão Computacional permite analisar imagens e identificar características relevantes. Entre os dispositivos de baixo custo que podem ser utilizados nesse contexto está o ESP-32, um microcontrolador que pode integrar recursos como Wi-Fi, Bluetooth e câmera, facilitando a comunicação com diferentes periféricos.

A convergência da área de IoT com o setor industrial cria a possibilidade de aprimorar soluções tradicionais. Em particular, Controladores Lógicos Programáveis (CLPs) são consolidados na automação de máquinas e controle de processos. Se ocorrer falha de comunicação ou não for possível conectar o dispositivo à sala de controle, acompanhar seu estado por meio do painel de LEDs (Light Emitting Diode; Diodo Emissor de Luz) presente no próprio dispositivo é a única alternativa. Nesse caso, para evitar o deslocamento presencial de um técnico – o que poderia gerar atrasos na linha de produção e prejuízos financeiros –, é proposto um sistema de captura de dados utilizando um dispositivo ESP32 com câmera, e técnicas de Visão Computacional, para identificar automaticamente os estados dos LEDs e transmiti-los ao operador, atuando efetivamente como um sistema redundante e reduzindo ociosidade na produção.

O projeto está vinculado ao Laboratório de Tecnologias Industriais (LIND), onde se desenvolvem pesquisas na área de aplicações industriais, unindo Inteligência Artificial e sistemas embarcados. O plano é que em uma aplicação real, o dispositivo possa capturar

imagens do painel do CLP e utilizar um modelo de IA embarcado para identificar a configuração dos LEDs, determinando quais estão ligados ou desligados. A configuração identificada será então comparada com uma tabela de estados previamente definida, na qual cada combinação de LEDs representa uma condição específica do equipamento. A partir dessa comparação, o sistema poderá determinar se o funcionamento está normal ou se existe uma falha, identificando também o tipo de falha correspondente. Essa lógica pode ser implementada diretamente no ESP32, por meio de estruturas condicionais ou outras formas de consulta à tabela, ou a configuração dos LEDs pode ser enviada a outro sistema para realizar essa interpretação. Em caso de falha, o dispositivo poderá comunicar a ocorrência e seu respectivo tipo a outro sistema por meio de protocolos como MQTT, permitindo o acompanhamento e o tratamento da ocorrência. A escolha do protocolo e da forma de processamento deverá considerar as limitações de memória, processamento, distância de comunicação e consumo de energia do dispositivo. Mas destacamos que a aplicação ainda está em fase de desenvolvimento, apresenta-se aqui os testes com IA embarcada realizados, desafios encontrados e indicações de trabalhos futuros. Por questões de tempo, não foram feitos testes mais longos.

Como fundamentação teórica sobre modelos de Visão Computacional, seu uso em dispositivos com recursos limitados, técnicas de processamento de imagens e aplicações industriais semelhantes à proposta deste projeto, destacam-se: HOWARD (2017), que apresentam a MobileNet, uma Rede Neural Convolucional de baixo custo computacional, com operações otimizadas, adequada a Visão Computacional; QUACH (2024), que comparam a MobileNet e a YOLO em uma tarefa de Visão Computacional; RAMAN (2020), que propõem o processamento e a classificação do estado de LEDs de um CLP por meio de imagens, sem o uso de redes neurais; e ZHANG (2024), que utiliza uma Rede Neural Convolucional (CNN) para detectar o estado de um comutador.

Metodologia

O trabalho define-se como uma pesquisa aplicada, de caráter experimental, com objetivo de provar o conceito para um sistema automático de identificação de alertas de falha em equipamentos industriais por meio do monitoramento dos LEDs do painel de um CLP. Seu objeto de estudo são modelos de visão computacional embarcada para a tarefa de leitura e interpretação do estado do LED. A pesquisa foi dividida em dois ciclos metodológicos: criar modelo para tarefa única de classificação e criar modelo para tarefa de classificação mais localização dos LEDs.

O trabalho usou uma estrutura no LIND, composta pelo CLP modelo SIMATIC S7-1200 e um tripé fixo com o ESP32CAM para capturar as imagens usadas no treinamento e teste do modelo. São 9 LEDs observados da região de input e output “A”, conforme Figura 1, com 8 configurações possíveis de acesos e apagados, codificados em “0”s e “1”s, como na Figura 2. Foram utilizadas duas placas de desenvolvimento ESP32: o ESP32CAM,

mais simples, com 4 MB de memória flash e 4 MB de PSRAM, menor e adequado para as capturas; e o ESP32S3 maior, com 16 MB de memória flash e 8MB de PSRAM, hardware otimizado para IA. Ambos com câmeras OV2640 (2 megapixels), mas, via software a resolução é reduzida para 240x240, economizando memória. O treinamento, conversão e quantização dos modelos é feito no Google Colab utilizando TensorFlow/Keras e otimizado para o embarcado via TensorFlow Lite Micro (TFLite RT).

No primeiro ciclo, a tarefa concentrou-se na identificação direta do estado dos LEDs. Capturaram-se 175 imagens RGB 240x240 via ESP32-CAM. Aplicou-se rotações aleatórias de ±10.8 graus, zoom de ±10% e brilhos de ±10%, conforme KERAS (2026), para expandir o conjunto de dados em 700 imagens (628 para treino e 72 para validação). Dos 9 LEDs, 7 foram anotados, ignorando “DI_a0” e “DI_a1” pois não variam de estado.

Optou-se pela MobileNet devido às suas otimizações, ao controle proporcional da largura e do tamanho da rede por meio do parâmetro α e à disponibilidade de versões pré-treinadas, que podem ser ajustadas ao problema específico por fine-tuning.

A arquitetura utiliza a MobileNetV1 pré-treinada no conjunto de dados ImageNet, usual em tarefas de classificação de imagens, e (α = 0,25). A saída da MobileNetV1 é conectada a uma camada GlobalAveragePooling2D (responsável por realizar a média espacial dos mapas de características) + Dense (32 neurônios, ativação ReLU) e 7 saídas Dense (2 neurônios, função de ativação softmax). Em cada saída Dense, os dois neurônios representam, respectivamente, “aceso” e “apagado”, explicitando as classes. O treinamento ocorreu ao longo de 200 épocas, com tamanho de lote 32, otimizador Adam, taxa de aprendizado 0,0001 e função de perda Sparse Categorical Crossentropy. A avaliação quantitativa empregou Matriz de Confusão, Acurácia e F1-Score. Conforme Figura 3.

No segundo ciclo a tarefa foi reformulada para detectar a posição do LED e em seguida classificar. Foram capturadas 404 imagens RGB 320x240 (recortadas para 240x240). Destas, somente 82 são usadas no conjunto de dados, pela dificuldade de anotar manualmente as coordenadas centrais (x,y) dos 9 LEDs. Nesse ciclo foram adotadas outras duas abordagens para tentar obter melhores resultados.

A primeira abordagem utilizou a MobileNetV1 com pesos do ImageNet (α = 0,25), seguida por SeparableConv2D (18 filtros, kernel 5x5, strides 1x1, função de ativação Leaky ReLU) + SeparableConv2D (18 filtros, kernel 3x3, strides 1x1, função de ativação) + Flatten (responsável por achatar os dados) + Dense (18 neurônios, função de ativação

linear), são 18 valores numéricos de saída para estimar as coordenadas por regressão. Mais um código para extrair recortes 3x3 de cada LED a partir da imagem original, e um Perceptron de um neurônio com ativação sigmóide. Alternativamente, testou-se um algoritmo para classificação baseado na média da intensidade das cores. No treinamento, o modelo de regressão usa 62 imagens de treino, 20 de validação, 10.000 épocas e função de perda MSE, enquanto o Perceptron usa 517 imagens de treino, 221 de validação, 1.000 épocas e função de perda Binary CrossEntropy. Para quantização, somente a MobileNetV1 quantizada em INT8, enquanto SeparableConv2D, Dense e Perceptron mantidas em Float32, pois a quantização da SeparableConv2D e Dense provoca perda de precisão maior comparativamente (medido pelo MAE e RMSE). Após a quantização, SeparableConv2D + Dense passam por fine-tuning. A avaliação qualitativa empregou o Erro Médio Absoluto (MAE), a Raiz do Erro Quadrático Médio (RMSE), a matriz de confusão, o F1-Score, e o tempo de inferência no ESP32-S3-CAM usando uma imagem salva na memória do dispositivo. Conforme Figura 4. Por último, com uma nova captura de cenário real, observa-se se o modelo classifica e localiza o LED corretamente.

A segunda abordagem, resumida em (DIAGRAMA_CICLO2_YOLO), faz a comparação com o modelo mais robusto YOLOv8n (JOCHER, 2023) treinado utilizando as 82 imagens originais (320x240). As coordenadas (x,y) foram convertidas em caixas delimitadoras (bounding boxes) de 3x3 pixels, normalizadas. Os rótulos de configuração são iguais. Internamente, a biblioteca utilitária Ultralytics converte-as para 256x256 com reescala e preenchimento de uma borda cinza. Além disso, a biblioteca filtra a saída, removendo detecções de baixa confiança ou duplicadas, para obter os 9 LEDs. O treino executou por 50 épocas (lote 16, otimizador Adam, taxa 0,001) usando as perdas IoU, BCE e Distribution Focal Loss. Por fim, a YOLOv8n original e com quantização INT8 foram embarcadas no ESP32-S3-CAM e mediu-se o tempo de inferência com uma imagem alocada na memória.

A performance de localização espacial foi mensurada pelas métricas MAE e RSME, enquanto a assertividade de classificação usou F1-Score e matriz de confusão. Avaliou-se também a latência de inferência embarcada no ESP32-S3.

Resultados e Discussões

O resultado do primeiro ciclo foi um modelo baseado MobileNet para identificação direta de 7 LEDs, com tamanho aproximado de 0,888 MB.A sua acurácia média foi 0,71, mas com um desvio padrão de 0,17, com acurácia mínima de 0,45 para o LED “DI_a4” e máxima de 0,88 para “DO_a1”. Observou-se que os LEDs com maior acurácia apresentavam maior proporção de exemplos apagados, indicando influência do desbalanceamento das classes sobre a métrica. Como exemplo, o LED DO_a0 apresentou VP = 19, FP = 13, FN = 21 e VN = 19, enquanto o “DI_a4” apresentou VP = 2, FP = 0,

FN = 6 e VN = 64. Usando a métrica F1-Score que pondera o desbalanceamento das classes “aceso” e “apagado”, obteve-se que o F1-Score médio é 0,33.

Portanto, apesar das imagens por configuração de LED estarem balanceadas, elas ainda estavam desbalanceadas segundo as classes aceso/apagado. Assim, os resultados indicaram dois problemas principais: a influência do desbalanceamento das classes sobre as métricas e a dificuldade de identificar LEDs muito pequenos na imagem, uma vez que cada LED ocupa aproximadamente 10×10 pixels, correspondendo a menos de 0,2% da área de uma imagem de 240×240 pixels.

Para o segundo ciclo, o objetivo foi treinar um modelo que tente aprender e usar a informação da posição dos LEDs para melhorar o acerto. A primeira abordagem usou a MobileNet junto de duas SeparableConv2D e uma Dense para estimar as coordenadas dos LEDs. Resultando em um modelo sem quantização de 0,915 MB. As coordenadas estimadas obtiveram MAE de 0,30 e RMSE de 0,30, correspondendo a aproximadamente 0,31 pixel de erro por coordenada. A tarefa de classificação é feita por um único Perceptron recebendo recortes 3x3 como entrada. Ele possui aproximadamente 2 KB e obteve acurácia de 98%. Alternativamente testou-se um algoritmo de classificação baseado na média de intensidade de cor da imagem, mas ele obteve uma acurácia de 80%. A avaliação do conjunto obteve um F1-Score de 97%. Seu tempo de inferência medido no ESP32-S3-CAM foi de 16 segundos. Com a quantização INT8 da MobileNet + SeparableConv2D + Dense o tamanho foi reduzido de 0,915 MB para 0,330 MB e o tempo de inferência reduziu para 1 segundo. Entretanto, provocou um erro médio de localização de 16 pixels nas coordenadas estimadas pelo modelo.

A análise pelo “Mean Absolute Error” e “Root Mean Squared Error” mostrou que, após a quantização, as features produzidas pela MobileNet apresentavam pouca alteração com um MAE de 0,26 e um RMSE de 0,74, enquanto a saída das camadas SeparableConv2D + Dense apresentava diferenças maiores com um MAE de 12,40 e um RMSE de 13,30, resultando em recortes incorretos. Aplicando a quantização somente na MobileNet e realizando o fine-tuning das camadas SeparableConv2D + Dense com treino utilizando as features produzidas pela MobileNet quantizada como entrada, o erro de estimativa passou para um MAE de 0,52 e um RMSE de 0,66. Em seguida, a avaliação do conjunto para identificar a configuração dos LEDs obteve uma matriz de confusão com 320 verdadeiros positivos (VP), 25 falsos negativos (FN), 26 falsos positivos (FP), 367 verdadeiros negativos (VN) e um F1-score de 92%. Um exemplo de uma saída com a localização e classificação dos LEDs pode ser visualizada na Figura 6.

O modelo foi embarcado no ESP32-S3-CAM extraindo os pesos do Perceptron e implementando-o via código, já que um único neurônio pode ser reduzido a uma equação linear e sua função de ativação. O tempo de inferência foi de aproximadamente 1 segundo.

Entretanto, ao posicionar o sistema diante do CLP e realizar a captura no local, a classificação da configuração foi incorreta, variando sem a configuração dos LEDs mudar. A imagem foi salva e posteriormente analisada no Google Colab, permitindo observar que o erro ocorreu na localização dos LEDs, como pode ser visto na Figura 7. Esse resultado indica que o modelo ainda apresenta limitações de generalização diante de variações de iluminação, posição da câmera e condições do ambiente.

Na segunda abordagem, testou-se a YOLOv8n original de 10 MB e a com quantização INT8 de 3 MB. Na avaliação, ambas obtiveram F1-Score de 100%, já o tempo de inferência medido no ESP32-S3-CAM foi aproximadamente de 3 minutos para o original, enquanto o modelo quantizado levou cerca de 16 segundos. Dessa forma, não foi realizado o teste da YOLO em cenário real. Embora tenha apresentado maior precisão que a abordagem com MobileNet, seu tempo de inferência no ESP32-S3-CAM permanece acima de 1 segundo, ainda que possa ser usado para outras aplicações, prefere-se dedicar o trabalho a modelos de maior velocidade.

Conclusão

Como pode ser visto, foi desenvolvido um modelo com uso da MobileNet que classificava LEDs sem aprender a informação de sua posição. Observou-se insuficiências dessa abordagem, o modelo tinha dificuldade de acertar o estado do LED devido seu pequeno tamanho, o que tornava a previsão do modelo enviesada para classe majoritária. Essas insuficiências foram superadas por meio de modelos que identificavam a posição dos LEDs e usavam essa informação para facilitar a classificação, mas que geravam outros desafios. Sob essa perspectiva, foram desenvolvidas uma abordagem com a MobileNet capaz de classificar corretamente os LEDs e outra com a YOLO, também classificando corretamente. Dessas abordagens, foram testadas variações usando quantização para reduzir seu tamanho. Somente a abordagem com uso da MobileNet e quantização mostrou obter um tempo de inferência de 1 segundo. No entanto, ela ainda apresenta limitações para um cenário real, pois o conjunto de dados criados tornou o modelo de IA muito sensível à configuração da câmera, iluminação e posição. Por questões de tempo, não foi possível realizar testes mais longos e com mais dados. Para trabalhos futuros, propõem-se tornar o conjunto de dados mais robusto, trocar a câmera por uma de resolução maior.

Referências

HOWARD, Andrew G. et al. MobileNets: efficient convolutional neural networks for mobile vision applications. arXiv preprint, arXiv:1704.04861, 2017. Disponível em: https://arxiv.org/abs/1704.04861. Acesso em: 01 de maio 2026.

JOCHER, Glenn; CHAURASIA, Ayush; QIU, Jing. Ultralytics YOLOv8. Version 8.0.0. 2023. Disponível em: https://github.com/ultralytics/ultralytics. Acesso em: 31 ago. 2026.

KERAS. Keras API documentation. 2026. Disponível em: https://keras.io/api/. Acesso em: 31 ago. 2026.

QUACH, Luyl-Da; QUOC, Khang Nguyen; QUYNH, Anh Nguyen; NGOC, Hoang Tran; THAI-NGHE, Nguyen. Tomato Health Monitoring System: tomato classification, detection, and counting system based on yolov8 model with explainable mobilenet models using grad-cam++. Ieee Access, [S.L.], v. 12, p. 9719-9737, 2024. Institute of Electrical and Electronics Engineers (IEEE). http://dx.doi.org/10.1109/access.2024.3351805.

RAMAN, Karthik. Image Analysis of PLC Diagnosis System. International Journal Of Advanced Trends In Computer Science And Engineering, [S.L.], v. 9, n. 2, p. 1973-1977, 25 abr. 2020. The World Academy of Research in Science and Engineering. http://dx.doi.org/10.30534/ijatcse/2020/165922020.

ZHANG, Jieping; ZHANG, Jieping. Analysis of the Application of Machine Vision in the Automation Control of Power Electronic Equipment. Archives Des Sciences, [S.L.], v. 74, n. 1, p. 123-132, 10 mar. 2024. ScienceStride Publishing. http://dx.doi.org/10.62227/as/74115.

Anexos



Figura 3 - Fluxo de trabalho do ciclo 1 Fonte: Gerado por IA. Gemini. 31 de agosto de 2026.



Figura 4 - Fluxo de trabalho do ciclo 2 para abordagem com Mobilenet Fonte: Gerado por IA. Gemini. 31 de agosto de 2026.



Figura 5 - Fluxo de trabalho do ciclo 2 para abordagem com YOLO Fonte: Gerado por IA. Gemini. 31 de agosto de 2026.

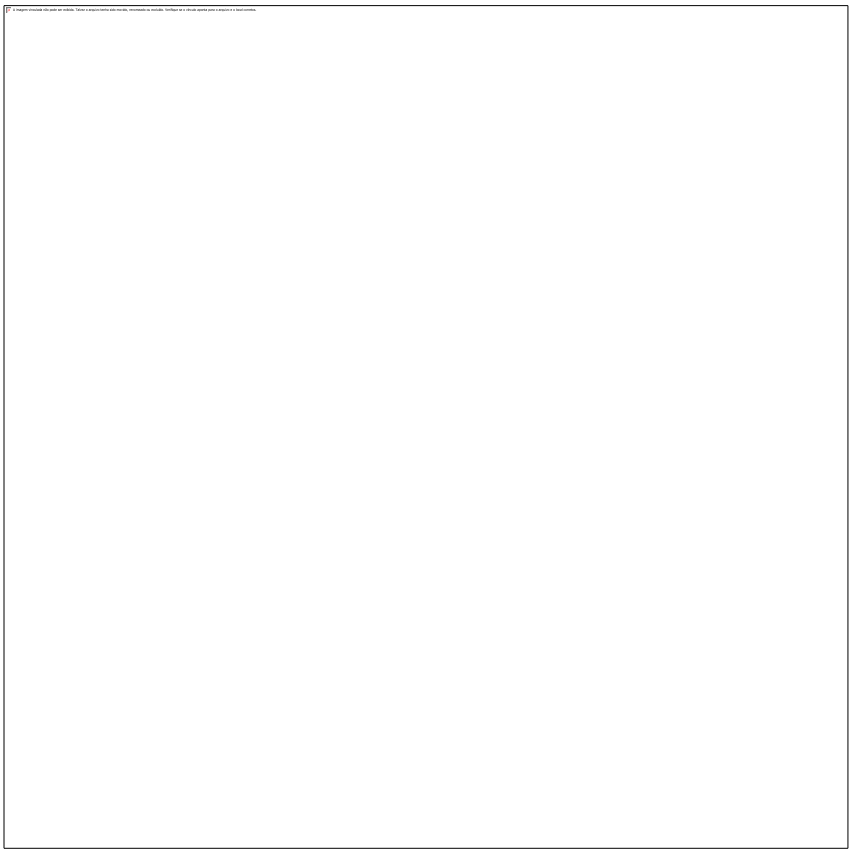


Figura 6 - Exemplo de saída da abordagem com Mobilenet Fonte: Elaborado pelo autor (2026).

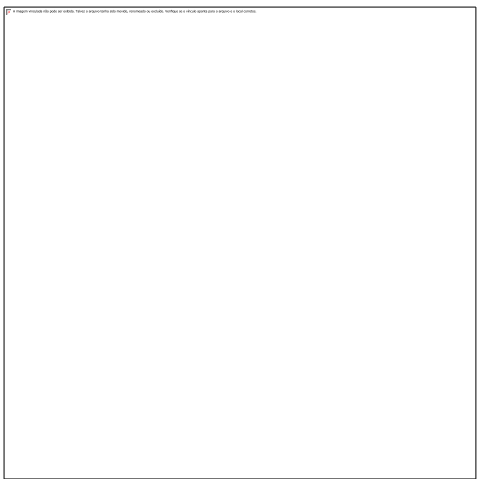


Figura 7 - Localização dos recortes com nova captura Fonte: Elaborado pelo autor (2026).

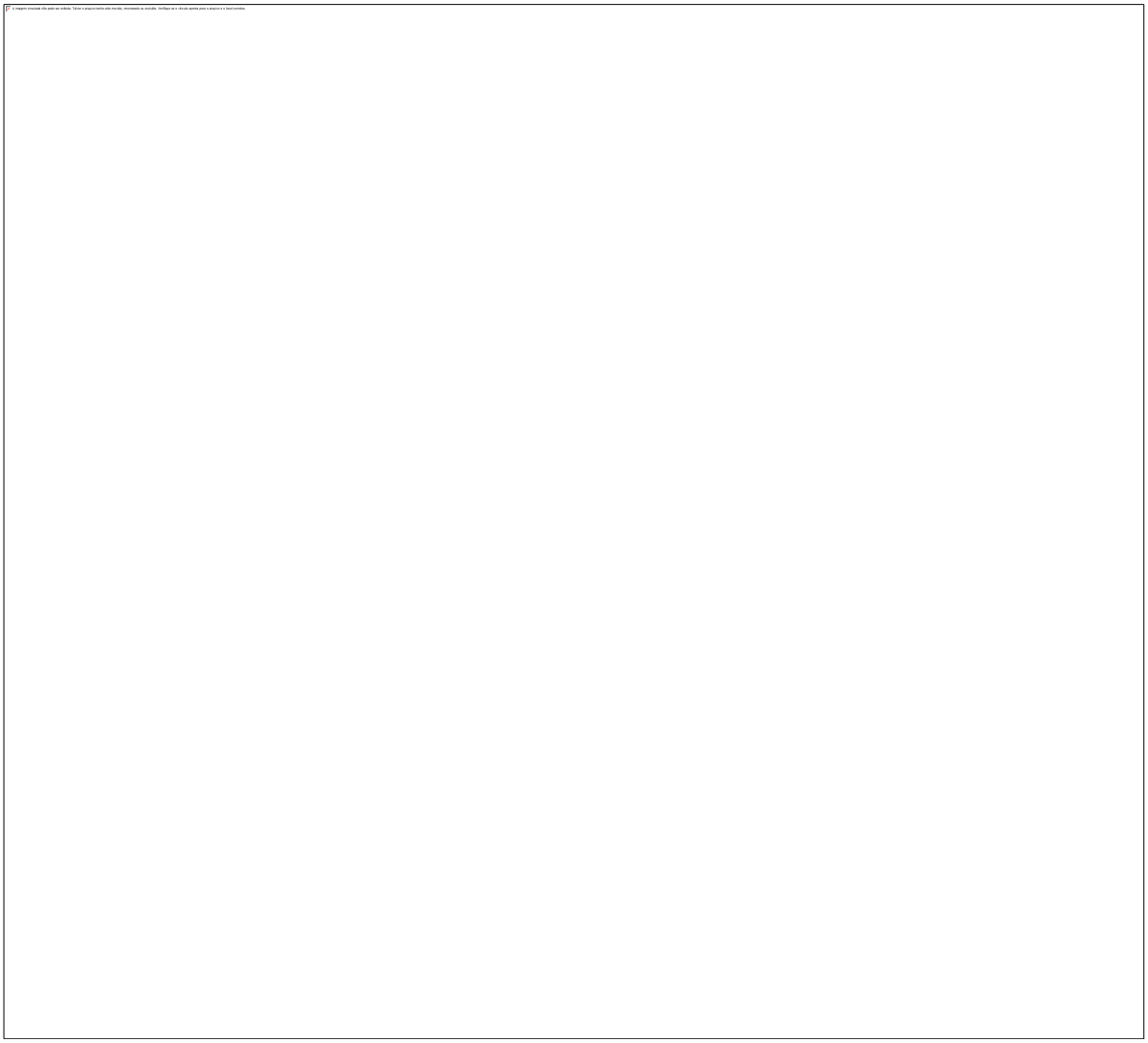


Figura 1 - Leds do CLP com nomeação Fonte: Elaborado pelo autor (2026).



Figura 2 - Leds do CLP por configuração Fonte: Elaborado pelo autor (2026).

CÓDIGO: ET1611

AUTOR: ARAMYS ANDREEW TAVARES DE PAIVA

ORIENTADOR: TIBERIO AZEVEDO PEREIRA

TÍTULO: Exploração e Aplicação de LLMs Locais: Chatbots e RAG com Ollama

Resumo

Este relatório apresenta os resultados da primeira fase de um projeto de Iniciação Científica (PIBIC) voltado à exploração de Modelos de Linguagem de Grande Escala (LLMs) executados localmente. A dependência de APIs proprietárias em nuvem traz desafios de custo, latência e privacidade de dados, o que motiva o estudo de alternativas locais, como a plataforma Ollama. Nesta primeira fase, o trabalho concentrou-se na fundamentação teórica sobre IA generativa e LLMs e, principalmente, no uso prático de LLMs via API. A partir disso, foram conduzidos quatro experimentos exploratórios com o modelo gemma3:270m, avaliando habilidades de raciocínio lógico, resolução de problemas, tradução e suporte multilíngue. Os resultados indicaram bom desempenho em tarefas de raciocínio simples e de tradução, com respostas corretas e fluentes, mas revelaram limitações em tarefas aritméticas de múltiplas etapas, nas quais o modelo entrou em um padrão de repetição sem chegar a uma resposta correta e em respostas multilíngues, por vezes incompletas. Esses achados são coerentes com o porte reduzido do modelo testado e reforçam a importância de avaliar criteriosamente modelos compactos antes de aplicá-los em tarefas que exijam maior precisão.

Palavras-chave: Modelos de Linguagem; LLM; Ollama; API; Inteligência Artificial Genera

TITLE: Modelos de Linguagem; LLM; Ollama; API; Inteligência Artificial Generativa.

Abstract

This report presents the results of the first phase of a Scientific Initiation project (PIBIC) focused on exploring locally executed Large Language Models (LLMs). Reliance on proprietary cloud APIs raises challenges related to cost, latency, and data privacy, which motivates the study of local alternatives such as the Ollama platform. This first phase focused on the theoretical foundations of generative AI and LLMs and, mainly, on the practical use of LLMs via API. Based on this, four exploratory experiments were conducted with the gemma3:270m model, assessing logical

reasoning, problem solving, translation, and multilingual support. Results showed good performance on simple reasoning and translation tasks, with correct and fluent responses, but revealed limitations in multi-step arithmetic tasks — where the model fell into a repetitive pattern without reaching a correct answer — and in multilingual responses, which were sometimes incomplete. These findings are consistent with the reduced size of the tested model and reinforce the importance of carefully evaluating compact models before applying them to tasks that require higher precision.

Keywords: Language Models; LLM; Ollama; API; Generative Artificial Intelligence.

Introdução

Os Grandes Modelos de Linguagem (LLMs) transformaram radicalmente o panorama da Inteligência Artificial, oferecendo capacidades avançadas de compreensão e geração de texto. No entanto, a dependência predominante de APIs proprietárias hospedadas na nuvem gera desafios críticos, como o alto custo de inferência, a latência de rede e, principalmente, a falta de privacidade de dados sensíveis. Nesse contexto, o uso de LLMs em ambientes locais surge como uma solução robusta e necessária.

Este projeto propõe o estudo e a aplicação de LLMs executados localmente através da plataforma Ollama, uma ferramenta que simplifica a orquestração desses modelos. O projeto, em sua concepção mais ampla, tem como horizonte a construção de pipelines de Geração Aumentada por Recuperação (RAG) e de assistentes virtuais contextuais. Este relatório, contudo, refere-se especificamente à primeira fase da bolsa, dedicada à fundamentação teórica e à exploração prática do uso de LLMs locais via API.

Objetivos desta fase:

• Compreender, em nível introdutório, o funcionamento de modelos de linguagem de grande escala e da IA generativa;

• Configurar e operar um ambiente local de inferência utilizando a plataforma Ollama e sua API;

• Conhecer os diferentes modelos de linguagem disponíveis gratuitamente em plataformas locais como o Ollama;

• Explorar, na prática, distintas habilidades oferecidas por um modelo de linguagem local — raciocínio, resolução de problemas, tradução e suporte multilíngue.

Metodologia

O estudo teórico desta fase baseou-se na obra de Bahree (2024), abrangendo os três capítulos iniciais, com ênfase no terceiro. Foram revisados os conceitos fundamentais de IA generativa, os princípios de funcionamento dos LLMs — incluindo a arquitetura Transformer, o processo de tokenização e o treinamento por previsão de próximo token — e, principalmente, o uso prático de LLMs por meio de APIs REST: o envio de requisições contendo um prompt e parâmetros de controle da geração, e o processamento das respostas retornadas pelo modelo.

A etapa prática foi conduzida utilizando o Ollama, plataforma que permite baixar e executar localmente diversos modelos de linguagem de pesos abertos e disponibilização gratuita — como Llama 3, Mistral, Phi-3 e a família Gemma. Nesta fase, os testes práticos foram realizados com o modelo gemma3:270m, escolhido por ser compacto e adequado à exploração inicial em hardware limitado.

A interação com o modelo foi realizada por meio de requisições HTTP à API local do Ollama (endpoint /api/generate), utilizando a biblioteca requests em Python, em ambiente Jupyter Notebook. Foram elaborados quatro experimentos exploratórios, cada um direcionado a uma habilidade distinta do modelo:

• Raciocínio lógico: aplicação de um problema simples de lógica, avaliando a capacidade do modelo de seguir uma cadeia de raciocínio;

• Resolução de problemas: aplicação de um problema matemático simples, avaliando a correção da solução apresentada;

• Tradução: solicitação de tradução de um trecho de texto entre dois idiomas, avaliando fidelidade e fluência;

• Suporte multilíngue: interação em um idioma diferente do português, avaliando a capacidade de compreensão e resposta coerente.

Em cada experimento foram registrados o tempo de resposta, o prompt utilizado, a resposta gerada e uma avaliação qualitativa do resultado (correto/incorreto, coerente/incoerente).

Resultados e Discussões

RESULTADOS E DISCUSSÃO: 20000

Raciocínio lógico

O que foi testado: o modelo recebeu uma pergunta de lógica simples — se todo gato é mamífero e todo mamífero tem coração, então um gato tem coração — para avaliar se conseguia chegar a uma conclusão correta e explicar o raciocínio.

Prompt utilizado

Resposta do modelo

Correto?

Tempo

Se todos os gatos são mamíferos, e todo mamífero tem coração, então um gato tem coração? Explique seu raciocínio.

"Sim, um gato tem coração."

Sim (parcial)

3,72 s

Fonte: elaborado pelo autor (2026), a partir da saída do modelo gemma3:270m via API do Ollama.

Resultado: o modelo chegou à conclusão logicamente correta, mas respondeu de forma direta e sucinta, sem explicitar a cadeia de raciocínio (silogismo) solicitada no prompt. Isso sugere que, embora o modelo seja capaz de inferir a resposta correta, sua capacidade de verbalizar o processo inferencial passo a passo é limitada — comportamento esperado em um modelo de porte reduzido como o gemma3:270m.

Resolução de problemas

O que foi testado: o modelo recebeu um problema aritmético simples — o cálculo do novo preço de um produto de R$ 80 após um desconto de 25% —, para avaliar se conseguia realizar corretamente um cálculo com múltiplas etapas.

Prompt utilizado

Resposta do modelo (resumida)

Correto?

Tempo

Uma loja vende um produto por R$ 80. Durante uma promoção, o preço foi reduzido em 25%. Qual o novo preço? Mostre o cálculo.

"R$ 80 - 25 * R$ 80 = R$ 80 - 200 * R$ 80 = ..." (padrão repetido até o limite de tokens)

Não

12,43 s

Fonte: elaborado pelo autor (2026). Nota: resposta truncada na tabela por conter repetição extensa; a saída completa está registrada em resultados_experimentos.json.

Figura 3. Execução do experimento de resolução de problemas no notebook, evidenciando o padrão de repetição na saída do modelo.

Fonte: elaborado pelo autor (2026).

Resultado: diferentemente do experimento anterior, o modelo não conseguiu resolver corretamente o problema (resposta esperada: R$ 60,00). Em vez de aplicar a redução percentual, ele entrou em um padrão de repetição do mesmo trecho de cálculo até atingir o limite de tokens da resposta, sem nunca chegar a um valor final.

Esse comportamento, além de indicar erro de conteúdo, evidencia uma limitação técnica comum em modelos muito compactos: a degeneração da geração em loops repetitivos, especialmente em tarefas que exigem múltiplas etapas de raciocínio aritmético. O tempo de resposta consideravelmente maior (12,43 s frente a menos de 4 s nos demais experimentos) também é compatível com esse comportamento de repetição.

Tradução

O que foi testado: solicitou-se ao modelo a tradução de uma frase curta do português para o inglês, para avaliar fidelidade ao sentido original e fluência do texto gerado.

Texto original

Tradução gerada

Qualidade

Tempo

"A inteligência artificial está transformando a forma como estudamos e trabalhamos."

"Artificial intelligence is transforming how we study and work."

Fiel e fluente

1,74 s

Fonte: elaborado pelo autor (2026).

Figura 4. Execução do experimento de tradução no notebook e saída gerada pelo modelo gemma3:270m.

Fonte: elaborado pelo autor (2026).

Resultado: a tradução gerada foi correta, fluente e fiel ao sentido do texto original, produzida em um tempo de resposta baixo. Esse foi o melhor resultado qualitativo entre

os quatro experimentos, indicando que tarefas de tradução direta de frases curtas estão bem dentro da capacidade do modelo compacto testado.

Testes adicionais de tradução, com frases de maior complexidade e outros pares de idiomas, estão em andamento para enriquecer esta análise em versões futuras deste relatório.

Suporte multilíngue

O que foi testado: o modelo recebeu uma pergunta em francês (não em português), pedindo a capital da França e a razão de sua importância, para avaliar compreensão e resposta coerente em outro idioma.

Idioma testado

Prompt

Resposta do modelo

Completa/coerente?

Francês

"Quelle est la capitale de la France, et pourquoi cette ville est-elle importante ?"

"La capitale de la France est Paris."

Coerente, porém incompleta

Fonte: elaborado pelo autor (2026).

Resultado: o modelo compreendeu corretamente o prompt em francês e respondeu de forma coerente e gramaticalmente correta no mesmo idioma, demonstrando capacidade multilíngue básica. Contudo, a resposta abordou apenas a primeira parte da pergunta (identificar a capital), sem discutir a importância da cidade, como havia sido solicitado.

O tempo de resposta muito baixo (0,59 s) é compatível com uma geração interrompida precocemente, o que sugere uma limitação do modelo em manter respostas multilíngues mais longas e completas.

Síntese dos resultados

De forma geral, os quatro experimentos evidenciam um padrão: o modelo gemma3:270m apresenta bom desempenho em tarefas diretas e de menor complexidade — como inferências lógicas simples e tradução de frases curtas — mas revela limitações relevantes em tarefas que exigem múltiplas etapas de raciocínio (resolução de problemas aritméticos) ou respostas mais longas e completas (suporte multilíngue). Esses resultados são coerentes com o porte reduzido do modelo e motivam a

comparação, em etapas futuras, com modelos de maior porte disponíveis na mesma plataforma, como Llama 3, Mistral e Phi-3.

Conclusão

A primeira fase desta Iniciação Científica teve como foco a construção de uma base de conhecimento sobre modelos de linguagem de grande escala e sobre o uso prático de LLMs locais via API, utilizando a plataforma Ollama. Os experimentos exploratórios permitiram observar, na prática, tanto capacidades quanto limitações de um modelo de linguagem local compacto: bom desempenho em raciocínio lógico simples e em tradução, e dificuldades em resolução de problemas aritméticos de múltiplas etapas e em respostas multilíngues completas.

A próxima fase do projeto avançará para o tratamento e leitura de arquivos (documentos, PDFs), a integração de LLMs locais com conteúdo de websites, e o desenvolvimento de um chatbot funcional — aproximando o projeto da arquitetura de Geração Aumentada por Recuperação (RAG) prevista no plano de trabalho original.

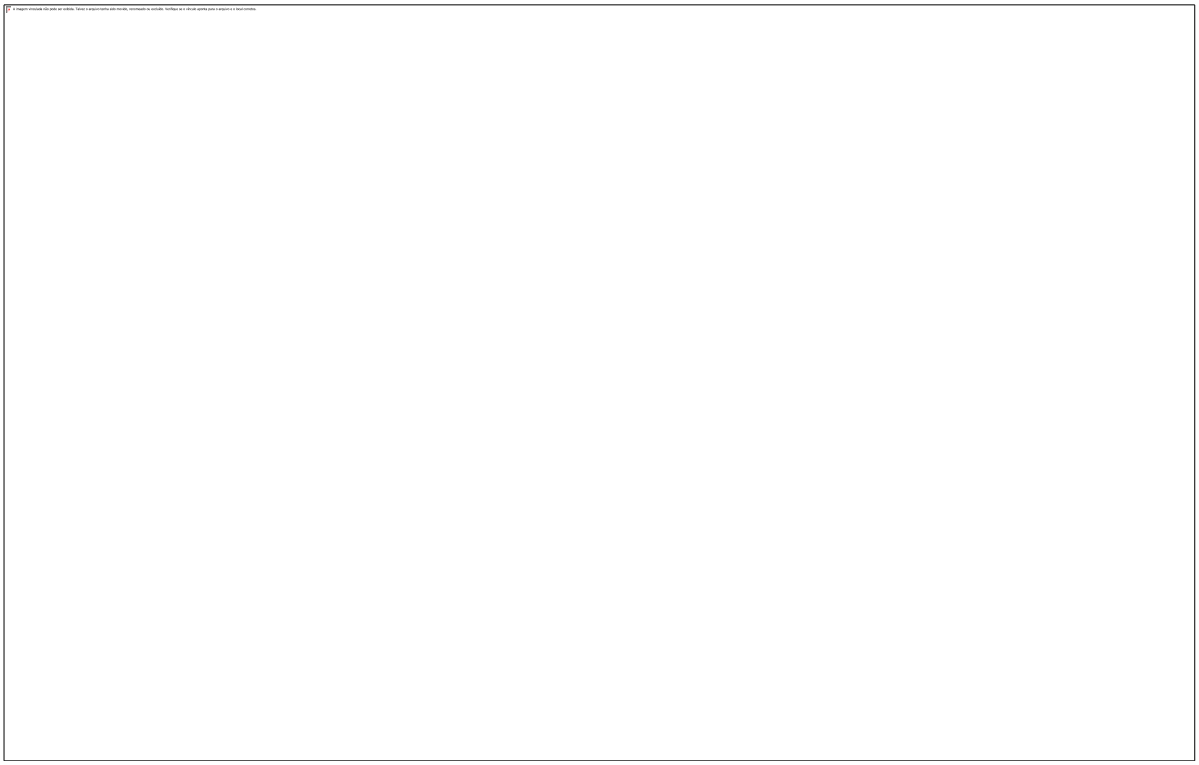
Referências

BAHREE, Amit. Generative AI in Action. Manning, 2024.

OLLAMA. Documentation. 2025. Disponível em: https://ollama.com.

RASCHKA, Sebastian. Build a Large Language Model (From Scratch). Manning, 2025.

Anexos



Graficos01

CÓDIGO: ET1646

AUTOR: YASMIN GIORDANO SANTOS

ORIENTADOR: ROGER KREUTZ IMMICH

TÍTULO: Blockchain e Smart Contracts para Mitigação de Ataques de Envenenamento em Aprendizado Federado Baseado em Internet das Coisas: Um Mapeamento Sistemático da Literatura

Resumo

O avanço da Internet das Coisas (IoT) tem impulsionado a adoção do Aprendizado Federado (FL), contudo a vulnerabilidade a ataques de envenenamento e a dependência de servidores centrais comprometem a segurança dessas redes. Embora blockchain e smart contracts descentralizem o sistema por meio de um ledger distribuído, imutável e auditável, a proteção da rede esbarra em um trade-off entre resiliência e alto custo computacional imposto aos dispositivos de borda. Uma análise do estado da arte evidencia que o gargalo central reside em uma assimetria crítica nas arquiteturas atuais: enquanto a validação do conteúdo dos gradientes apresenta alto rigor formal, a gestão de identidade e o provisionamento são omitidos em quase um terço da amostra. Essa negligência na camada de entrada compromete a integridade on-chain, evidenciando que a falta de autenticação eficaz prévia agrava a vulnerabilidade a ataques. Conclui-se que a mitigação efetiva em ambientes restritos exige uma abordagem em níveis, de forma a integrar autenticação identitária, validação estatística de baixo custo e reputação acumulada.

Palavras-chave: Blockchain, Smart Contracts, Aprendizado Federado, IoT, Envenenamento.

TITLE: Blockchain and Smart Contracts for Mitigating Poisoning Attacks in Internet of Things-Based Federated Learning: A Systematic Literature Mapping

Abstract

The advancement of the Internet of Things (IoT) has driven the adoption of Federated Learning (FL), however, vulnerability to poisoning attacks and reliance on central servers compromise the security of these networks. Although blockchain and smart contracts decentralize the system through a distributed, immutable, and auditable ledger, network protection faces a trade-off between resilience and the high computational cost imposed on edge devices. An analysis of the state of the art demonstrates that the primary bottleneck lies in a critical asymmetry within current architectures: while the validation of gradient content exhibits high formal rigor, identity management and provisioning are omitted in nearly a third of the sample. This oversight at the entry layer compromises on-chain integrity, highlighting that the lack of effective prior authentication exacerbates vulnerability to attacks. It is concluded that effective mitigation in resource-constrained environments requires a layered approach, integrating identity authentication, low-cost statistical validation, and accumulated reputation.

Keywords: Blockchain, Smart Contracts, Federated Learning, IoT, Poisoning.

Introdução

A rápida evolução da Internet das Coisas (Internet of Things – IoT) transformou radicalmente a infraestrutura de coleta e processamento de dados (ATZORI; IERA; MORABITO, 2010), com dispositivos heterogêneos alocados na borda da rede gerando e armazenando volumes massivos de dados sensíveis (AL-FUQAHA et al., 2015). Na arquitetura computacional tradicional de Inteligência Artificial (IA) e Aprendizado de Máquina (Machine Learning – ML), a criação de modelos preditivos exige a transferência contínua desses dados brutos para servidores em nuvem ou centros de processamento centralizados (MCMAHAN et al., 2017). Esse modelo, no entanto, introduz alguns gargalos operacionais: alto consumo de largura de banda na rede, latências de comunicação inadequadas para sistemas em tempo real (SHI et al., 2016) e exposição de usuários e organizações a violações regulatórias de privacidade, tais como as estabelecidas pelo Regulamento Geral sobre a Proteção de Dados (UNIÃO EUROPEIA, 2016) e pela Lei Geral de Proteção de Dados (BRASIL, 2018).

Para superar essas barreiras, o Aprendizado Federado (Federated Learning – FL) emergiu como um paradigma para o treinamento colaborativo e distribuído de algoritmos de ML na borda (MCMAHAN et al., 2017; KAIROUZ et al., 2021). No paradigma do FL, os nós não transferem seus dados brutos para um repositório central. Em vez disso, cada um executa o treinamento do modelo localmente com base em suas próprias amostras de dados, compartilhando apenas os parâmetros atualizados do modelo com um servidor agregador (MCMAHAN et al., 2017; LI et al., 2020). Este consolida as contribuições de múltiplos clientes por meio de algoritmos de agregação, gerando um modelo global que é subsequentemente redistribuído para os nós participantes para novos ciclos iterativos de treinamento (MCMAHAN et al., 2017).

Embora o Aprendizado Federado busque promover a privacidade, a descentralização do treinamento e a alocação física de dispositivos IoT em ambientes desprotegidos expandem a superfície de ataque do sistema (RODRÍGUEZ-BARROSO et al., 2023). Dispositivos comprometidos transformam-se em nós adversários, viabilizando Ataques de Envenenamento (Poisoning Attacks), divididos entre envenenamento de dados (data poisoning, como label flipping e injeção de ruído) e envenenamento de modelo ou gradientes (model/gradient poisoning, como scaling attacks), além de variações evasivas como ataques de backdoor e Sybil. Somado a isso, as arquiteturas convencionais de FL dependem de um servidor central agregador, o que estabelece um ponto único de falha e compromete a rastreabilidade auditável das atualizações transmitidas.

Para mitigar essas vulnerabilidades, a integração de blockchain e smart contracts reestrutura a topologia do FL ao descentralizar a agregação ou atuar como uma camada externa de governança sobre o agregador central por meio de um ledger distribuído, imutável e transparente. Essa infraestrutura atua na auditabilidade do histórico de parâmetros e hashes (WANG et al., 2024), na validação on-chain de gradientes, no cálculo de pontuações de reputação (Trust Scores) e na gestão de identidade para o provisionamento seguro de nós IoT. Contudo, a viabilização dessas arquiteturas exige a calibração de trade-offs entre robustez defensiva, latência de consenso, custos de armazenamento e consumo energético em nós de recursos restritos.

Diante do crescimento de pesquisas a partir de 2025 (JUNIOR et al., 2025) e da fragmentação do corpo de conhecimento sobre a mitigação de ataques de envenenamento em FL-IoT, justifica-se a condução deste Mapeamento Sistemático da Literatura (MSL). O objetivo geral deste estudo é identificar, classificar e sintetizar as evidências científicas sobre arquiteturas integradas de Aprendizado Federado, blockchain e contratos inteligentes projetadas para a mitigação, auditoria e isolamento de ataques de envenenamento em ecossistemas da Internet das Coisas.

Metodologia

O presente Mapeamento Sistemático da Literatura (MSL) foi conduzido segundo as diretrizes metodológicas estabelecidas por Kitchenham e Charters (2007) para revisões sistemáticas em Engenharia de Software, complementadas pelo arcabouço para Systematic Mapping Studies de Petersen, Vakkalanka e Kuzniarz (2015) O protocolo estruturou-se em três fases fundamentais: Planejamento (escopo, perguntas de pesquisa e critérios de seleção); Condução (busca, triagem e avaliação de qualidade); e Relatório (extração, síntese e categorização dos dados).

2.1 PICOC e Perguntas de Pesquisa

A delimitação do escopo fundamentou-se na estrutura PICOC (Population, Intervention, Comparison, Outcome, Context) (KITCHENHAM; CHARTERS, 2007), alinhando os conceitos de Aprendizado Federado em IoT às infraestruturas de tecnologia de registro distribuído (DLT):

- **Population** (P):ecossistemas de Internet das Coisas (IoT), Internet das Coisas Industrial (IIoT) e dispositivos de computação na borda (edge computing).
- **Intervention** (I): arquiteturas de Aprendizado Federado (FL) integradas a redes blockchain e smart contracts.
- **Comparison** (C): servidores agregadores centrais de FL, algoritmos de defesa convencionais sem uso de DLT e sistemas IoT sem mecanismos defensivos.
- **Outcome** (O): mitigação de ataques de envenenamento (data poisoning, model/gradient poisoning, backdoors), auditabilidade de atualizações e resiliência da acurácia do modelo.
- **Context** (C): provisionamento, autenticação e monitoramento de nós em ambientes de borda com restrição de recursos

2.2 Perguntas de Pesquisa

A partir das lacunas identificadas na literatura e dos componentes da PICOC, formularam-se cinco perguntas de pesquisa:

- **RQ1**: Quais são as principais estratégias de ataques de envenenamento direcionadas a ambientes de Federated Learning em IoT documentadas na literatura?
- **RQ2**: Como a tecnologia blockchain e os smart contracts são arquitetados para auditar as atualizações locais, validar gradientes e isolar/revogar os nós atacantes?
- **RQ3**: Qual é o papel dos mecanismos de registro, gerenciamento de identidade e autenticação (provisionamento) baseados em blockchain na prevenção inicial da entrada de nós IoT maliciosos na rede de Aprendizado Federado?
- **RQ4**: Quais algoritmos ou técnicas são executados de forma integrada à blockchain para detectar as anomalias nos gradientes?
- **RQ5**: Quais são os impactos e trade-offs quantificados na literatura entre a resiliência contra os ataques de envenenamento e a eficiência computacional/energética de dispositivos IoT restritos?

2.3 Estratégia de Busca nas Bases de Dados

A estratégia de busca seguiu o processo recomendado por Kitchenham e Charters (2007) de derivação de termos via PICOC, com testes exploratórios de calibração da string em cada base, de modo a equilibrar sensibilidade e precisão da busca, resultando na seguinte string genérica: ("Internet of Things" OR "IoT" OR "Edge") AND ("Federated Learning" OR "FL") AND ("poisoning" OR "backdoor") AND ("blockchain" OR "smart contract") AND ("defense" OR "mitigation" OR "secure aggregation" OR "authentication" OR "provisioning").

A aplicação adaptada dessa consulta em cinco bases digitais recuperou 162 registros brutos: IEEE Xplore, 60 artigos (37,0%), com a string original; Scopus, 55 registros (34,0%), via TITLE-ABS-KEY; Web of Science, 36 artigos (22,2%), por Tópico (TS), ambas com filtros nativos de idioma e tipo de documento; ScienceDirect, 6 estudos (3,7%), com consulta sintetizada em razão do limite de oito operadores booleanos; e ACM Digital Library, 5 registros (3,1%), restrita ao campo Abstract após ruído da busca irrestrita.

2.4 Processo de Seleção e Critérios de Elegibilidade

A triagem adotou o fluxo PRISMA (PAGE et al., 2021), em três etapas: Identificação, com recuperação e remoção de duplicatas entre as cinco bases; Triagem, dividida em leitura de títulos e resumos e, posteriormente, leitura integral dos textos restantes; e Inclusão. A filtragem articulou-se a cinco critérios de inclusão (IC1–IC5) e oito de exclusão (EC1–EC8), cujo enunciado consta no Anexo 1, resultando em amostra final de 16 estudos primários (2022–2026). O detalhamento quantitativo e as justificativas de exclusão constam no Diagrama PRISMA, no Anexo 2.

2.5 Avaliação de Qualidade dos Estudos

Para mensurar a precisão metodológica e a relevância científica dos 16 estudos primários selecionados, aplicou-se um checklist de Avaliação de Qualidade (Quality Assessment) (DYBÅ; DINGSØYR, 2008) composto por cinco questões:

- **QA1**: O objetivo da pesquisa e a lacuna que ela propõe preencher estão claramente definidos?
- **QA2**: O artigo descreve detalhadamente o cenário do ataque de envenenamento e as capacidades do atacante?
- **QA3**: A arquitetura baseada em blockchain e o papel dos smart contracts na mitigação/isolamento são explicados de forma clara?
- **QA4**: A proposta foi validada por meio de experimentos quantitativos, simulações ou provas formais adequadas?
- **QA5**: A solução proposta leva em consideração as limitações de recursos (processamento, energia, armazenamento) dos dispositivos IoT ou da rede de borda?

Cada questão foi respondida em três níveis – Sim (1,0), Parcialmente (0,5) e Não (0,0) –, cuja soma compôs a nota final de qualidade de cada estudo, empregada como fator de ponderação na síntese dos resultados, com 3,5 como nota de corte.

2.6 Formulário de Extração de Dados

Para viabilizar uma síntese comparável entre os 16 estudos primários, Elaborou-se um formulário de extração com 15 atributos padronizados (Anexo 3), aplicado a cada estudo mediante leitura integral do texto completo, com registro literal ou parafraseado em planilha. Os atributos 1 a 4 registram a identificação bibliográfica do estudo; os atributos 5 e 6 alimentam a taxonomia de ataques da RQ1; os atributos 7 a 10 sustentam a caracterização arquitetural da RQ2, com os atributos 8 e 10 aprofundados também na RQ3; os atributos 11 e 12 fundamentam a RQ4; e os atributos 13 a 15, incluindo limitações declaradas pelos autores para a RQ5.

Resultados e Discussões

A amostra final deste mapeamento é composta por 16 estudos primários, publicados entre 2022 e 2026, extraídos das cinco bases digitais consultadas e submetidos aos 19 atributos do formulário de extração. A

distribuição temporal revela uma concentração acentuada de publicações em 2025 (9 estudos; 56,3% da amostra), seguida por 2026 (3 estudos; 18,8%), enquanto 2022 e 2024 contribuem, cada um, com apenas 2 estudos (12,5%). Essa curva de crescimento corrobora a justificativa metodológica do protocolo quanto à emergência recente do tema, sugerindo que a convergência entre blockchain, smart contracts e defesa de Aprendizado Federado contra envenenamento em IoT é uma linha de pesquisa em fase de consolidação, e não ainda amadurecida.

Quanto à arquitetura de rede, prevalece amplamente o modelo de blockchain de consórcio/permissionada (13/16 estudos), tipicamente implementada sobre Hyperledger Fabric com Membership Service Provider (MSP) e consenso PBFT (ISSA et al., 2025; KASYAP; TRIPATHY, 2024; RAHMAN; WANG; WEI, 2025). Arquiteturas híbridas (blockchain pública combinada a um canal privado) aparecem em 2 estudos (WANG et al., 2024; FENG et al., 2022), e apenas um estudo adota blockchain pública/permissionless de forma exclusiva (OLAPOJOYE et al., 2025), configuração que, notavelmente, é a única da amostra sem um mecanismo de identidade/registro detalhado.

Funcionalmente, a blockchain desempenha três papéis recorrentes: armazenamento imutável de gradientes/modelos para fins de auditoria (14/16 estudos); gerenciamento de reputação/pontuação de nós (8/16) ; e substituição total do agregador central por agregação descentralizada (também em 8/16). Essa sobreposição indica que a maioria das soluções combina ao menos duas dessas funções simultaneamente, e não as trata como alternativas mutuamente excludentes. Um achado de interesse é que, apesar dos smart contracts serem componente central desse mapeamento, três estudos (18,8%) não operacionalizam contratos inteligentes de forma explícita em sua arquitetura de defesa (WEI et al., 2026; ALHARBI, 2025; MUN et al., 2025), recorrendo predominantemente à lógica de consenso entre nós ou a autoridades externas, evidência de heterogeneidade conceitual relevante para a interpretação das RQs seguintes.

3.1 RQ1: Estratégias de Ataques de Envenenamento

A RQ1 buscou mapear as principais estratégias de envenenamento documentadas em FL-IoT. A extração revela uma taxonomia predominantemente bimodal, complementada por vetores mais especializados. O envenenamento de dados (data poisoning) está presente em 12/16 estudos (75%), manifestando-se sobretudo como label flipping e injeção de amostras falsas nos conjuntos de treinamento local. É o vetor mais testado empiricamente, com frações de nós maliciosos variando de 10% (WANG et al., 2024) a 90% (ROY; SINGHA; BERA, 2026), o que indica ausência de padronização metodológica entre os estudos quanto ao limiar de estresse adversarial considerado realista para cenários de borda. O envenenamento de modelo/gradiente (model/gradient poisoning) aparece com a mesma frequência, incluindo variantes específicas como scaling attacks, foco exclusivo de Olapojoye et al. (2025), e sign-flipping com fator ±1,5×, em Alharbi (2025): trata-se do vetor mais sofisticado quanto à fundamentação matemática, frequentemente acompanhado de provas formais de detecção como a desigualdade DKW em (OLAPOJOYE et al., 2025) e a separação de clusters (WEI et al., 2026).

Ataques de backdoor ocorrem em 5/16 estudos (31%), caracterizados por gatilhos de features específicas visando a classificação incorreta de uma classe-alvo. Destacam-se Roy, Singha e Bera (2026), que reportam taxa de sucesso de ataque (ASR) contida a aproximadamente 9% mesmo com 90% de clientes maliciosos, e Salim, Naaz e Choi (2025), cujo mecanismo de rollback reduz o ASR para menos de 5% em até quatro rodadas. AtaquesSybil, envolvendo múltiplas identidades falsas, aparecem em 3/16 estudos (19%) (SALIM; NAAZ; CHOI, 2025; ALHARBI, 2025; ALATAWI., 2025); em Alatawi (2025), 10% dos nós simulados adotam essa estratégia, sendo bloqueados via validação cruzada de identificadores no Smart Contract. Vetores híbridos, não capturados pela taxonomia clássica de poisoning attacks mas relevantes ao

ecossistema FL-blockchain, incluem ataques de inversão de gradiente para reconstrução de dados privados (WANG et al., 2024); free-riding e vazamento por inferência (SALIM; NAAZ; CHOI, 2025); falhas Bizantinas coordenadas multivetoriais (ALHARBI, 2025); e ataques de inferência de associação/reconstrução sobre gradientes em texto claro (FENG et al., 2022).

A análise cruzada revela um padrão relevante: estudos que testam frações extremas de nós maliciosos, como Cheng et al. (2022), com até 80%, e Roy, Singha e Bera (2026) com até 90%, dependem de mecanismos de validação por comitê ou de detecção comportamental (XAI/SHAP), sugerindo que a robustez Byzantina clássica (tolerância teórica de um terço a metade dos nós corrompidos) é insuficiente como garantia isolada quando o cenário de ameaça de IoT admite maiorias adversariais, exigindo camadas complementares de verificação comportamental ou de reputação acumulada. Adicionalmente, a coexistência frequente de data e gradient poisoning no mesmo estudo (9/16) indica que a literatura já reconhece a insuficiência de tratar esses vetores isoladamente, já que ataques realistas em borda tendem a se manifestar em múltiplos estágios do pipeline federado.

3.2 RQ2: Arquitetura de Blockchain e Smart Contracts

A RQ2 investiga como a blockchain e os smart contracts são arquitetados para auditar atualizações, validar gradientes e isolar nós atacantes. Os dados extraídos permitem identificar cinco funções recorrentes de Smart Contracts na amostra: automação de registro/autenticação, presente com granularidade variável em 10/16 estudos; disparo de detecção de anomalias e remoção de outliers (OLAPOJOYE et al., 2025; ROY; SINGHA; BERA, 2026, 2025; PENG et al., 2026; NIU et al., 2025); execução da própria lógica de agregação do modelo global, substituindo o agregador central (ISSA et al., 2025; KASYAP; TRIPATHY, 2024; FENG et al., 2022; PENG et al., 2026); gerenciamento de incentivos econômicos e penalização via staking (NIU et al., 2025; PENG et al., 2026; CHENG et al., 2022); e controle de acesso baseado em políticas, como ABAC (Attribute-Based Access Control) ou esquemas multiassinatura de limiar (RAHMAN; WANG; WEI, 2025; FENG et al., 2022).

Quanto aos algoritmos de detecção integrados on-chain, observa-se diversidade técnica considerável: métodos estatísticos de distância, como o filtro por Amplitude Mediana Absoluta (MAD) combinado a clustering hierárquico aglomerativo (ISSA et al., 2025), ou o filtro por Intervalo Interquartil (IQR), com pesos classificados como maliciosos quando fora do intervalo e sustentado por garantia probabilística via desigualdade DKW (OLAPOJOYE et al., 2025). Já Wei et al. (2026) combina Transformada Discreta de Cosseno (DCT) com similaridade de cosseno para filtrar o ruído de privacidade diferencial antes da detecção, abordagem que resolve o conflito frequentemente negligenciado entre privacidade e detectabilidade estatística de outliers. Técnicas criptográficas – criptografia homomórfica, ZKP, secret sharing – aparecem em 5/16 estudos, sendo notável a opção de Peng et al. (2026) por RSA assimétrico em vez de criptografia homomórfica, priorizando eficiência computacional em detrimento de propriedades criptográficas mais fortes. Abordagens baseadas em Inteligência Artificial Explicável (XAI) em Roy, Singha e Bera (2026), representam uma inovação pouco frequente na amostra, permitindo não apenas detectar o cliente malicioso, mas também identificar o rótulo-alvo do ataque de backdoor. Por fim, Alharbi (2025) propõe detecção via amplitude de estado quântico, ainda inteiramente simulada.

O achado mais relevante para RQ2 é que a robustez da defesa não decorre apenas da presença de Smart Contracts, mas da combinação entre tipo de rede, algoritmo de detecção e política de agregação. Estudos que integram múltiplas camadas – identidade, estatística e reputação –, como Issa et al. (2025) e Peng et al. (2026), reportam desempenho superior sob frações extremas de nós maliciosos, ao passo que soluções de camada única, puramente preventivas por identidade (FENG et al., 2022) ou puramente estatísticas sem

reputação acumulada (OLAPOJOYE et al., 2025), apresentam maior vulnerabilidade residual a ataques sofisticados ou coordenados ao longo de múltiplas rodadas.

3.3 RQ3: Provisionamento, Registro e Gestão de Identidade

A RQ3 busca compreender o papel dos mecanismos de identidade na prevenção da entrada de nós maliciosos, mas os dados revelam um espectro de maturidade tecnológica amplo e desigual. No extremo mais robusto, encontram-se implementações baseadas em infraestrutura de chave pública (PKI) com certificados X.509 emitidos por autoridades certificadoras via Hyperledger Fabric CA e MSP (RAHMAN; WANG; WEI, 2025; KASYAP; TRIPATHY, 2024), incluindo esquemas hierárquicos com uma CA por instituição participante em cenários multi-hospitalares (ROY; SINGHA; BERA, 2026). Em patamar intermediário, destacam-se esquemas de criptografia de curva elíptica (ECC) certificateless com rotação de pseudônimos e rekeying via Teorema Chinês do Resto, eliminando key escrow e garantindo sigilo forward/backward (SALIM; NAAZ; CHOI, 2025); gerenciamento de identidade federada cross-domain via emparelhamento bilinear e Key Generation Center por domínio (FENG et al., 2022); e pseudo-identidades leves baseadas em hash sem PKI tradicional (MUN et al., 2025). No extremo mais experimental, Alatawi (2025) emprega assinaturas digitais pós-quânticas (CRYSTALS-Dilithium) sob autenticação zero-trust, e Alharbi (2025) introduz "provas de confiança quânticas" – abordagens tecnicamente avançadas, porém validadas apenas em simulação.

Um resultado recorrente é a existência de uma lacuna significativa de detalhamento: 5/16 estudos (31%) – Olapojoye et al. (2025), Cheng et al. (2022), Wei et al. (2026), Lu et al. (2026) e, parcialmente, Niu et al. (2025) – não especificam um protocolo formal de registro, PKI ou autenticação robusta, limitando-se a assinaturas digitais genéricas ou funções de registro sem verificação aprofundada de identidade, revelando que parte da literatura trata o provisionamento como aspecto acessório. Esse desequilíbrio é reforçado pelo caso de Rahman, Wang e Wei (2025), cujo mecanismo ABAC filtra participação exclusivamente por atributos de identidade e contexto, sem avaliar o conteúdo da atualização do modelo; os próprios autores reconhecem explicitamente que clientes autorizados, porém maliciosos, não são detectados pelo mecanismo, exigindo defesas complementares que o estudo não implementa.

Em termos de maturidade tecnológica (TRL), a maioria dos estudos valida suas propostas por simulação sobre datasets padronizados (MNIST, CIFAR-10, Fashion-MNIST, ToN-IoT, SWaT/WADI), em estágio de prova de conceito (TRL 3 a 4). Apenas três estudos avançam para testes em hardware embarcado real, aproximando-se de um TRL 5: Rahman, Wang e Wei (2025), com cinco unidades heterogêneas; Feng et al. (2022), em Raspberry Pi 4; e, de forma indireta via estimativas de custo, Salim, Naaz e Choi (2025). A implementação quântica de Alharbi (2025) é inteiramente simulada via IBM Qiskit Aer, com validade limitada em hardware NISQ real de apenas 5 qubits, o que posiciona inovações criptográficas de ponta ainda distantes da viabilidade de implantação em campo para redes IoT restritas.

3.4 RQ4: Algoritmos e Técnicas de Detecção de Anomalias Integrados à Blockchain

A RQ4 aprofunda o mecanismo técnico por trás da validação on-chain descrita na RQ2, focando especificamente nos algoritmos executados para detectar anomalias nos gradientes, agrupáveis em seis famílias não mutuamente excludentes. Métricas estatísticas clássicas de distância aparecem em 9/16 estudos, incluindo filtro por Desvio Absoluto Mediano (MAD) sobre a norma L2 combinado a clustering hierárquico aglomerativo (ISSA et al., 2025), o filtro por Intervalo Interquartil sustentado pela desigualdade DKW (OLAPOJOYE et al., 2025), a similaridade de cosseno contra a mediana coordenada a coordenada

(SALIM; NAAZ; CHOI, 2025) e os pesos euclidiano/cosseno combinados no algoritmo particionamento vertical seguro, o VPSA (KASYAP; TRIPATHY, 2024). A agregação Bizantina robusta com clustering inclui o agrupamento hierárquico em duas features estatísticas (ISSA et al., 2025), o k-means sobre coeficientes de baixa frequência de uma Transformada Discreta de Cosseno (DCT) em Wei et al. (2026) – técnica que filtra especificamente o ruído de privacidade diferencial antes da detecção – e a classificação por agrupamento de centroides em (CHENG et al., 2022).

A criptografia para preservação de privacidade acoplada à detecção aparece sob forma de Differentially Private SGD com ruído laplaciano somado ao gradiente antes do armazenamento on-chain (WANG et al., 2024), secret sharing/mascaramento homomórfico em (FENG et al., 2022) e VPSA (KASYAP; TRIPATHY, 2024). Já a XAI é uma abordagem pouco frequente na amostra, representada apenas por Roy, Singha e Bera (2026), que combinam atribuição aditiva SHAP por classe com Divergência de Jensen-Shannon entre a distribuição de cada cliente e uma referência do modelo global, permitindo não apenas detectar o cliente malicioso, mas também identificar o rótulo-alvo do ataque de backdoor. Da mesma forma, a computação quântica aplicada à detecção é exclusiva de Alharbi (2025), que calcula um anomaly score como o módulo ao quadrado do produto interno entre o estado quântico de referência e o estado atual do dispositivo, complementado por uma Quantum Graph Attention Network para modelagem de confiança. Por fim, areputação probabilística acumulada aparece através de atualização via distribuição Beta combinando feedback de acurácia com o score anterior (MUN et al., 2025) e de fórmulas de reputação/contribuição baseadas em taxa de rejeição em validação cruzada (NIU et al., 2025).

Um achado relevante é que nem todos os estudos avaliam o conteúdo do gradiente: Rahman, Wang e Wei (2025), via ABAC, e, em grande medida, Feng et al. (2022), via autenticação multiassinatura, implementam defesa puramente preventiva/identitária, sem qualquer métrica estatística sobre os pesos submetidos – limitação reconhecida pelos próprios autores como central ao trabalho. Isso reforça, por via distinta, o mesmo padrão já identificado na RQ3: defesas de conteúdo e defesas de identidade raramente são tratadas como complementares obrigatórias dentro de uma mesma arquitetura, ainda que a evidência sugira que a combinação das duas produza resultados mais robustos do que qualquer uma isoladamente.

3.5 RQ5: Trade-offs entre Resiliência e Eficiência Computacional e Energética

A RQ5 investiga os impactos quantificados entre resiliência a ataques de envenenamento e eficiência operacional de dispositivos IoT restritos, evidenciando a fronteira de Pareto entre robustez defensiva e leveza computacional. No extremo de maior eficiência, destacam-se os mecanismos baseados em Criptografia de Curva Elíptica (ECC), que impõem sobrecarga mínima ao ecossistema: conforme Salim, Naaz e Choi (2025), a autenticação consome entre 1,6 ms e 1,9 ms, com overhead de metadados de apenas 68 bytes; de maneira análoga, Mun et al. (2025) reportam tempos de geração e verificação de assinatura de 0,154 ms e 0,225 ms, respectivamente, sob um overhead de comunicação de 212 bytes. Essa eficiência, contudo, tem como contrapartida detecção restrita ao plano da identidade e autenticidade criptográfica, com baixa granularidade para a inspeção direta do conteúdo dos gradientes.

Quanto ao consenso e à vazão, a otimização das arquiteturas de rede mitiga diretamente os gargalos de sincronização: o sharding hierárquico proposto por Wei et al. (2026) reduz a latência de transação em até 90,8% frente a arranjos SPDL com 100 nós, enquanto o consenso Raft aprimorado de Lu et al. (2026) é 45,95 a 53,73% mais rápido que o Proof-of-Work e de 26,25 a 35,08% mais rápido que o PBFT, reduzindo o consumo energético em cerca de 31% em relação ao BFLC e em aproximadamente 44% frente ao FL tradicional (Vanilla FL). Paralelamente, Salim, Naaz e Choi (2025) alcançaram uma vazão de 1800 tx/s com latência de 58 ms por bloco. Em contraposição, a criptografia homomórfica plena revela custos

proibitivos para dispositivos de borda: o esquema baseado em Paillier avaliado por Kasyap e Tripathy (2024) demanda mais de 2.000 s para criptografia e 60 s para decriptografia, contra os 1,32 s exigidos por esquemas de agregação sem homomorfismo pleno (VPSA). Em linha semelhante, o cálculo de explicabilidade via XAI/SHAP em Roy, Singha e Bera (2026) satura o processamento de CPU em cerca de 100%, obrigando o deslocamento dessa etapa operacional da borda para a camada de fog.

A escalabilidade é ainda limitada pelo custo assintótico não linear de algoritmos de filtragem avançados. A etapa de clustering de Issa et al. (2025), com complexidade $O(M^2 \log M)$, consolida-se como gargalo computacional em cenários de larga escala, enquanto a atribuição de contribuição via índice de Banzhaf em Peng et al. (2026), de complexidade O(2^N), torna-se inviável para execução on-chain acima de aproximadamente 20 participantes. Ainda assim, Peng et al. (2026) alcançam redução de 64,7% no tempo de treinamento e de 80,3% no volume de batches frente ao treinamento local tradicional, e Alharbi (2025) mantém o overhead da blockchain abaixo de 8,5% do custo total em 500 dispositivos.

Em síntese quantitativa, a literatura evidencia que a garantia de robustez defensiva sob frações adversariais elevadas (acima de 40–50%) requer camadas adicionais de validação cujo custo computacional cresce de forma não linear em função do número de participantes. Em contrapartida, os esquemas mais leves escalam adequadamente em nós restritos, mas concentram sua defesa na verificação de identidade e reputação probabilística simples. A ausência de propostas que conciliem simultaneamente alta resiliência contra maiorias adversariais, custo computacional aceitável e validação empírica em hardware IoT real estabelece uma lacuna de pesquisa central no estado da arte.

3.6 Discussão e Implicações

O cruzamento das cinco RQs revela uma literatura em duas velocidades: mecanismos de defesa baseados em conteúdo – estatísticos, criptográficos, comportamentais e, mais recentemente, explicáveis via XAI – apresentam diversidade técnica e rigor formal consideráveis, enquanto mecanismos de provisionamento e identidade permanecem comparativamente subdesenvolvidos ou tratados de forma genérica em quase um terço da amostra. Essa assimetria evidencia uma falha estrutural nas propostas atuais. Por um lado, a negligência com a identidade fragiliza o controle de acesso e a rastreabilidade da rede; por outro, a adoção de filtros identitários isolados mostra-se incapaz de deter nós maliciosos previamente credenciados (RAHMAN; WANG; WEI, 2025). Assim, consolida-se a necessidade de arquiteturas de defesa em camadas – que articulem autenticação forte na entrada e inspeção contínua de gradientes e reputação –, capazes de conciliar a resiliência contra ataques com a eficiência operacional exigida em ambientes IoT.

Conclusão

Este Mapeamento Sistemático da Literatura teve como objetivo identificar, classificar e sintetizar as evidências científicas sobre arquiteturas que integram Aprendizado Federado, blockchain e smart contracts para a mitigação de ataques de envenenamento em ecossistemas de Internet das Coisas, a partir da análise de 16 estudos primários publicados entre 2022 e 2026. Os achados relativos à RQ1 evidenciaram uma taxonomia predominantemente bimodal de ataques, com data poisoning e model/gradient poisoning respondendo por três quartos da amostra cada, frequentemente coexistentes no mesmo pipeline de treinamento e complementados por vetores especializados como backdoor, Sybil e inversão de gradiente. Quanto à RQ2, verificou-se o predomínio de blockchains de consórcio permissionadas (81,3%), nas quais

os smart contracts automatizam registro, detecção de anomalias, agregação descentralizada e penalização. Já a RQ3 revelou uma assimetria interessante observada neste mapeamento: enquanto implementações de ponta empregam PKI/X.509, ECC certificateless e assinaturas pós-quânticas, quase um terço da amostra não especifica protocolo formal de identidade. Por fim, os algoritmos de detecção mapeados na RQ4 demonstraram diversidade técnica considerável (de filtros estatísticos a clustering, criptografia homomórfica, XAI e computação quântica), enquanto a RQ5 confirmou uma limitação entre robustez defensiva e eficiência computacional, na qual mecanismos leves baseados em ECC escalam adequadamente, mas oferecem detecção rasa, enquanto camadas criptográficas ou comportamentais mais robustas impõem sobrecarga significativa a dispositivos de borda.

Pode-se entender que a principal contribuição deste estudo reside no mapeamento de uma literatura em duas velocidades: a sofisticação algorítmica da detecção de anomalias por conteúdo avança consideravelmente mais rápido do que a maturidade dos protocolos de provisionamento e identidade, os quais permanecem subdesenvolvidos ou tratados de forma genérica em parcela expressiva dos estudos. Essa assimetria é agravada pela evidência de que filtros de identidade isolados são insuficientes contra adversários já credenciados, assim como defesas puramente estatísticas negligenciam o vetor de entrada malicioso na rede. Consolida-se, portanto, a tese de que arquiteturas de defesa em camadas, combinando identidade robusta, validação estatística/comportamental e reputação acumulada, constituem o caminho mais promissor para a resiliência sob frações adversariais elevadas, superando soluções de camada única ao custo de maior complexidade arquitetural.

Referências

AL-FUQAHA, A. et al. Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications. IEEE Communications Surveys & Tutorials, v. 17, n. 4, p. 2347-2376, 2015. DOI: https://doi.org/10.1109/COMST.2015.2444095.

ALATAWI, M. N. EdgeGuard-IoT: 6G-Enabled Edge Intelligence for Secure Federated Learning and Adaptive Anomaly Detection in Industry 5.0. Computers, Materials and Continua, v. 85, n. 1, p. 695-727, 2025. DOI: https://doi.org/10.32604/cmc.2025.066606.

ALHARBI, S. QuantumTrust-FedChain: A Blockchain-Aware Quantum-Tuned Federated Learning System for Cyber-Resilient Industrial IoT in 6G. Future Internet, v. 17, n. 11, p. 493, 2025. DOI: https://doi.org/10.3390/fi17110493.

ATZORI, L.; IERA, A.; MORABITO, G. The Internet of Things: A survey. Computer Networks, v. 54, n. 15, p. 2787-2805, 2010. DOI: https://doi.org/10.1016/j.comnet.2010.05.010.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). Diário Oficial da União: Brasília, DF, 2018.

CHENG, X. et al. A Blockchain-Empowered Cluster-Based Federated Learning Model for Blade Icing Estimation on IoT-Enabled Wind Turbine. IEEE Transactions on Industrial Informatics, v. 18, n. 12, p. 9184-9195, 2022. DOI: https://doi.org/10.1109/TII.2022.3159684.

DYBÅ, T.; DINGSØYR, T. Empirical studies of agile software development: A systematic review. Information and Software Technology, v. 50, n. 9-10, p. 833-859, 2008. DOI: https://doi.org/10.1016/j.infsof.2008.01.006.

FENG, C. et al. Blockchain-Empowered Decentralized Horizontal Federated Learning for 5G-Enabled UAVs. IEEE Transactions on Industrial Informatics, v. 18, n. 5, p. 3582-3592, 2022. DOI: https://doi.org/10.1109/TII.2021.3116132.

ISSA, W. et al. DT-BFL: Digital Twins for Blockchain-enabled Federated Learning in Internet of Things networks. Ad Hoc Networks, v. 178, art. 103934, 2025. DOI: https://doi.org/10.1016/j.adhoc.2025.103934.

KAIROUZ, P. et al. Advances and Open Problems in Federated Learning. Foundations and Trends in Machine Learning, v. 14, n. 1-2, p. 1-210, 2021. DOI: https://doi.org/10.1561/2200000083.

KASYAP, H.; TRIPATHY, S. Privacy-preserving and Byzantine-robust Federated Learning Framework using Permissioned Blockchain. Expert Systems with Applications, v. 238, pt. E, art. 122210, 2024. DOI: https://doi.org/10.1016/j.eswa.2023.122210.

KITCHENHAM, B.; CHARTERS, S. Guidelines for performing Systematic Literature Reviews in Software Engineering. Keele: Keele University, 2007. (EBSE Technical Report, EBSE-2007-01).

LI, T. et al. Federated Learning: Challenges, Methods, and Future Directions. IEEE Signal Processing Magazine, v. 37, n. 3, p. 50-60, 2020. DOI: https://doi.org/10.1109/MSP.2020.2975749.

LU, J. et al. Energy-efficient serverless federated learning with blockchain-enhanced optimized raft consensus. Future Generation Computer Systems, v. 182, art. 108473, 2026. DOI: https://doi.org/10.1016/j.future.2026.108473

MCMAHAN, B. et al. Communication-Efficient Learning of Deep Networks from Decentralized Data. In: INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE AND STATISTICS, 20., 2017. Proceedings... PMLR, 2017. p. 1273-1282.

MUN, H. et al. Emerging Blockchain and Reputation Management in Federated Learning: Enhanced Security and Reliability for Internet of Vehicles (IoV). IEEE Transactions on Vehicular Technology, v. 74, n. 2, p. 1893-1908, 2025. DOI: https://doi.org/10.1109/TVT.2024.3456852.

NIU, H. et al. DSFL: a blockchain-based data sharing and federated learning framework. Applied Intelligence, v. 55, art. 474, 2025. DOI: https://doi.org/10.1007/s10489-025-06400-w.

OLAPOJOYE, R. et al. FedECPA: An Efficient Countermeasure Against Scaling-Based Model Poisoning Attacks in Blockchain-Based Federated Learning. Sensors, v. 25, art. 6343, 2025. DOI: https://doi.org/10.3390/s25206343.

PAGE, M. J. et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. Journal of Clinical Epidemiology, v. 134, p. 178-189, 2021. DOI: https://doi.org/10.1016/j.jclinepi.2021.03.001.

PENG, X. et al. A privacy-enhanced multi-party industrial control systems collaboration framework. Reliability Engineering & System Safety, v. 273, art. 112318, 2026. DOI: https://doi.org/10.1016/j.ress.2026.112318.

PETERSEN, K.; VAKKALANKA, S.; KUZNIARZ, L. Guidelines for conducting systematic mapping studies in software engineering: An update. Information and Software Technology, v. 64, p. 1-18, 2015. DOI: https://doi.org/10.1016/j.infsof.2015.03.007.

JUNIOR, A. J. K. Q. et al. A Systematic Review on the Practicality of Poisoning Defenses in Federated IoT Systems. IEEE Access, v. 13, p. 211109-211137, 2025. DOI: https://doi.org/10.1109/ACCESS.2025.3637093.

RAHMAN, S.; WANG, Y.; WEI, B. Trust at the Edge: ABAC-Secured Federated Learning for Smart Home Access Control Using Blockchain. IEEE Access, v. 13, p. 175094-175108, 2025. DOI: https://doi.org/10.1109/ACCESS.2025.3618270.

RODRÍGUEZ-BARROSO, N. et al. Survey on federated learning threats: Concepts, taxonomy on attacks and defences, experimental study and challenges. Information Fusion, v. 90, p. 148-173, 2023. DOI: https://doi.org/10.1016/j.inffus.2022.09.011.

ROY, S.; SINGHA, S. K.; BERA, D. X-BDFL: Explainable Blockchain-Enabled Distributed Federated Learning Against Poisoning Attack in Healthcare. IEEE Access, v. 14, p. 2726-2741, 2026. DOI: https://doi.org/10.1109/ACCESS.2025.3648434.

SALIM, M. M.; NAAZ, F.; CHOI, K. Lightweight ECC-Based Self-Healing Federated Learning Framework for Secure IIoT Networks. Sensors, v. 25, art. 6867, 2025. DOI: https://doi.org/10.3390/s25226867.

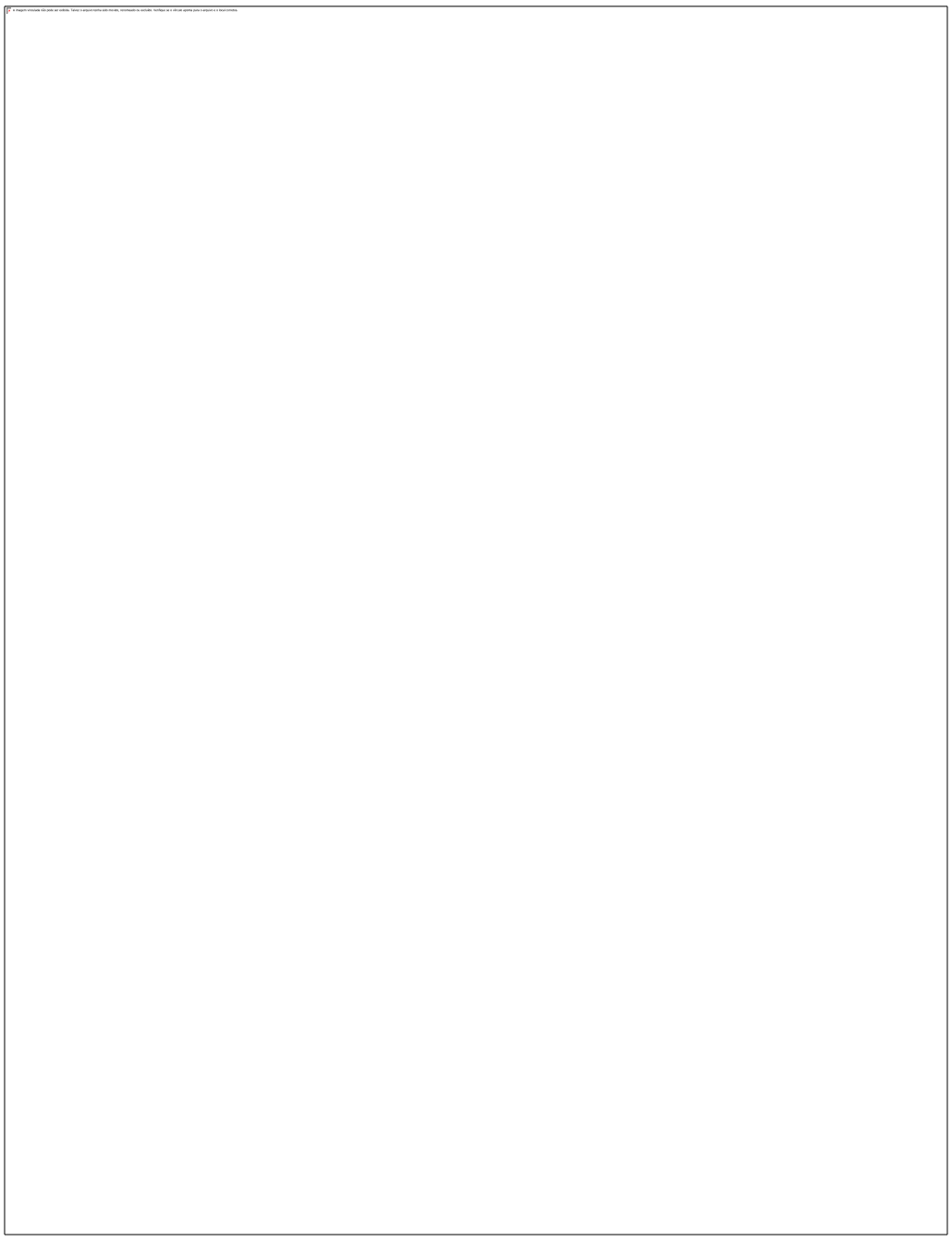
SHI, W. et al. Edge Computing: Vision and Challenges. IEEE Internet of Things Journal, v. 3, n. 5, p. 637-646, 2016. DOI: https://doi.org/10.1109/JIOT.2016.2579198.

UNIÃO EUROPEIA. Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho, de 27 de abril de 2016 (Regulamento Geral sobre a Proteção de Dados – GDPR). Jornal Oficial da União Europeia, L 119, p. 1–88, 2016.

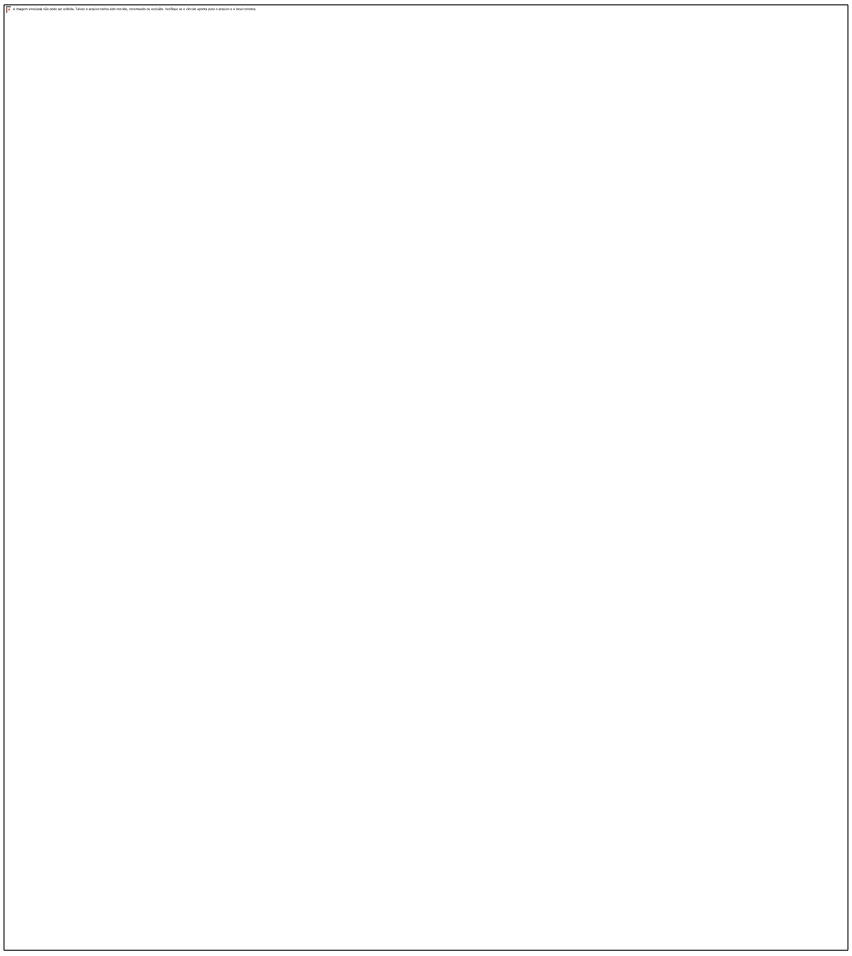
WANG, M. et al. Blockchain-Based Gradient Inversion and Poisoning Defense for Federated Learning. IEEE Internet of Things Journal, v. 11, n. 9, p. 15667-15681, 2024. DOI: https://doi.org/10.1109/JIOT.2023.3347899.

WEI, Z. et al. Secure and Trustworthy Federated Learning Framework via Hierarchical Sharding Blockchain. IEEE Internet of Things Journal, v. 13, n. 4, p. 7527-7541, 2026. DOI: https://doi.org/10.1109/JIOT.2025.3639396.

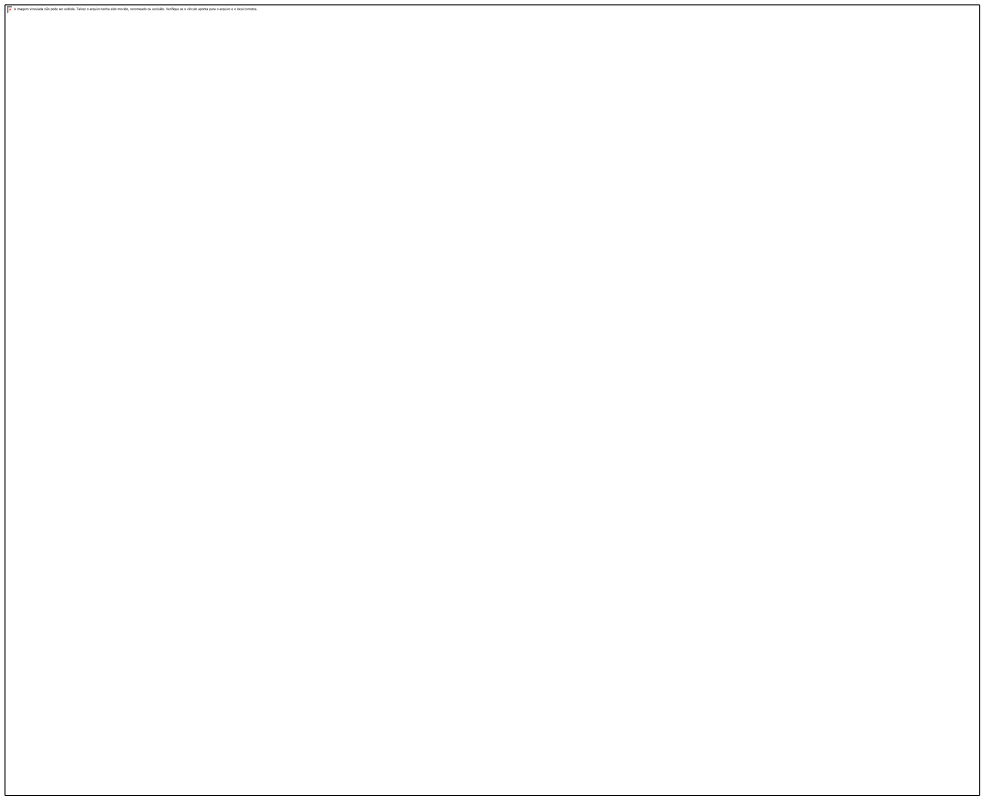
Anexos



Anexo 1 – Tabela de Critérios de Seleção



Anexo 2 – Diagrama PRISMA



Anexo 3 – Tabela de Avaliação de Qualidade

CÓDIGO: ET1653

AUTOR: LEONARDO MACHADO MOREIRA

COAUTOR: MATTEO PEIXOTO

COAUTOR: VICENTE ANGELO DE SOUSA JUNIOR

COAUTOR: NELSON ION DE OLIVEIRA

ORIENTADOR: AUGUSTO JOSE VENANCIO NETO

TÍTULO: Monitoramento do ciclo de vida de aplicações elásticas entregues por Cloud-Network Slices

Resumo

Este trabalho propõe um sistema de monitoramento transcamada para avaliar o desempenho do Minecraft — jogo sandbox sensível a atraso — sobre uma infraestrutura 5G NR real end-to-end com acesso sem fio fixo (FWA). O testbed integra um núcleo 5G privado (5GC) hospedando um servidor PaperMC gerenciado pelo controlador Crafty, e um gNB conectado a um dispositivo FWA que fornece acesso Wi-Fi a headsets Meta Quest 3. A carga de trabalho combina três bots Mineflayer com um a três headsets usados simultaneamente por jogadores. Um framework automatizado correlaciona KPIs de rede (throughput, RTT e jitter) com métricas de QoS da aplicação (ticks por segundo e duração do tick). A principal novidade é a sincronização entre rastros sem fio de baixo nível e dados de desempenho do motor de jogo de alto nível em ambiente 5G FWA. A contribuição é um framework robusto de avaliação para serviços sensíveis a atraso, que valida a viabilidade do 5G FWA como conectividade primária para jogos imersivos e identifica o impacto dos atrasos induzidos pela rede na experiência do jogador.

Palavras-chave: 5G, FWA, Minecraft, Meta Quest 3, QoE,QoS, KPIs

TITLE: Monitoring the lifecycle of elastic applications delivered by Cloud-Network Slices

Abstract

This work proposes a cross-layer monitoring system to evaluate the performance of Minecraft — a delay-sensitive sandbox game — over a real end-to-end 5G NR infrastructure using fixed wireless access (FWA). The testbed integrates a private 5G

core (5GC) hosting a PaperMC server managed by the Crafty controller, and a gNB connected to an FWA device that provides Wi-Fi access for Meta Quest 3 headsets. The workload combines three Mineflayer bots with one to three headsets used simultaneously by players. An automated framework correlates network KPIs (throughput, RTT, and jitter) with application-specific QoS metrics (ticks per second and tick duration). The main novelty is the synchronization of low-level wireless network traces with high-level game engine performance data within a 5G FWA environment. The contribution is a robust evaluation framework for delay-sensitive services, validating the feasibility of 5G FWA as a primary connectivity bridge for immersive gaming and identifying the impact of network-induced delays on player experience.

Keywords: 5G, FWA, Minecraft, QoE, Meta Quest 3,minecraft-prometheus-exporter

Introdução

O ecossistema do 5G New Radio (NR) foi concebido para atender requisitos que se estendem além das altas taxas de transmissão, permitindo assim serviços emergentes como a realidade estendida (XR). A XR é um termo abrangente que engloba realidade virtual (VR), realidade aumentada (AR) e realidade mista (MR), que integram fluidamente ambientes físicos e digitais. O mercado de XR, avaliado em US$ 142,39 bilhões em 2023, projeta-se crescer significativamente (US$ 1,069 bilhão em 2030 [1] e US$ 2,127,81 bilhões em 2034 [2]), impulsionado por investimentos substanciais nos setores automotivo, de jogos e entretenimento [3]–[6]. Entre eles, jogos sandbox altamente interativos como o Minecraft se destacam como aplicações demandantes que requerem latência ultrabaixa e alta confiabilidade para manter a imersão do jogador e a sincronização em um mundo digital compartilhado. Apesar de seu potencial, a implantação generalizada de serviços de XR em redes móveis permanece desafiadora devido à natureza instável do meio sem fio [7]. Aplicações de RV interativas, como jogos imersivos, impõem requisitos rigorosos de qualidade de serviço (QoS); por exemplo, latências superiores a 20 ms podem comprometer a experiência e levar a enjoo ou perda de interatividade [8]–[10]. Redes comerciais 4G frequentemente não conseguem atingir esses limiares, tornando infraestruturas 5G personalizadas essenciais para prover o isolamento e desempenho necessários [8].
Além disso, displays comerciais populares montados na cabeça (HMDs), como o Meta Quest 3, atualmente não possuem suporte nativo à conectividade celular, criando uma lacuna significativa para a adoção móvel de XR [9], [10].
Para preencher essa lacuna de conectividade, o acesso sem fio fixo (FWA) emergiu como uma solução primária, atuando

como um ponto de acesso Wi-Fi com conectividade 5G para integrar dispositivos de RV na rede móvel. Embora estudos recentes tenham explorado a viabilidade de transmissão de vídeo em 360 graus sobre FWA 5G [7], [11], [12], o desempenho de jogos multiplayer em tempo real, onde o estado do mundo deve ser sincronizado entre vários usuários, permanece uma área crítica de pesquisa. Especificamente, o enlace ascendente frequentemente torna-se um gargalo significativo ao multiplexar atualizações frequentes de pose e comandos do jogador de vários usuários simultâneos [12]. Este estudo apresenta uma análise de desempenho do Minecraft operando sobre uma infraestrutura real de acesso sem fio fixo (FWA) 5G ponta-a-ponta (E2E). O ambientes de teste experimental proposto hospeda um servidor Minecraft dentro de um núcleo privado 5G (5GC), entregando o serviço via um gNodeB (gNB) para um dispositivo FWA que fornece conectividade para um a três jogadores usando headsets Meta Quest 3 simultaneamente. A avaliação utiliza um framework automatizado inovador que correlaciona Indicadores de Desempenho Chave da Rede (KPIs), como tempo de ida e volta (RTT) capturado via script Tshark, com métricas de qualidade da experiência (QoS) no nível da aplicação, como ticks por segundo (TPS), extraídas através do plugin minecraft-prometheus-exporter. As principais contribuições deste trabalho são:

• Projeto e implementação de uma plataforma de teste E2E 5G real baseada em tecnologias de código aberto para suportar jogos imersivos multiusuário.
• Validação do 5G FWA como uma ponte de conectividade para múltiplos headsets Meta Quest 3, superando a falta de suporte nativo 5G.
• Introdução de uma metodologia de monitoramento cruzado de camadas automatizada que sincroniza a coleta de KPIs de rede de baixo nível com métricas específicas do motor de jogo via o plugin minecraft-prometheus-exporter.
• Avaliação experimental do impacto multiusuário (de 1 a 3 jogadores simultâneos e 3 bots Mineflayer) na estabilidade da rede e na experiência resultante do jogador em ambientes sandbox sensíveis à latência.

Metodologia

O testbed é uma rede 5G privada end-to-end que integra 5GC, gNB e um gateway FWA para conectar headsets VR. O servidor Minecraft roda no mesmo host físico do núcleo 5G, decisão tomada para minimizar a latência de transporte.

Infraestrutura e serviço. O núcleo 5G usa Open5GS v2.7.0 sobre Ubuntu Server 22.04, orquestrado com Docker, hospedando o Crafty Controller v4, que gerencia um servidor Paper MC v1.21.5 (JRE/JDK 21+). O suporte a VR vem dos plugins Vivecraft Spigot Extension e FlyCraft, este último responsável por regular a velocidade de voo dos jogadores. O RAN emprega srsRAN v25.04 com uma USRP B210 configurada em 51 PRBs, 20 MHz, faixa de 3740–3760 MHz, MIMO 2x2 e 256-QAM. O gNB se conecta a um gateway FWA TP-Link NX620v, que fornece Wi-Fi aos headsets Meta Quest 3 rodando QuestCraft v6.0.0. Um Motorola Moto G50 5G atua como UE nos testes de caracterização.

Framework de monitoramento. A coleta é centralizada em um dashboard Grafana 13.1.3, alimentado por três agentes: Prometheus 3.13.2 na camada de aplicação, Zabbix Agent 7.2.15 para desempenho do 5GC e um coletor nativo de KPIs de RAN. O Crafty Controller expõe endpoints Prometheus com status do host e dos servidores. No Paper MC, o plugin minecraft-prometheus-exporter (v3.1.2) coleta TPS, tick duration e métricas de chunks, enquanto o prometric (v0.1.0) mede o ping por jogador. No 5GC, scripts Python acionados via Zabbix geram telemetria JSON estruturada. O gNB envia streams UDP/JSON ao srsRAN Metrics Server v1.7.3, que persiste os dados em InfluxDB 2.7.12 para visualização no Grafana. Um script próprio, exportador_rede.js, mede RTT e throughput em nível de fio.

Geração automatizada de carga. O gerador bots.js, escrito em Node.js v20.20.2 com Mineflayer v4.25.0, simula múltiplos jogadores interagindo no mundo, permitindo medir o impacto da densidade de usuários sobre os KPIs de rede e a estabilidade do jogo.

A avaliação é dividida em dois conjuntos de experimentos.

Conjunto 1 — Caracterização do sistema. Caracteriza o desempenho da rede 5G subjacente para estabelecer limites superiores de throughput e limites inferiores de latência. Usa iperf3 para estimativa de throughput TCP e ping para medição de RTT. Duas configurações são comparadas, de modo a isolar a sobrecarga introduzida pela ponte FWA: (i) conectividade 5G direta, com o Moto G50 conectado diretamente ao gNB; e (ii) conectividade com ponte FWA, com o UE acessando o TP-Link NX620v por Wi-Fi. Cada configuração é testada em 10 repetições de 5 minutos, garantindo relevância estatística e captura de flutuações do meio sem fio.

Conjunto 2 — Cenário de jogos imersivos. Avalia o serviço Minecraft sob carga, usando o framework de monitoramento proposto. A carga combina três agentes bots.js com um a três operadores humanos em Meta Quest 3 com QuestCraft, formando três casos de teste independentes: 3 bots + 1 jogador VR, 3 bots + 2 jogadores VR e 3 bots + 3 jogadores VR. Para maximizar a carga sobre o motor Paper MC e sobre a rede, bots e jogadores executam tarefas de exploração — movimento contínuo por áreas não renderizadas do mapa a aproximadamente 21,78 blocos/s. Isso obriga o servidor a gerar e carregar novos chunks em tempo real, elevando o uso de CPU e a transmissão de dados. Para garantir consistência entre execuções, o estado do mundo é reiniciado antes de cada experimento, limpando o cache de chunks carregados. Cada caso de teste é repetido 3 vezes, com 10 minutos por execução.

Resultados e Discussões

Esta seção apresenta os resultados experimentais do nosso estudo, organizados em três subseções principais. Primeiramente, são apresentadas as funcionalidades e a visibilidade intercamadas do painel de monitoramento do framework proposto. Em seguida, o Conjunto de Experimentos 1 caracteriza o desempenho de linha de base da rede 5G, comparando a conectividade direta com a configuração FWA-bridged. Por fim, o Conjunto de Experimentos 2 analisa o impacto das cargas de trabalho multiusuário sobre as métricas do motor de jogo Minecraft e a experiência imersiva dos jogadores.

A. Painel de Monitoramento do Framework

O framework de monitoramento proposto, visualizado através do painel Grafana ilustrado na Fig. 4, atua como uma plataforma centralizada de observabilidade para todo o ecossistema 5G FWA, da infraestrutura de núcleo à experiência do usuário final. A interface foi arquitetada para fornecer visibilidade intercamadas, permitindo a correlação da saúde da infraestrutura, desempenho do rádio e QoS no nível de aplicação. O painel é organizado em três domínios funcionais primários:

1) Fig. 4(a) (métricas 5GC): Esta visão fornece telemetria em tempo real do Open5GS e do ambiente de hospedagem. Especificamente, ela rastreia a utilização de CPU e memória para contêineres de serviço críticos, incluindo o Crafty Controller, Prometheus e o próprio Grafana. Adicionalmente, ela exibe o throughput no nível do núcleo (bytes enviados e recebidos por segundo) para refletir a capacidade da infraestrutura backbone durante experimentos de jogos imersivos.

2) Fig. 4(b) (métricas gNB): Esta seção apresenta dados de RAN de alta fidelidade coletados da pilha de protocolo srsRAN. Os KPIs de telecomunicações incluem o número de UEs ativos, taxas de bits de uplink e downlink, e a utilização média do Esquema de Modulação e Codificação (MCS). Também incorpora telemetria de taxa de erro, tais como Taxa de Erro de Bloco no Enlace Descendente (BLER) e Relatórios de Status de Buffer (BSR), que são vitais para diagnosticar latências induzidas por sem fio ou perda de pacotes que podem originar-se nos segmentos 5G ou Wi-Fi.

3) Fig. 4(c) (métricas específicas do Minecraft): O domínio final foca na camada de serviço, utilizando métricas extraídas via o exporter sladkoff/minecraft-prometheus-exporter e os plugins

prometric executados no Paper MC. Por exemplo, ela rastreia a TPS e a duração do tick/MSPT (tempo médio de processamento de estado).

A principal contribuição técnica deste painel é sua capacidade de sincronizar funcionalmente KPIs de rede de baixo nível com o desempenho do motor de jogo de alto nível. Ao alinhar dados com carimbo de tempo provenientes do gNB e 5GC com a saúde interna da simulação do servidor Minecraft, o framework identifica exatamente como flutuações no meio de rádio 5G, como mudanças no MCS ou aumento do BSR, se traduzem em degradação na camada de aplicação, como uma queda na TPS.

B. Conjunto de Experimentos 1: Caracterização da Rede 5G

Para avaliar o impacto da infraestrutura FWA, realizamos uma análise comparativa entre a conectividade 5G direta (5G Only) e a implantação com ponte (5G FWA). A Fig. 5 ilustra esses resultados através da função de distribuição cumulativa (CDF) de throughput e medidas temporais de RTT.

(a) Throughput. (b) RTT.

Fig. 5: Caracterização do desempenho de linha de base do impacto da FWA no cenário comparado com apenas 5G (Motorola Moto G50 5G): (a) CDF de capacidade de throughput; (b) medidas temporais de RTT e valores médios.

A capacidade de throughput, mostrada na Fig. 5. (a), revela uma redução significativa quando o dispositivo FWA é introduzido. Essa lacuna de desempenho representa o custo de gateway para a conversão e encaminhamento necessários dos dados entre a interface 5G NR e o ponto de acesso Wi-Fi fornecido aos headsets de RV. Ao analisar o 50° percentil (mediana) das curvas de CDF, o throughput cai de aproximadamente 73 Mbps na configuração apenas 5G para cerca de 44 Mbps no cenário 5G FWA. Isso representa uma perda de throughput de aproximadamente 40% no nível da mediana.

(a) Métricas do 5GC. (b) Métricas da gNB. (c) Métricas Específicas do Minecraft.

Fig. 4: Ilustração do Painel de Monitoramento.

Um compromisso de desempenho semelhante é observado quanto à latência na Fig. 5. (b). Embora o enlace 5G direto mantenha um RTT médio estável de 27,38 ms, a introdução da ponte FWA aumenta o RTT médio para aproximadamente 39,04 ms (aproximadamente um aumento de 44%). Estes resultados indicam que, embora o FWA efetivamente preencha a lacuna de conectividade para dispositivos de RV não celulares como o Meta Quest 3, ele introduz uma sobrecarga mensurável tanto em largura de banda quanto em atraso que deve ser considerada em aplicações imersivas altamente sensíveis.

C. Conjunto de Experimentos 2: Cenário de Jogo Imersivo

A análise do cenário de jogo imersivo, representado na Fig.

6, demonstra a correlação entre o desempenho da rede e a estabilidade do motor de jogo à medida que a densidade de usuários aumenta de um para três jogadores VR (mais três bots automatizados).

Um tick é um ciclo de execução fundamental do loop do jogo do lado do servidor do Minecraft, no qual o estado do mundo (movimento de entidades, atualizações de blocos, redstone e física) é avançado em um passo fixo [13]. O servidor visa uma duração de tick de 50 ms, resultando em um TPS nominal de 20. Como o TPS é limitado a esta taxa e só pode cair quando o processamento por tick excede 50 ms, ele serve como um proxy direto, em tempo real, para a saúde computacional do lado do servidor. Uma queda sustentada abaixo de 20 TPS indica que o servidor não consegue manter o ritmo com as demandas de simulação do mundo, manifestando-se como lag de mob/física, interações de blocos atrasadas e responsividade percebida pelo jogador degradada, independentemente das condições de rede.

Conforme mostrado na CDF da Fig. 6(a), o throughput aumenta proporcionalmente com o número de usuários ativos no mundo compartilhado. Essa tendência indica que a largura de banda disponível fornecida pela infraestrutura 5G FWA é suficiente para atender às demandas de dados de múltiplos clientes imersivos, confirmando que o throughput não é o gargalo primário nesta configuração específica.

No entanto, o impacto do aumento da carga torna-se evidente ao analisar a saúde interna do motor de jogo. A Fig. 6(b) ilustra um aumento progressivo na duração do tick à medida que mais jogadores são adicionados ao ambiente. Sob a maior carga testada (3 bots mais 3 VRs), a duração máxima do tick excede o limite crítico de 50 ms. Essa degradação de desempenho reflete diretamente os níveis de latência estabelecidos na fase de caracterização do sistema. Com a RTT de rede estando na faixa de 40 ms em condições sem carga (como mostrado na Fig. 5(b)), o tempo cumulativo de processamento e transmissão para o loop do lado do servidor começa a exceder seu orçamento nominal.

Essa relação é ainda mais validada pelas CDFs de TPS na Fig. 6(c). À medida que a carga de jogadores aumenta, os valores de TPS caem consistentemente abaixo da marca ideal de 20 TPS, entrando em níveis de desempenho inferiores conforme o motor luta para sincronizar o estado do mundo entre todas as seis entidades ativas. Ao capturar com sucesso a sincronização entre KPIs de rede de baixo nível e telemetria do motor de jogo de alto nível, estes resultados validam a eficácia do framework de monitoramento em camadas (cross-layer) proposto na identificação dos fatores específicos que comprometem a QoE em ambientes imersivos 5G.

Conclusão

Este artigo apresentou uma avaliação abrangente do Minecraft, um jogo sandbox sensível à latência, implantado sobre uma infraestrutura 5G E2E privada utilizando FWA. A contribuição central deste trabalho foi a proposta e validação de um framework de monitoramento automatizado cross-layer projetado para sincronizar KPIs de rede sem fio de baixo nível com a telemetria do motor de jogo de alto nível. Ao integrar uma 5GC privada, um gNB e um gateway FWA para fornecer conectividade para headsets Meta Quest 3, o estudo demonstrou com sucesso a viabilidade do 5G FWA como uma ponte de conectividade primária para jogos imersivos em cenários onde o suporte nativo celular atualmente não está disponível.

Os resultados experimentais proporcionaram insights críticos sobre os trade-offs de desempenho inerentes às implantações de FWA. A fase de caracterização do sistema identificou um custo de gateway composto por uma redução de 40% no throughput mediano e um aumento de 44% na RTT média (aproximadamente 12 ms) ao introduzir a ponte FWA. Embora o throughput tenha sido encontrado como suficiente para cargas de trabalho multiusuário, a análise do cenário de jogos imersivos revelou que o aumento da densidade de jogadores (até três bots e três usuários VR) impacta significativamente a saúde do lado do servidor. Especificamente, a maior carga testada empurrou a duração máxima do tick além do limite crítico de 50 ms, levando a uma degradação no TPS abaixo da taxa nominal de 20. Essas descobertas validam a eficácia do framework proposto na correlação de latências induzidas pela rede com a estabilidade do loop interno do jogo, o que é essencial

(a) Throughput.
(b) Duração do tick.
(c) Ticks por Segundo.

Fig. 6: Resultados de Validação.

para manter a imersão do jogador. Trabalhos futuros vão envolver um estudo completo sobre a análise de desempenho, incorporando todo o escorpo das métricas coletadas pelo framework de monitoramento em experimentos com maior tempo de duração. Pesquisas futuras vão explorar diversas configurações de hardware e software para implementação do gNB, comparando especificamente a USRP B210 e N310 com diferentes protocolos open stacks (OCUDU and DURANTA). Além disso, a análise objetiva de QoS apresentada neste trabalho será complementada por avaliações subjetivas de QoE envolvendo jogadores humanos reais para avaliar o impacto percebido de flutuações da rede na jogabilidade e na imersão.

Referências

[1] G. V. Research, “Extended Reality Market (2024 - 2030),” 2024. [Online]. Available: https://www.grandviewresearch.com/industry-analysis/extended-reality-xr-market-report
[2] F. Business, “Extended Reality Market Size, Share & Industry Analysis and Regional Forecast, 2026–2034,” 2026. [Online]. Available: https://www.fortunebusinessinsights.com/extended-reality-market-106637
[3] A. Clérigo, G. Silva, M. Schrapel, P. Rito, S. Sargento, and A. Vinel, “Cooperative Augmented Reality: Displaying Occluded Vehicles using V2X Communications,” in IEEE VNC, 2025.
[4] K. Kieanwatana and R. Vongvit, “Virtual reality in tourism: The impact of virtual experiences and destination image on the travel intention,” Results in Engineering, vol. 24, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2590123024018930
[5] R. Kim, “Retail After COVID-19: Use Virtual Reality to Enhance Ecommerce,” in 2022 IEEE Technology and Engineering Management Conference (TEMSCON EUROPE), 2022, pp. 118–123.
[6] S. Machała, N. Chamier-Gliszczy´nski, and T. Królikowski, “Application of AR/VR Technology in Industry 4.0.” Procedia Computer Science, vol. 207, pp. 2990–2998, 2022.
[7] E. Stafidas and F. Foukalas, “A Survey on Enabling XR Services in Beyond 5G Mobile Networks,” IEEE Access, vol. 12, pp. 59 170–59 197, 2024.
[8] M. Susloparov, A. Krasilov, and E. Khorov, “Providing High Capacity for AR/VR traffic in 5G Systems with Multi-Connectivity,” in IEEE BlackSeaCom, 2022.
[9] O. S. Peñaherrera-Pulla, C. Baena, S. Fortes, E. Baena, and R. Barco, “KQI Assessment of VR Services: A Case Study on 360-Video Over 4G and 5G,” IEEE Transactions on Network and Service Management, vol. 19, no. 4, pp. 5366–5382, 2022.
[10] T. Taleb, N. Sehad, Z. Nadir, and J. Song, “VR-Based Immersive Service Management in B5G Mobile Systems: A UAV Command and Control Use Case,” IEEE Internet of Things Journal, vol. 10, no. 6, pp. 5349–5363, March 2023.
[11] F. Hu, Y. Deng, H. Zhou, T. H. Jung, C.-B. Chae, and A. H. Aghvami, “A Vision of an XR-Aided Teleoperation System toward 5G/B5G,” IEEE Communications Magazine, vol. 59, no. 1, pp. 34–40, 2021.
[12] J. K. Sundararajan, H.-J. Kwon, O. Awoniyi-Oteri, Y. Kim, C.-P. Li, J. Damnjanovic, S. Zhou, R. Ma, Y. Tokgoz, P. Hande, T. Luo, K. Mukkavilli, and T. Ji, “Performance Evaluation of Extended Reality Applications in 5G NR System,” in 2021 IEEE 32nd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), 2021, pp. 1–7.
[13] Minecraft Wiki, “Tick,” https://minecraft.wiki/w/Tick, 2026, accessed: 2026-08-17.

Anexos

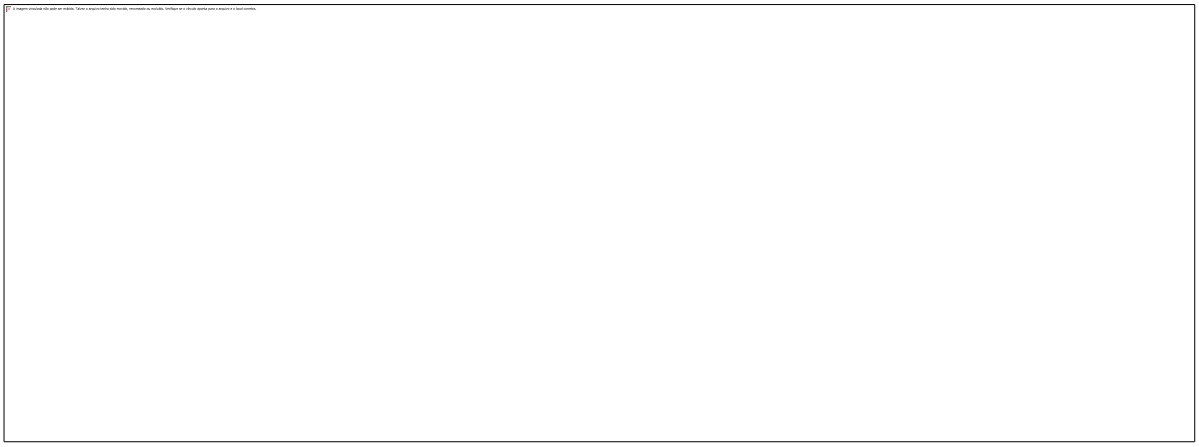


Fig. 1

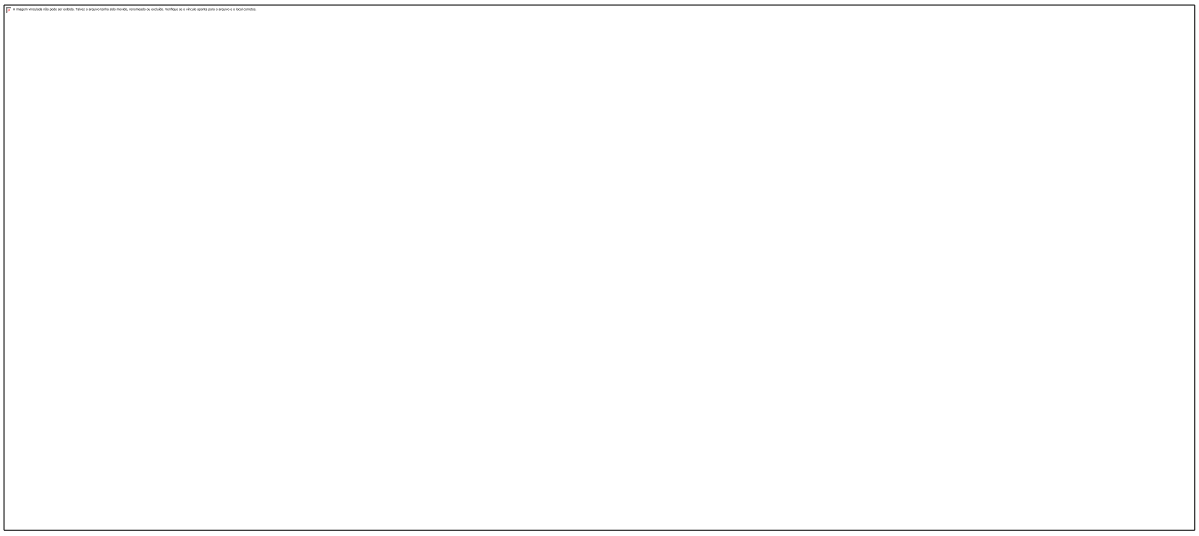


Fig. 2



Fig. 3



Fig. 4 (a)



Fig. 4 (b)



Fig. 4 (c)

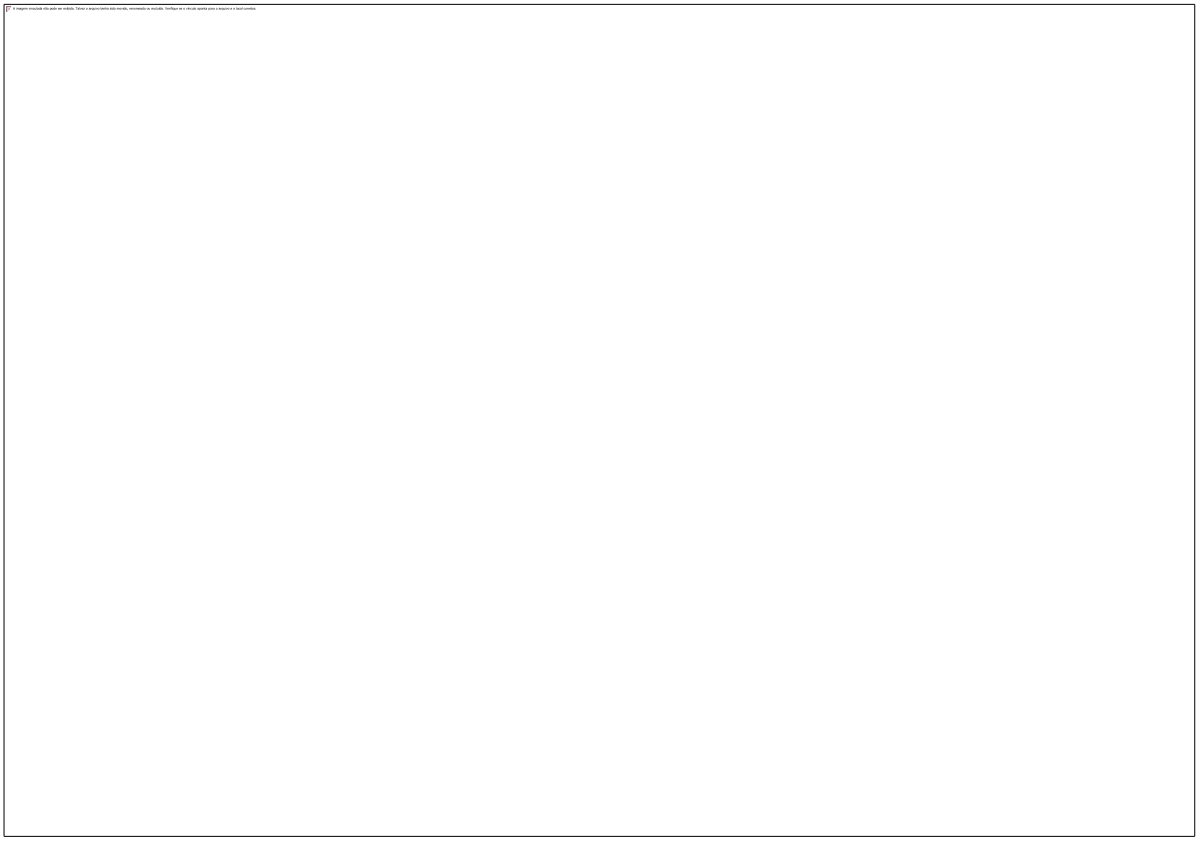


Fig. 5 (a)

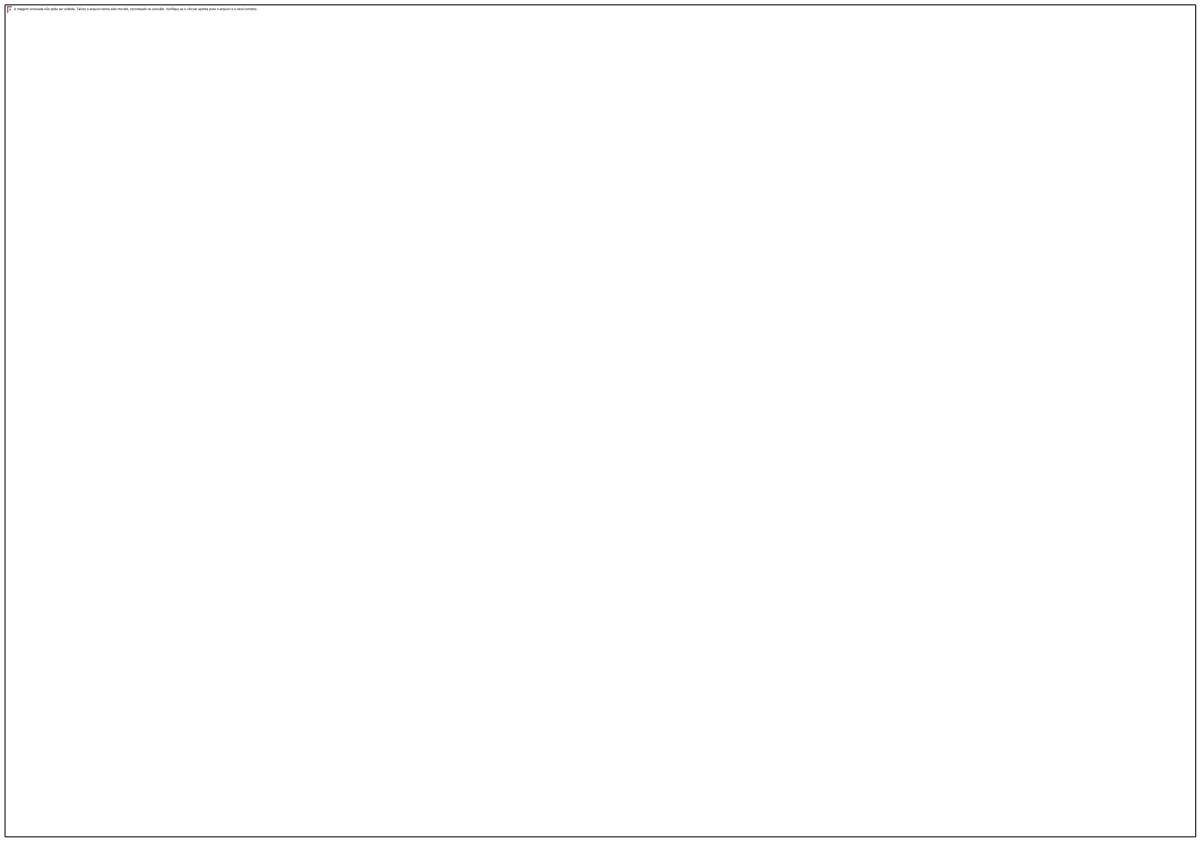


Fig. 5 (b)

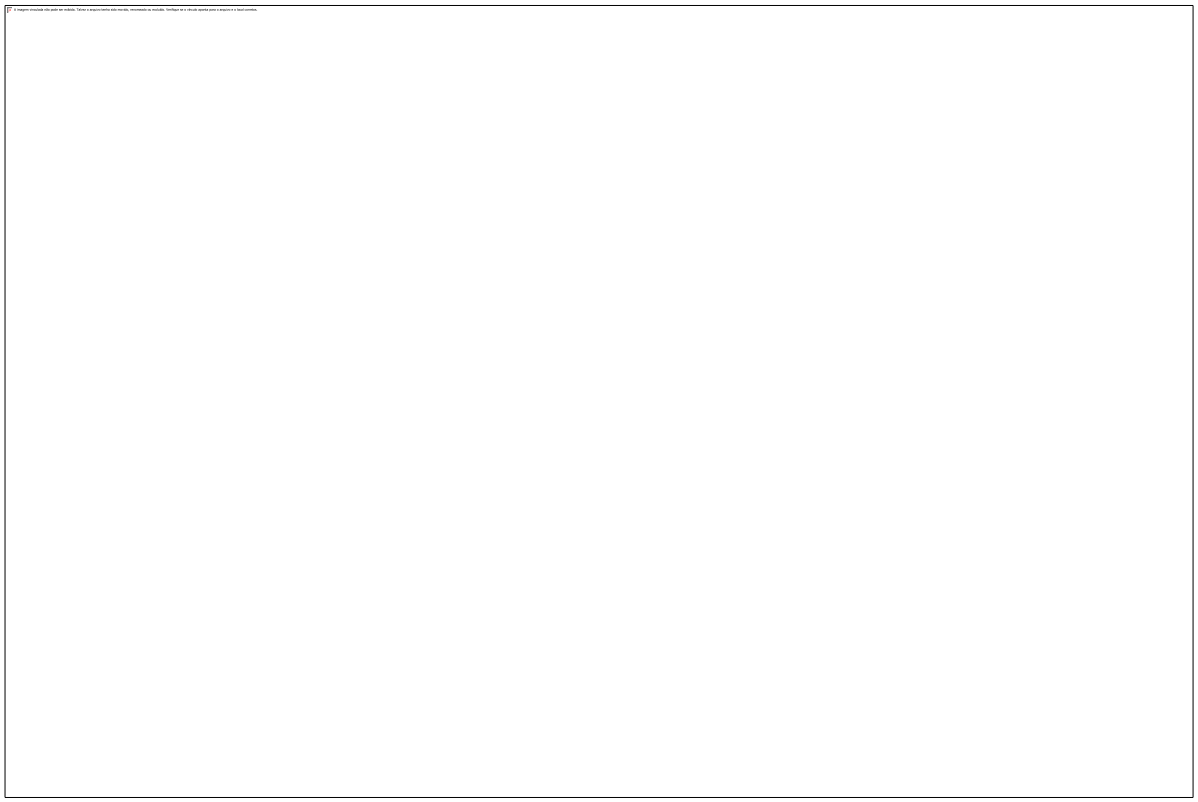


Fig. 6 (a)

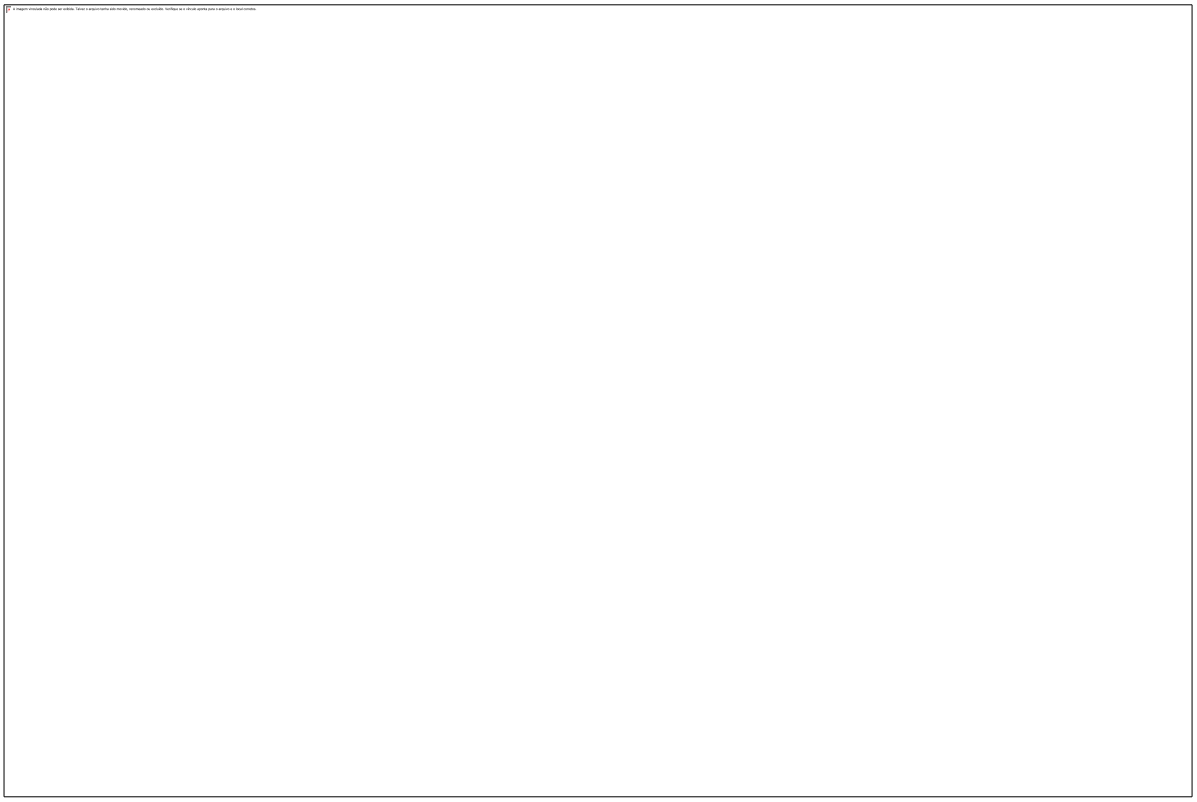


Fig. 6 (b)

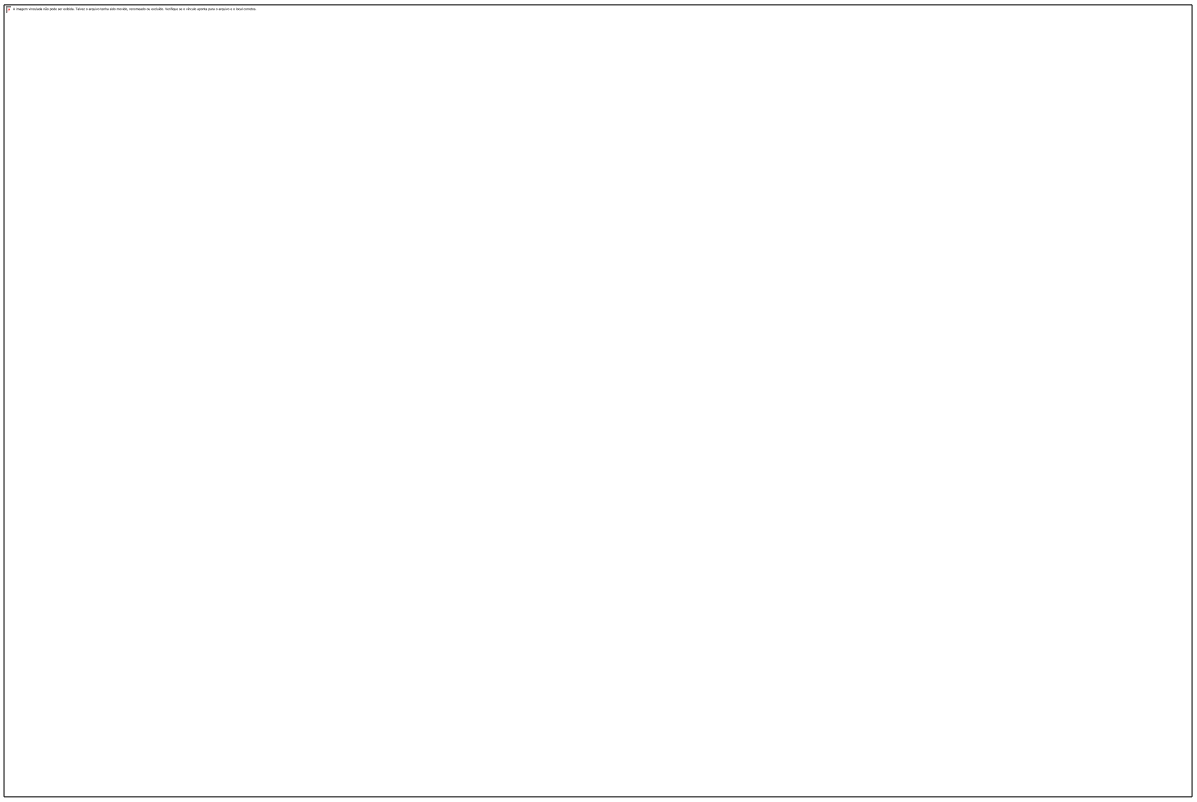


Fig 6 (c)

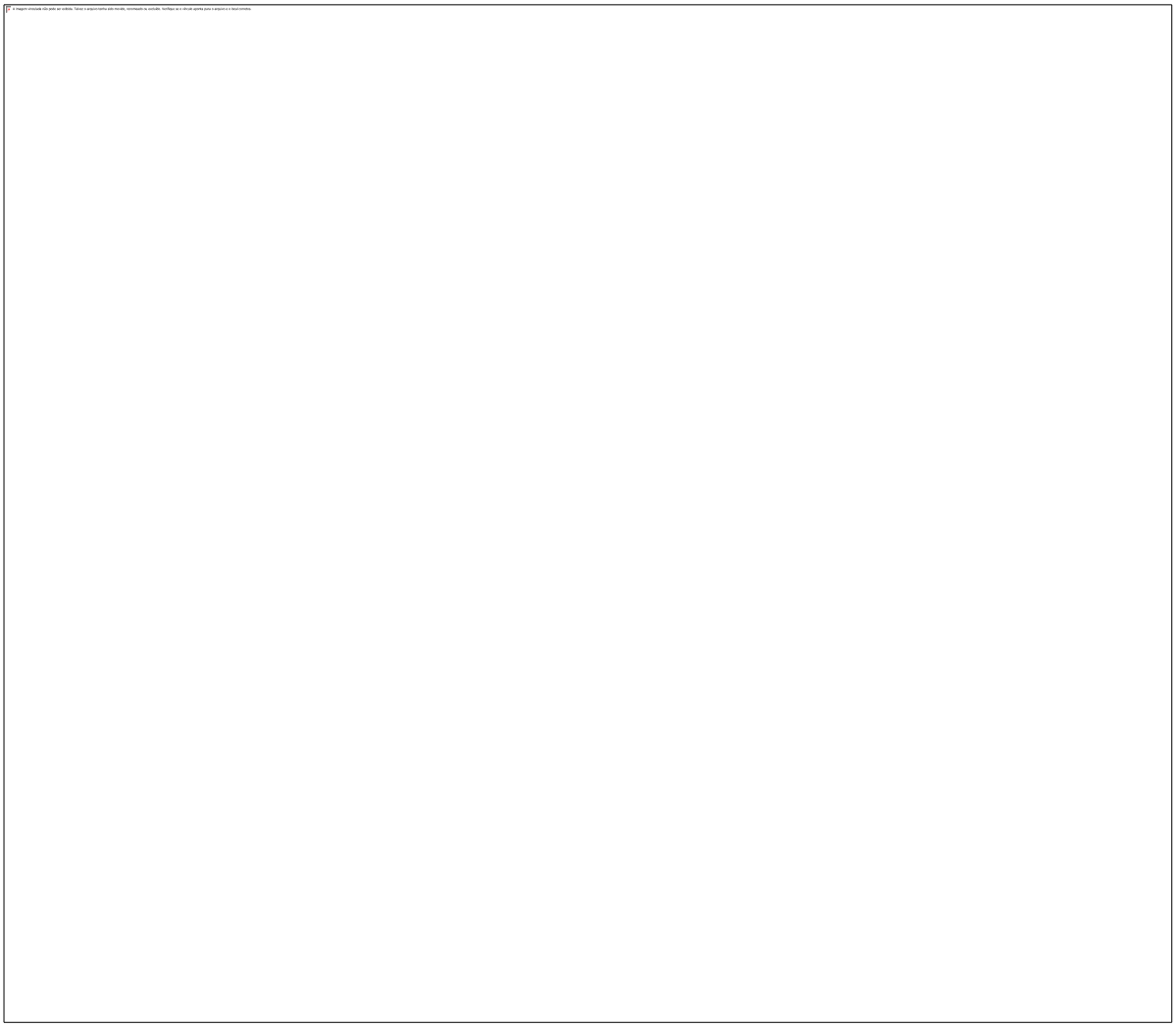


Tabela I

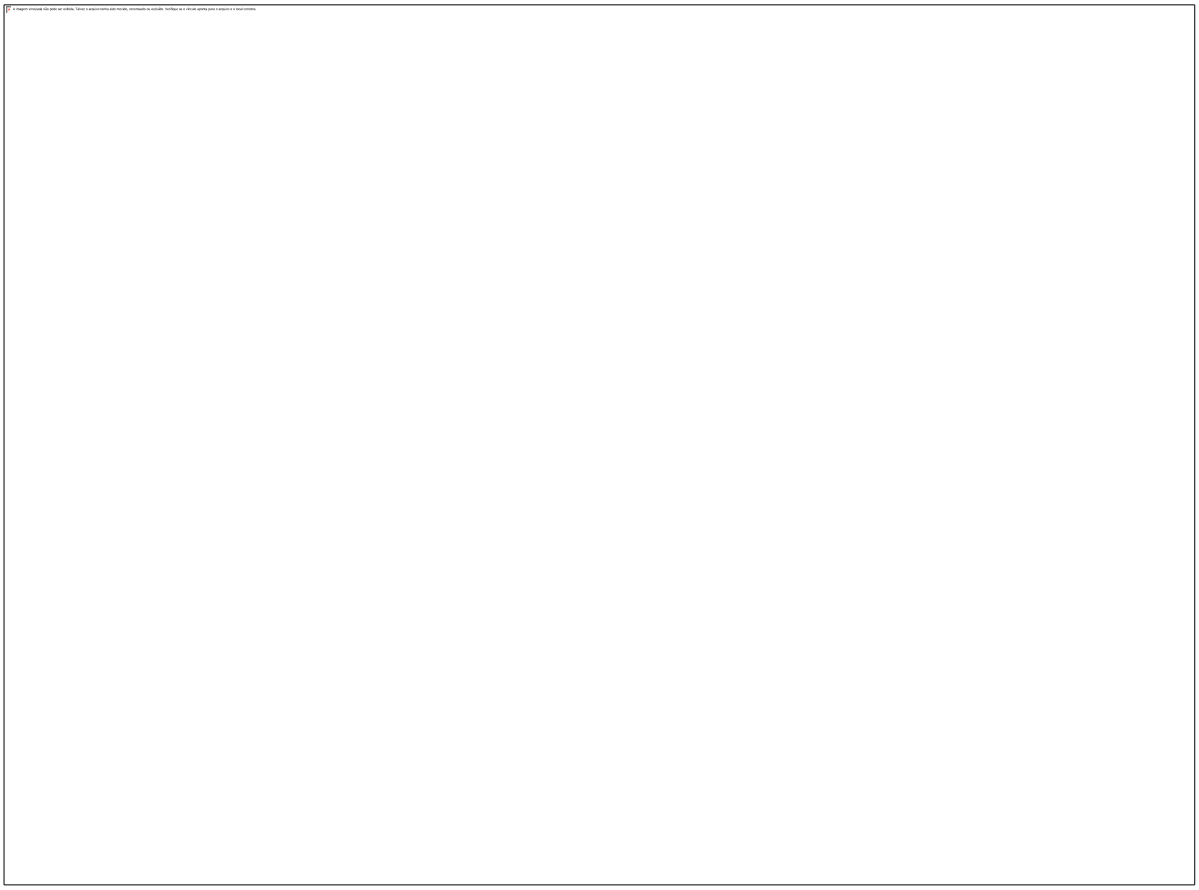


Tabela II

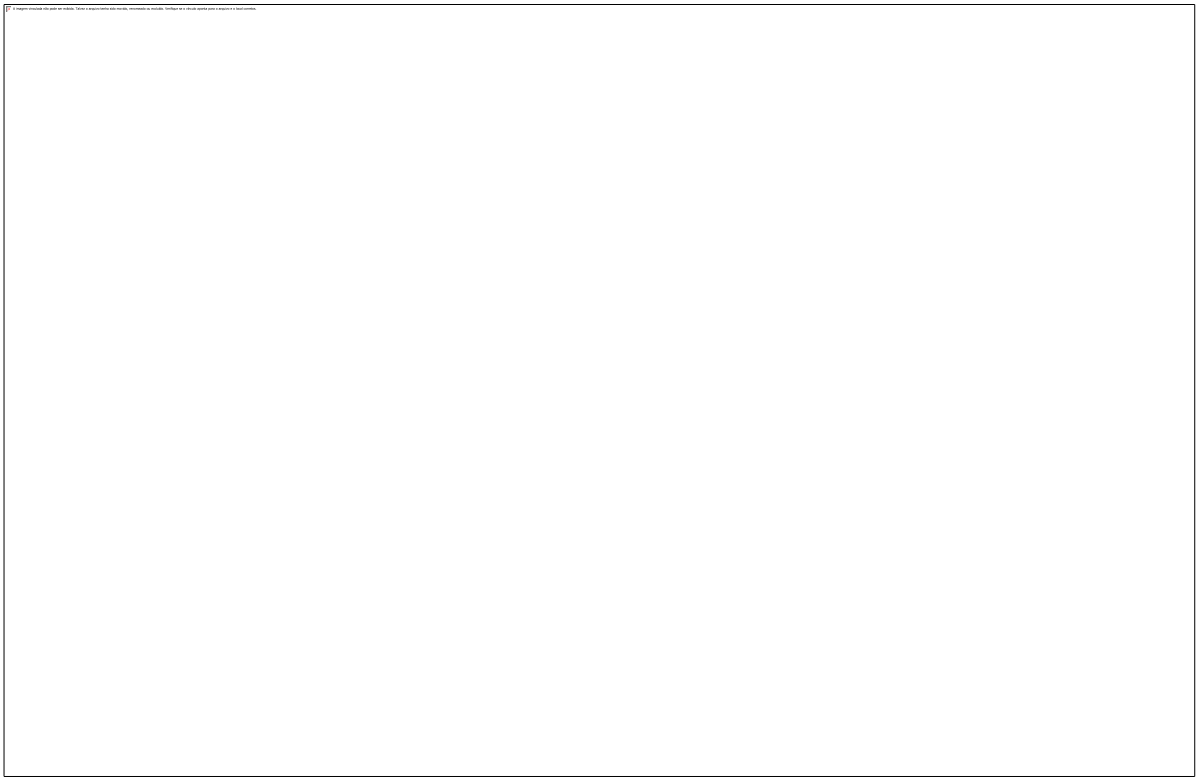


Tabela III

CÓDIGO: ET1669

AUTOR: ISABELE DE OLIVEIRA FERREIRA

ORIENTADOR: ISABEL DILLMANN NUNES

TÍTULO: Desenvolvimento de plataforma de ensino de pensamento computacional apoiada por inteligência artificial voltada para pessoa idosa

Resumo

Diante dos avanços das tecnologias digitais, é necessário dominar ferramentas tecnológicas para realizar tarefas do cotidiano. Entretanto, muitos grupos ainda são excluídos da cidadania digital, entre eles se destaca o público 60+ [Bernardo et al., 2024]. A população com mais de 60 anos vem aumentando a cada ano [IBGE. 2022], evidenciando a necessidade de iniciativas voltadas para esse público. Esse trabalho visa apresentar o desenvolvimento de uma plataforma web de ensino de programação apoiada por inteligência artificial voltada para o público 60+, contextualizado nas demandas do Projeto de Extensão de Inclusão Digital para Pessoas Idosas (ProEIDI/UFRN). A ferramenta busca atenuar barreiras comuns em ambientes tradicionais por meio de uma interface adaptada, além de utilizar a linguagem de programação Delégua, estruturada integralmente em português. A arquitetura de software é composta por front-end em Next.js hospedado na Vercel e back-end em Node.js com banco de dados MySQL, integrado ao modelo de linguagem Llama 3.3 por meio da infraestrutura Groq. A plataforma disponibiliza conteúdos teóricos contextualizados, lições práticas em formato de quiz, editor de código e compilador online, integrados a um tutor de inteligência artificial que fornece feedbacks formativos e imediatos.

Palavras-chave: Pensamento Computacional. Inteligência Artificial. Pessoas Idosas.

TITLE: Development of an AI-supported computational thinking learning platform aimed at older adults

Abstract

In light of advances in digital technologies, mastering technological tools has become necessary to perform everyday tasks. However, many groups are still excluded from digital citizenship, with older adults (60+) standing out among them [Bernardo et al., 2024]. The population aged 60 and over has been increasing every year [IBGE, 2022], highlighting the need for initiatives aimed at this demographic. This work aims to present the development of a web platform for teaching programming supported by artificial intelligence aimed at the 60+ audience, contextualized within the demands of the Digital Inclusion Extension Project for Older Adults (ProEIDI/UFRN). The tool seeks to mitigate common barriers in traditional environments through an adapted interface, in addition to using the Delégua programming language, entirely structured in Portuguese. The software architecture is composed of a Next.js front-end hosted on Vercel and a Node.js back-end with a MySQL database, integrated with the Llama 3.3 language model through the Groq acceleration infrastructure. The platform provides contextualized theoretical content, practical lessons in quiz format, a code editor, and an online compiler, integrated with an artificial intelligence tutor that delivers formative and immediate feedback.

Keywords: Computational Thinking. Artificial Intelligence. Elderly People.

Introdução

Diante dos avanços das tecnologias digitais, é necessário dominar ferramentas tecnológicas para realizar tarefas do cotidiano. Estar incluído digitalmente virou sinônimo de conseguir se comunicar, estar informado sobre o que acontece no mundo e ter autonomia para resolver problemas diários. Entretanto, muitos grupos ainda são excluídos da cidadania digital, entre eles se destaca o público 60+ [Bernardo et al., 2024]. A população com mais de 60 anos vem aumentando a cada ano [IBGE, 2022], evidenciando a necessidade de iniciativas voltadas para esse público.

Visando esse cenário, o Projeto de Extensão de Inclusão Digital para Pessoas Idosas (ProEIDI), desenvolvido na Universidade Federal do Rio Grande do Norte (UFRN), oferece cursos voltados para o público 60+ abordando os principais conceitos envolvendo o uso do smartphone e computador. Essa iniciativa tem como objetivo incentivar a autonomia, desenvolvimento cognitivo e inclusão digital das pessoas idosas participantes do projeto. Com a evolução do projeto, mais cursos foram sendo incorporados, dentre eles os de Pensamento Computacional e Inteligência Artificial [Medeiroset. al., 2024].

Essa mudança estrutural evidencia que a necessidade do público idoso ao relacionar-se com a tecnologia vai além do simples manuseio de ferramentas digitais, adentrando também no entendimento de paradigmas como o pensamento computacional. O Pensamento Computacional é um conceito desenvolvido como uma estratégia para resolução de problemas baseado em 4 pilares, sendo eles: abstração (capacidade de focar na parte importante do problema), decomposição (dividir o problema em partes menores), reconhecimento de padrões (identificar características repetidas entre os problemas) e algoritmos (sequência lógica de passos para resolver um problema). Essa estratégia é usada para além da computação, sendo possível ser aplicada na resolução de problemas do cotidiano como uma habilidade fundamental humana [Wing, 2006]. Com

isso, esse conteúdo ajuda no desenvolvimento de habilidades executivas, como memória, raciocínio lógico e resolução de problemas [Lucena, 2020]. Ao estimular essas funções, a pessoa idosa desenvolve maior criticidade e criatividade, ajudando a criar maior autonomia no uso de ferramentas computacionais.

O ensino de programação para pessoas idosas inclui desafios específicos, uma vez que ambientes de desenvolvimento convencionais apresentam vocabulário muito técnico e interfaces que dificultam a usabilidade. Além disso, no contexto das salas de aula do ProEIDI, o tempo é insuficiente para os alunos realizarem muitos exercícios práticos e sanarem todas as suas dúvidas em sala, o que reforça a necessidade de ferramentas complementares de apoio fora do horário de aula. É nesse cenário que a incorporação de recursos de Inteligência Artificial (IA) generativa e modelos de tutoria pedagógica adaptativa desponta como uma alternativa para o acompanhamento contínuo do ensino de programação para esse público.

Tendo em vista esse contexto, o presente trabalho tem como objetivo geral propor, projetar e iniciar o desenvolvimento de uma plataforma web de ensino de pensamento computacional voltada para o público idoso, apoiada por tutoria pedagógica baseada em inteligência artificial, no contexto das demandas das turmas de Pensamento Computacional do ProEIDI.

Metodologia

A presente pesquisa aplicada investiga a utilização da Inteligência Artificial e de interfaces acessíveis no ensino de programação voltada para o público 60+. Quanto a sua abordagem, pode ser caracterizada como mista (qualitativa e quantitativa), abrangendo tanto a análise conceitual dos processos de aprendizagem, quanto a futura coleta de métricas do uso do ambiente digital em desenvolvimento. Do ponto de vista dos objetivos, o trabalho é descritivo e propositivo, estruturado a partir da observação das turmas do ProEIDI.

Por meio da revisão de literatura e do histórico de observação dentro das salas de aulas, identificou-se a necessidade de desenvolver uma interface com baixa densidade de elementos, botões com áreas ampliadas de clique, ajuste no tamanho de fontes, alto contraste de cores e eliminação de termos técnicos em língua estrangeira. Essas diretrizes visam diminuir a carga de memória necessária para utilizar a plataforma, bem como reduzir a ansiedade associada ao medo de cometer erros operacionais.

Sobre a modelagem da arquitetura de software da plataforma, o front-end foi desenvolvido utilizando framework Next.js e hospedado na plataforma Vercel, já o back-end foi feito a partir de um ambiente de execução Node.js em comunicação com o banco de dados relacional MySQL. Para a execução das práticas de programação integrou-se o interpretador da linguagem Delégua, que utiliza comandos em português brasileiro. Essa adição permite que o público alvo foque integralmente no algoritmo e na lógica sequencial, diminuindo o atrito com vocabulários estrangeiros. Além disso, o back-end conecta-se à API da plataforma Groq para processar as requisições com o

modelo de linguagem Llama 3.3, viabilizando o retorno de feedbacks rápidos e personalizados a cada exercício submetido pelo aluno. O comportamento do modelo foi definido via Engenharia de Prompt, atribuindo-lhe a função de um monitor atencioso e empático, o qual deve avaliar o código submetido pelo estudante. A IA valida os acertos com mensagens encorajadoras e, em caso de erro, fornece dicas sem entregar a solução pronta, estimulando a autonomia na depuração dos exercícios.

Resultados e Discussões

Os resultados obtidos no âmbito deste trabalho mostram a concepção, o projeto de arquitetura e o desenvolvimento da plataforma web Senior Code. A ferramenta foi concebida para atender a dois perfis distintos de usuários: o perfil de "Professor" e o perfil de "Aluno". Ao autenticar-se, o professor dispõe de interfaces para o gerenciamento pedagógico, podendo cadastrar e atualizar os conteúdos teóricos, formular novos testes de múltipla escolha (quizzes) e parametrizar exercícios práticos de programação com as diretrizes a serem validadas pela feedback gerado por inteligência artificial.

Por sua vez, o perfil de aluno dá acesso direto a uma área visual limpa e descomplicada, organizada em três módulos centrais: Conteúdo Teórico, Lições Práticas e Espaço de Treinamento. Essa segmentação permite que a pessoa idosa navegue de forma linear e autônoma, minimizando o risco de desorientação espacial durante o uso da plataforma, confundindo onde fica cada função da ferramenta.

Na seção dos conteúdos teóricos, os conceitos estruturantes da computação foram desmembrados em lições progressivas: introdução ao pensamento computacional, decomposição, abstração, reconhecimento de padrões e algoritmos, além de tópicos elementares de programação como variáveis, tipos de dados, operadores, estruturas de repetição e condicionais. Ao desenvolver uma plataforma voltada para o ensino de programação para a pessoa idosa, é necessário abordar os conteúdos de forma contextualizada para as vivências desse público, buscando trazer os conceitos de pensamento computacional para o seu cotidiano [Nunes, 2025]. Com isso, os conteúdos cadastrados buscam utilizar sempre uma linguagem simples, clara e sem termos infantilizados.

Na parte de lições práticas, os exercícios dividem-se em dois formatos: quizzes e questões práticas de programação. Os exercícios cadastrados seguem o padrão de serem desenvolvidos de forma análoga às vivências da pessoa idosa, transformando seu dia a dia em questões de programação para serem resolvidas utilizando os pilares do pensamento computacional. Dessa maneira, as pessoas idosas têm a oportunidade de treinar a resolução de problemas reais relacionados a sua vida utilizando os conhecimentos adquiridos com plataforma, assim trabalhando suas habilidades cognitivas, como memória e raciocínio lógico. Um exemplo implementado aborda o reconhecimento de padrões na administração de medicamentos: o aluno analisa uma sequência temporal de dosagens (6h, 12h, 18h) para identificar o intervalo regular (+6 horas) e deduzir o próximo horário (24h/Meia-noite).

Na última sessão da plataforma, o espaço de treinamento contém um compilador online em que o usuário pode digitar e compilar códigos em Delégua em tempo real, sem a obrigatoriedade de cumprir enunciados fixos, incentivando a prática contínua e a curiosidade investigativa.

Os achados deste trabalho dialogam diretamente com a literatura seminal de Wing (2006), reafirmando o pensamento computacional como uma competência humana transversal de resolução de problemas, e não como uma habilidade restrita à formação de profissionais de computação. No âmbito da estimulação cognitiva para o público idoso, as escolhas arquiteturais e didáticas convergem com as pesquisas de Lucena (2020) e Oliveira e Pasqualotti (2021), cujos resultados evidenciam que o exercício de abstração, lógica e decomposição auxiliam na preservação da memória de trabalho e da atenção. Ademais, ao confrontar a Senior Code com plataformas educacionais genéricas observa-se que ferramentas convencionais impõem barreiras ao público 60+ devido a interfaces sobrecarregadas e foco predominante em linguagens de sintaxe em inglês. A combinação de uma interface acessível, comandos em português e tutoria por IA generativa supre o desafio da limitação de tempo para atendimento individualizado nas aulas do ProEIDI (Medeiros et al., 2024), proporcionando uma alternativa viável para a continuidade da aprendizagem fora do ambiente presencial.

Conclusão

O presente trabalho apresentou a concepção, o projeto de arquitetura e o desenvolvimento da plataforma web Senior Code, voltada para o ensino de pensamento computacional e lógica de programação para pessoas idosas, apoiada por tutoria adaptativa baseada em inteligência artificial. A pesquisa demonstrou a viabilidade técnica de articular tecnologias web modernas (Next.js, Node.js e MySQL), uma linguagem com sintaxe em língua portuguesa (Delégua) e serviços de IA (Groq/Llama 3.3 70B) para construir um ecossistema educacional acessível, seguro e inclusivo. O ambiente permite o contato da pessoa idosa com a programação, superando o letramento digital meramente instrumental e fomentando a autonomia, o raciocínio crítico e o estímulo às funções executivas. Como trabalhos futuros e desdobramentos da pesquisa, propõe-se a aplicação sistemática de testes empíricos de usabilidade, coleta de percepções qualitativas e quantitativas junto aos alunos das próximas turmas de Pensamento Computacional do ProEIDI e a condução de avaliações estruturadas de ganho cognitivo (protocolos de pré-teste e pós-teste) para mensurar a evolução no domínio dos pilares do pensamento computacional e na retenção de memória. Além disso, é necessário o aprimoramento do módulo de IA e a finalização do protótipo desenvolvido durante essa pesquisa.

Referências

Bernardo, J., Moura, B., Mattos, R., Fialho, I., Brito, K., Pinheiro, M., Pessoa, M. (2024) “Defensor Digital 60+: uma estratégia de inclusão digital da pessoa idosa”, In: Anais do XXX Workshop de Informática na Escola (WIE 2024), Sociedade Brasileira de Computação, Porto Alegre, p. 635–645.

Azevedo, A. L. M. DOS S. IBGE - Educa | Jovens. Disponível em: <https://educa.ibge.gov.br/jovens/conheca-o-brasil/populacao/18318-piramide-etaria.html>. Acesso em 07 dez. 2025.

Nunes, I. C., França, R. S. de. (2025) “Pensamento Computacional Para Todas As Idades: Uma Análise Comparativa Do Aprendizado De Programação Entre Crianças E Pessoas Idosas”. In: WORKSHOP SOBRE EDUCAÇÃO EM COMPUTAÇÃO (WEI), 33. , 2025, Maceió/AL. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, 2025 . p. 971-981. ISSN 2595-6175. DOI: https://doi.org/10.5753/wei.2025.8652.

Oliveira, C. R. de, Ferreira, I. de O., Oliveira, W. M. S. de, Macena, F. J. dos S., Medeiros, V. A., Martins, P. L. de S., Nunes, I. D. (2025). “Senior Code AI: Estimulação Cognitiva para Idosos por Meio da Programação”. In: CONCURSO APPS.EDU - CATEGORIA PROTÓTIPO - CONGRESSO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO (CBIE), 14. , 2025, Curitiba/PR. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, 2025. p. 220-223. DOI: https://doi.org/10.5753/cbie_estendido.2025.13897.

Wing, J. M. 2006. Computational thinking. Commun. ACM 49, 3 (March 2006), 33–35. https://doi.org/10.1145/1118178.1118215

Medeiros, B. M., Lima, F. C. N. de., Cardoso, R. M. M., Campos, S. C. G. de, Fernandes, V. E. G., Nunes, I. D. (2024) “Da Teoria à Prática: Resultados de uma Metodologia de Ensino de Pensamento Computacional para Pessoas Idosas”. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO (SBIE), 35. , 2024, Rio de Janeiro/RJ. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, 2024 . p. 2506-2518. DOI: https://doi.org/10.5753/sbie.2024.242639.

LUCENA, Daniel Araújo de. Pensamento computacional como intervenção para desenvolvimento cognitivo
em idosos. 2020. Dissertação de Mestrado. Universidade Federal do Rio Grande do Norte.

CÓDIGO: ET1730

AUTOR: MARIA EDUARDA DOS SANTOS CARVALHO

ORIENTADOR: ADRIAO DUARTE DORIA NETO

TÍTULO: Implantação de Algoritmos de Diagnóstico de Falhas em Dispositivos IoT de Baixo Custo: Estudo de Caso com o ESP32S3

Resumo

Este trabalho apresenta uma abordagem baseada em redes neurais para identificação e diagnóstico de falhas em um sistema experimental de tanques acoplados. Inicialmente, foi desenvolvido um modelo neural autorregressivo para representar o comportamento normal do processo a partir do histórico da tensão aplicada à bomba e dos níveis dos dois tanques, os dados foram organizados em janelas temporais de cinco amostras e utilizados para treinamento, validação e teste do modelo de identificação. Em seguida, foram analisadas oito condições de falha, envolvendo vazamentos, entupimentos parciais e totais e falha da bomba, além da condição normal de operação. Para o diagnóstico, foi treinado um classificador utilizando informações provenientes dos níveis medidos e estimados pelo modelo. O classificador alcançou acurácia de 88,74%, a análise da matriz de confusão mostrou elevado desempenho para a maior parte das classes, embora tenha sido observada dificuldade na distinção entre falha da bomba e vazamento do tanque 1 para o reservatório, devido à semelhança entre os efeitos produzidos por essas condições sobre a dinâmica do sistema. Os resultados indicam a viabilidade da abordagem para o diagnóstico das falhas investigadas.

Palavras-chave: Diagnóstico de falhas. Redes neurais. Tanques acoplados.

TITLE: Implementation of Fault Diagnosis Algorithms on Low-Cost IoT Devices: A Case Study with the ESP32S3

Abstract

This paper presents a neural network-based approach for fault identification and diagnosis in an experimental coupled-tank system. Initially, an autoregressive neural model was developed to represent the process's normal behavior based on the history of the voltage applied to the pump and the levels of the two tanks; the data were organized into five-sample time windows and used for training, validation, and testing the identification model. Subsequently, eight fault conditions were analyzed—including leaks, partial and total blockages, and pump failure—alongside the normal operating condition. For diagnosis, a classifier was trained using information derived from both measured and model-estimated levels. The classifier achieved an accuracy of 88.74%; confusion matrix

analysis revealed high performance for most classes, although difficulty was observed in distinguishing between pump failure and a leak from tank 1 into the reservoir, due to the similarity of the effects these conditions have on system dynamics. The results demonstrate the feasibility of the approach for diagnosing the investigated faults.

Keywords: Fault diagnosis. Neural networks. Coupled tanks.

Introdução

A contínua integração de sensores, sistemas de comunicação e dispositivos computacionais tem ampliado as possibilidades de monitoramento e automação de processos industriais comuns (HAQIQ et al., 2026). Ao mesmo tempo, o avanço da inteligência artificial tem permitido a aplicação de modelos de aprendizado de máquina em tarefas como identificação de sistemas, monitoramento de processos e detecção e diagnóstico de falhas (FERNANDES; CORCHADO; MARREIROS, 2022).

Normalmente, parte significativa do processamento necessário para essas aplicações é realizada de maneira integrada, utilizando computadores e controladores. No entanto, a evolução dos microcontroladores tem possibilitado transferir tarefas para dispositivos próximos aos sensores e atuadores. Essa abordagem é associada à computação de borda e pode reduzir a dependência de infraestrutura externa e permitir que informações importantes sejam processadas de maneira local (HAQIQ et al., 2026). Além da computação de borda, observa-se atualmente um crescimento expressivo no uso de inteligência artificial (IA) em diversas aplicações, desde sistemas de monitoramento até o diagnóstico de falhas. Nesse contexto, a IA também tem sido cada vez mais incorporada a dispositivos de borda, beneficiando-se das características apresentadas anteriormente, como a redução da latência e a menor dependência de recursos computacionais remotos. Para viabilizar esse tipo de aplicação, dispositivos como Single Board Computers (SBCs) e microcontroladores vêm sendo explorados para a execução de modelos de IA diretamente na borda. Um exemplo é o ESP32-S3, um microcontrolador de baixo custo que apresenta recursos computacionais suficientes para executar determinados modelos de IA, tornando-o uma alternativa para aplicações de inteligência artificial em dispositivos embarcados.

Entre as aplicações possíveis está a detecção e o diagnóstico de falhas em processos industriais – alterações provocadas por vazamentos, obstruções, falhas em atuadores – que podem modificar o comportamento dinâmico de um processo e comprometer a operação. A identificação prévia dessas anomalias permite que ações de manutenção sejam tomadas a fim de reduzir possíveis impactos sobre o sistema. Uma solução para esse problema consiste em utilizar modelos capazes de aprender o comportamento esperado do processo em condições normais e analisar os desvios observados quando o sistema passa a operar sob condições anormais (FERNANDES; CORCHADO; MARREIROS, 2022).

Neste trabalho, utiliza-se como estudo de caso um sistema didático de tanques acoplados disponível no Laboratório de Avaliação de Medição em Petróleo da Universidade Federal do Rio Grande do Norte. A partir de dados experimentais de tensão aplicada à bomba e níveis dos dois tanques, foram utilizadas redes neurais para modelar o comportamento

dinâmico do processo, em seguida, foram analisados dados obtidos mediante a introdução de diferentes condições de falha, incluindo vazamentos, entupimentos parciais e totais e falha da bomba.

O objetivo deste trabalho é avaliar o desempenho de uma abordagem baseada em redes neurais para a identificação e o diagnóstico de falhas em um sistema de tanques acoplados, como etapa preliminar à sua posterior implantação em um microcontrolador ESP32-S3. Para isso, foram realizadas etapas de captura e pré-processamento dos dados, identificação do comportamento, treinamento dos modelos e análise do comportamento do sistema diante das falhas introduzidas experimentalmente.

Metodologia

O trabalho foi desenvolvido utilizando um sistema de tanques acoplados disponível no Laboratório de Avaliação de Medição em Petróleo da Universidade Federal do Rio Grande do Norte(Figura. 1). O processo é composto por dois tanques interligados, cujos níveis são monitorados durante a operação, a alimentação do sistema é realizada por uma bomba, sendo consideradas como variáveis para a identificação do processo a tensão associada ao seu acionamento e os níveis dos tanques 1 e 2.

A medição dos níveis dos tanques foi realizada a partir da pressão de fundo realizada pela coluna de água, utilizando instrumentos industriais instalados no sistema, as informações obtidas desses instrumentos são transmitidas, por meio do protocolo MODBUS, a um ESP32 responsável pelo controle do processo e também pela geração das tensões aplicadas à bomba. A comunicação entre o ESP32 e a aplicação desenvolvida no ambiente Node-RED é realizada por meio do protocolo MQTT, e essa aplicação é responsável por gerar as variações de tensão aplicadas na bomba e capturar os níveis correspondentes dos tanques, que serão usados posteriormente para o treinamento dos modelos.

De início, foram adquiridos dados experimentais do sistema em condições normais de operação, esses dados foram utilizados para o desenvolvimento de um modelo neural de identificação, responsável por estimar o comportamento dos níveis dos tanques. Os registros passaram por etapas de pré-processamento e organização em conjuntos destinados ao treinamento, validação e teste dos modelos.

Foram utilizadas janelas contendo cinco instantes consecutivos das três variáveis do processo, tensão da bomba, nível do tanque 1 e nível do tanque 2, o que totaliza 15 valores de entrada. A partir desses valores, o modelo realiza a estimativa dos níveis dos dois tanques no instante subsequente.

Um aspecto importante do pré-processamento foi o tratamento das descontinuidades existentes entre as capturas, uma vez que algumas delas foram interrompidas e retomadas posteriormente, fazendo com que o último estado registrado em uma captura não correspondesse ao primeiro estado da captura seguinte. Sem o tratamento dessas descontinuidades, a concatenação desses estados resultaria em transições anormais, o que

prejudicaria a aprendizagem da dinâmica correta do sistema. Desse modo, cada sequência foi processada de maneira independente para que, somente depois, os conjuntos fossem concatenados, preservando, assim, a continuidade dos dados.

Para representar o comportamento normal do processo foram construídas redes neurais com características autorregressiva do tipo NARX (Nonlinear Autoregressive with Exogenous Inputs). O modelo de identificação foi treinado com dados referentes à operação normal e seu desempenho foi analisado pela comparação entre os níveis medidos experimentalmente e os níveis estimados. Após a identificação do comportamento normal do sistema, foram consideradas oito classes de falha, além da condição de operação normal, totalizando nove classes, as falhas simuladas foram: entupimento parcial do tanque 1, entupimento parcial do tanque 2, entupimento total do tanque 1, entupimento total do tanque 2, vazamento do tanque 1 para o reservatório, vazamento do tanque 1 para o tanque 2, vazamento do tanque 2 para o reservatório e falha da bomba.

As falhas foram introduzidas de forma controlada no sistema, os vazamentos foram provocados com a ajuda de uma mangueira de silicone, utilizada para criar um caminho adicional de escoamento, os entupimentos por sua vez, foram realizados utilizando borracha de calafetagem para restringir parcial ou totalmente o escoamento da água nos orifícios correspondentes, enquanto a falha da bomba foi simulada computacionalmente no ambiente Node-RED.

Para o diagnóstico das condições de operação, foram utilizadas informações obtidas a partir da comparação entre o comportamento medido do sistema e aquele estimado pelo modelo de identificação. Os dados foram organizados em janelas de cinco amostras consecutivas, considerando os níveis medidos e estimados dos dois tanques, assim foram utilizadas 4 sequências de 5 amostras, nível medido e estimado do tanque 1 e do tanque 2, o que totalizou 20 valores de entrada para o modelo de classificação. Esse conjunto de 20 valores foi associado à condição de operação presente no intervalo correspondente, tendo como possíveis saídas, uma das nove condições comentadas anteriormente.

O sistema de identificação utiliza um modelo NARX com 15 entradas, correspondentes a três variáveis — tensão e níveis dos tanques 1 e 2 — em uma janela de cinco amostras ((t) a (t-4)), para previsão dos níveis no instante (t+1). A arquitetura é composta por dois submodelos densos independentes, cada um com duas camadas ocultas de 10 neurônios, ativação ReLU e uma saída linear de um neurônio. O primeiro utiliza tensão e nível do tanque 1, enquanto o segundo utiliza os níveis dos tanques 1 e 2, permitindo a modelagem individual dos tanques devido à maior suscetibilidade do tanque 2 a ruídos. As saídas são concatenadas para formar a previsão conjunta.

O diagnóstico utiliza 20 entradas, correspondentes aos níveis reais e previstos dos dois tanques, também considerando cinco amostras temporais. A arquitetura possui uma camada oculta de 10 neurônios com ReLU e uma camada de saída de nove neurônios com Softmax, correspondentes às oito condições de falha e à operação normal. O modelo de identificação foi treinado por 2000 épocas, utilizando MSE como função de perda e MAE como métrica, enquanto o diagnóstico foi treinado por 1500 épocas, utilizando Sparse Categorical Crossentropy e acurácia. Ao final, o classificador foi avaliado por meio da matriz de confusão(Figura 4).

Resultados e Discussões

Inicialmente, foi avaliado o desempenho do modelo empregado na identificação da dinâmica do sistema de tanques. O modelo foi treinado para estimar os níveis dos tanques 1 e 2 no instante subsequente a partir do histórico temporal da tensão aplicada à bomba e dos níveis dos dois tanques.

Para o modelo de identificação, foram obtidos valores de erro quadrático médio(MSE) de 1,0912, 0,7500 e 1,4393 para os conjuntos de treinamento, validação e teste, respectivamente, também foram obtidos valores de erro absoluto médio(MAE) de 0,4125, 0,4941 e 0,6793 para os conjuntos de treinamento, validação e teste, respectivamente. Essas métricas quantificam a diferença entre os níveis estimados pelo modelo e aqueles medidos fisicamente, onde valores mais próximos de zero indicam maior proximidade entre as estimativas e as medições. Os erros de predição para os níveis dos tanques 1 e 2 no conjunto de teste (Figuras. 5 e 6) permitem observar a diferença entre os valores medidos e aqueles estimados pelo modelo ao longo das amostras. Os gráficos de comparação entre os níveis medidos e estimados reforçam esse resultado(Figuras. 2 e 3), para ambos os tanques, as curvas de predição permanecem visivelmente próximas dos níveis registrados durante as capturas no sistema físico, acompanhando adequadamente as variações dos níveis ao longo do conjunto de teste. Esse comportamento indica que o modelo conseguiu representar de maneira adequada a relação entre a tensão aplicada à bomba e os níveis dos tanques.

Após a obtenção do modelo do comportamento do sistema em condições de normalidade, foram analisados os dados coletados quando o sistema operava em condições de falhas, com isso para o diagnóstico foram consideradas nove classes: operação normal, entupimento parcial do tanque 1, entupimento parcial do tanque 2, entupimento total do tanque 1, entupimento total do tanque 2, vazamento do tanque 1 para o reservatório, vazamento do tanque 1 para o tanque 2, vazamento do tanque 2 para o reservatório e falha da bomba.

A preparação dos dados para o classificador corresponde à divisão de aproximadamente 70% dos dados para treinamento, 15% para validação e 15% para teste, realizada para cada condição de operação antes de os dados correspondentes de todas as classes serem agrupados para compor os conjuntos usados no treinamento, validação e teste do classificador. O modelo apresentou acurácia de aproximadamente 84,66% no treinamento, 84,66% na validação e 88,74% no conjunto de teste.

No cenário geral, esses foram os dados obtidos. Já na análise individual das classes, o entupimento parcial do tanque 2 e os entupimentos totais dos tanque 1 e 2 apresentaram taxa de identificação de 100% no conjunto de teste, o vazamento do tanque 2 para o reservatório também apresentou um desempenho elevado, com aproximadamente 94% de identificação correta. Para o funcionamento normal, cerca de 92% das amostras foram corretamente reconhecidas, enquanto aproximadamente 8% foram classificadas como entupimento parcial do tanque 1. O entupimento parcial do tanque 1 apresentou aproximadamente 91% de acerto.

Dentre os resultados obtidos, o que mais se destacou na matriz de confusão foi o referente à falha da bomba, onde apenas 50% das amostras dessa classe foram identificadas corretamente, enquanto cerca de 49% foram classificadas como vazamento do tanque 1 para o reservatório. Da mesma forma, aproximadamente 10% das amostras do vazamento do tanque 1 para o reservatório foram classificadas como falha da bomba. Isso mostra que existe uma limitação principal do classificador, que não ocorreu de maneira aleatória entre as classes, mas concentrou-se particularmente na separação dessas duas condições específicas(Figura. 4).

A confusão entre falha da bomba e vazamento do tanque 1 para o reservatório apresenta coerência com a dinâmica física do sistema, pois em condições normais, a bomba fornece água ao tanque 1 e quando ocorre um vazamento desse tanque para o reservatório, parte da vazão que deveria contribuir para sua elevação é desviada, reduzindo a quantidade de água disponível no tanque, uma redução na atuação da bomba pode produzir efeito semelhante, uma vez que também diminui a quantidade de água fornecida ao tanque 1. Dessa forma, ambas as condições podem gerar evoluções similares dos níveis, confundindo o classificador. A dificuldade do classificador em separar essas duas classes indica uma limitação das informações disponibilizadas para o diagnóstico, e não necessariamente que o modelo não consegue reconhecer as falhas. Diferentemente das classes do entupimento parcial do tanque 2 e os entupimentos totais dos dois tanques que alcançaram taxa de identificação de 100% no conjunto de teste, o que evidencia que as alterações geradas por cada falha conseguiram produzir padrões suficientemente distintos para serem reconhecidos pelo modelo.

De maneira geral, os resultados indicam que a identificação do comportamento do sistema e a classificação das condições de operação desse sistema foi capaz de distinguir a maior parte das falhas investigadas. A acurácia de 88,74% no conjunto de teste mostra a aplicabilidade dessa abordagem para o sistema estudado.

Conclusão

Este trabalho investigou a utilização de redes neurais para a identificação e o diagnóstico de falhas em um sistema de tanques acoplados. O modelo de identificação foi capaz de representar de forma satisfatória o comportamento dinâmico do sistema, apresentando baixos valores de erro e mantendo as estimativas dos níveis dos tanques próximas às medições obtidas experimentalmente. A partir dessa representação, foi possível realizar a classificação das diferentes condições de operação consideradas no estudo.

O modelo de diagnóstico apresentou acurácia de 88,74% no conjunto de teste e foi capaz de distinguir adequadamente a maior parte das condições analisadas. Algumas classes, como os entupimentos, apresentaram elevado desempenho de identificação. A principal limitação observada ocorreu na diferenciação entre a falha da bomba e o vazamento do tanque 1 para o reservatório, condições que produzem efeitos semelhantes sobre a dinâmica dos níveis e, consequentemente, podem gerar padrões de comportamento próximos para o classificador.

Os resultados obtidos demonstram a viabilidade da abordagem neural para o diagnóstico das falhas investigadas, constituindo uma etapa preliminar para a avaliação de sua implantação em um microcontrolador ESP32-S3. A implementação e validação dos modelos diretamente no dispositivo não foram concluídas, permanecendo como etapa necessária para avaliar efetivamente a viabilidade da solução embarcada.

Referências

Haqiq N, Zaim M, Haqiq A, Sbihi M, El Ouaazizi A. Internet das Coisas para a Indústria 4.0: Uma Revisão Sistemática da Literatura sobre Tecnologias, Arquiteturas, Aplicações e Desafios. IoT . 2026; 7(2):46. https://doi.org/10.3390/iot7020046

Fernandes M, Corchado JM, Marreiros G. Machine learning techniques applied to mechanical fault diagnosis and fault prognosis in the context of real industrial manufacturing use-cases: a systematic literature review. Appl Intell (Dordr). 2022;52(12):14246-14280. doi:10.1007/s10489-022-03344-3

Anexos

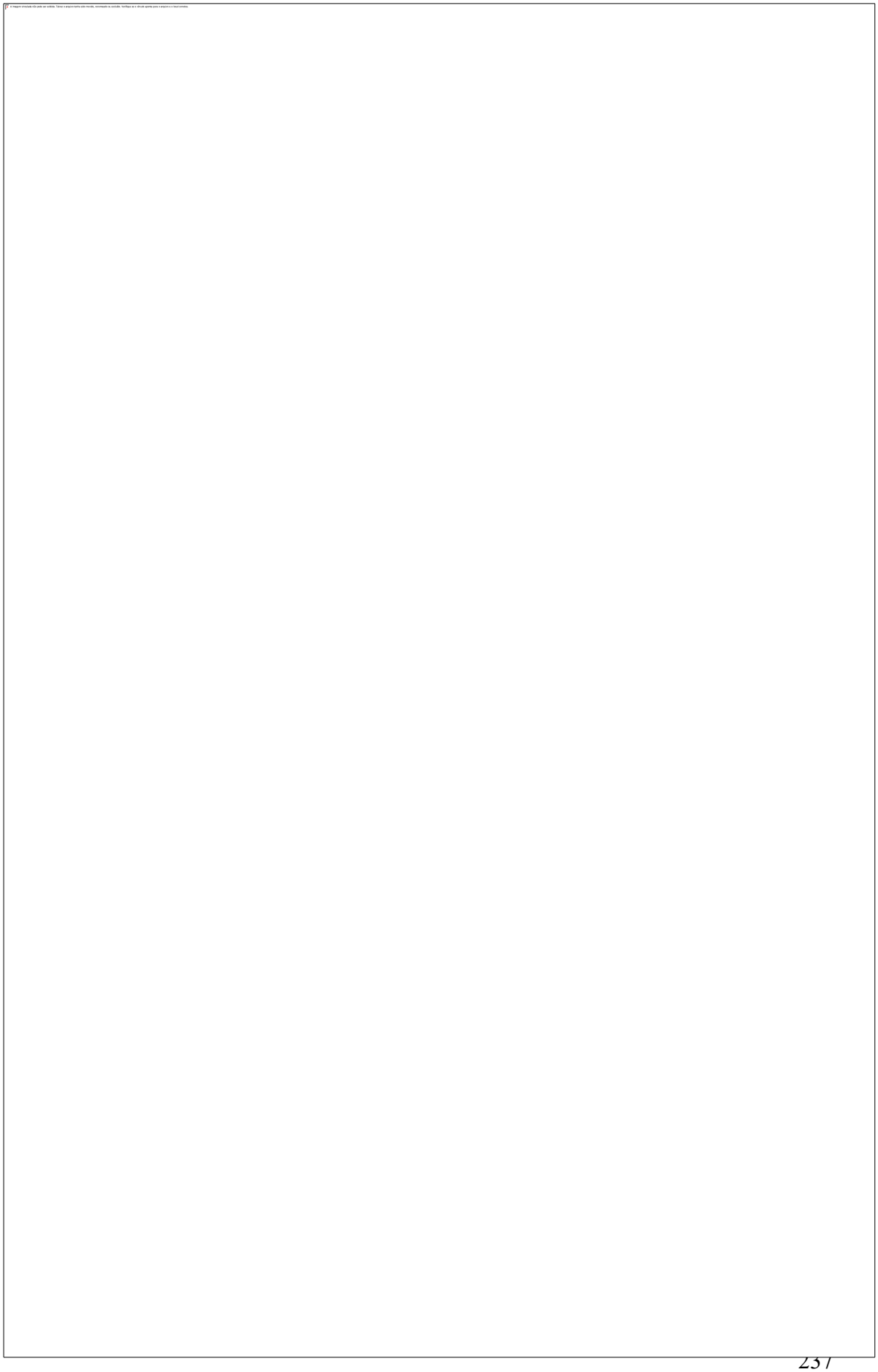


Figura 1 – Sistema de tanques acoplados utilizado para aquisição dos dados e realização dos ensaios de falha. Fonte: Autoria própria (2026).

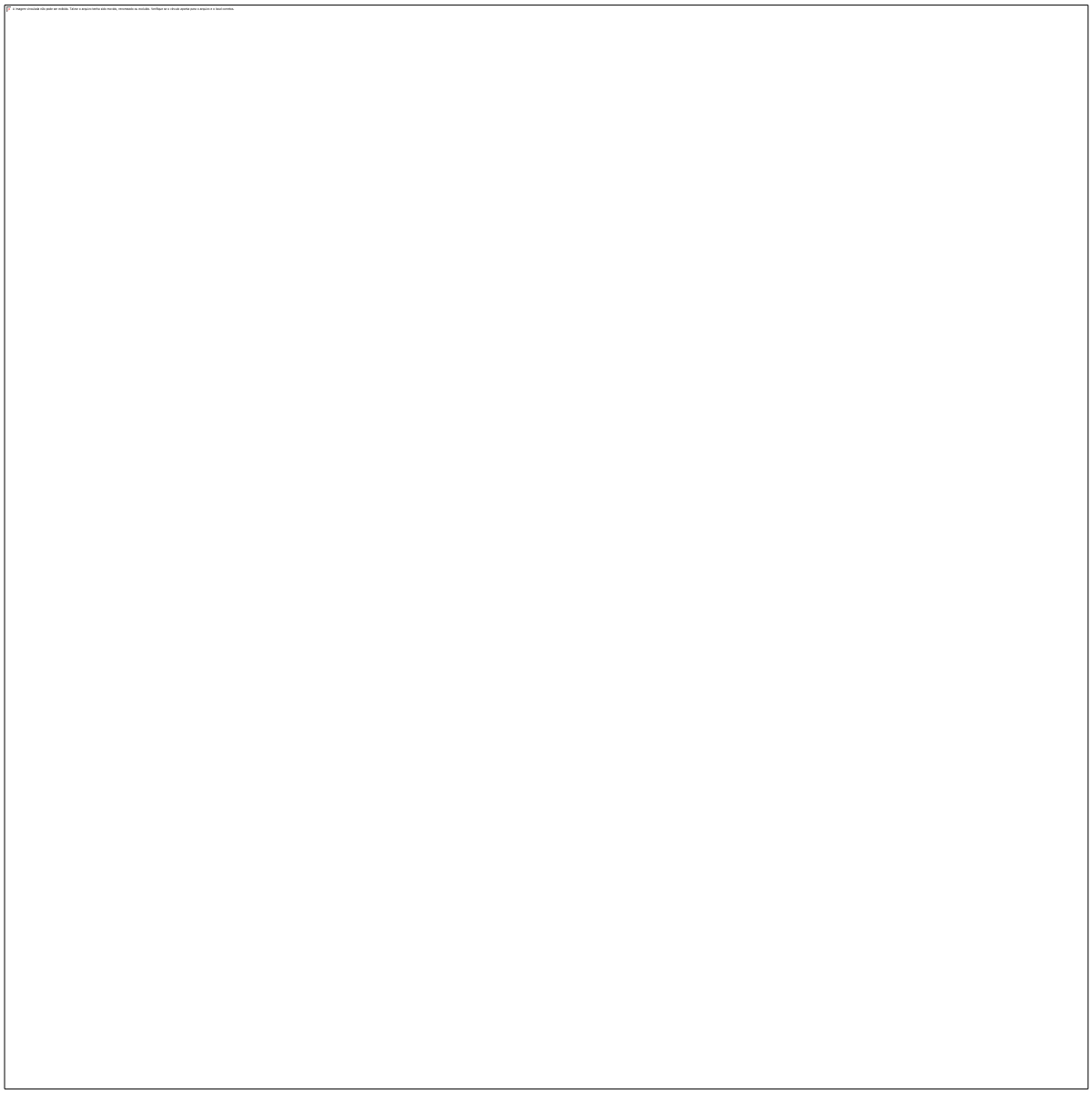


Figura 2 – Comparação entre o nível medido e o nível estimado pelo modelo para o tanque 1 no conjunto de teste.

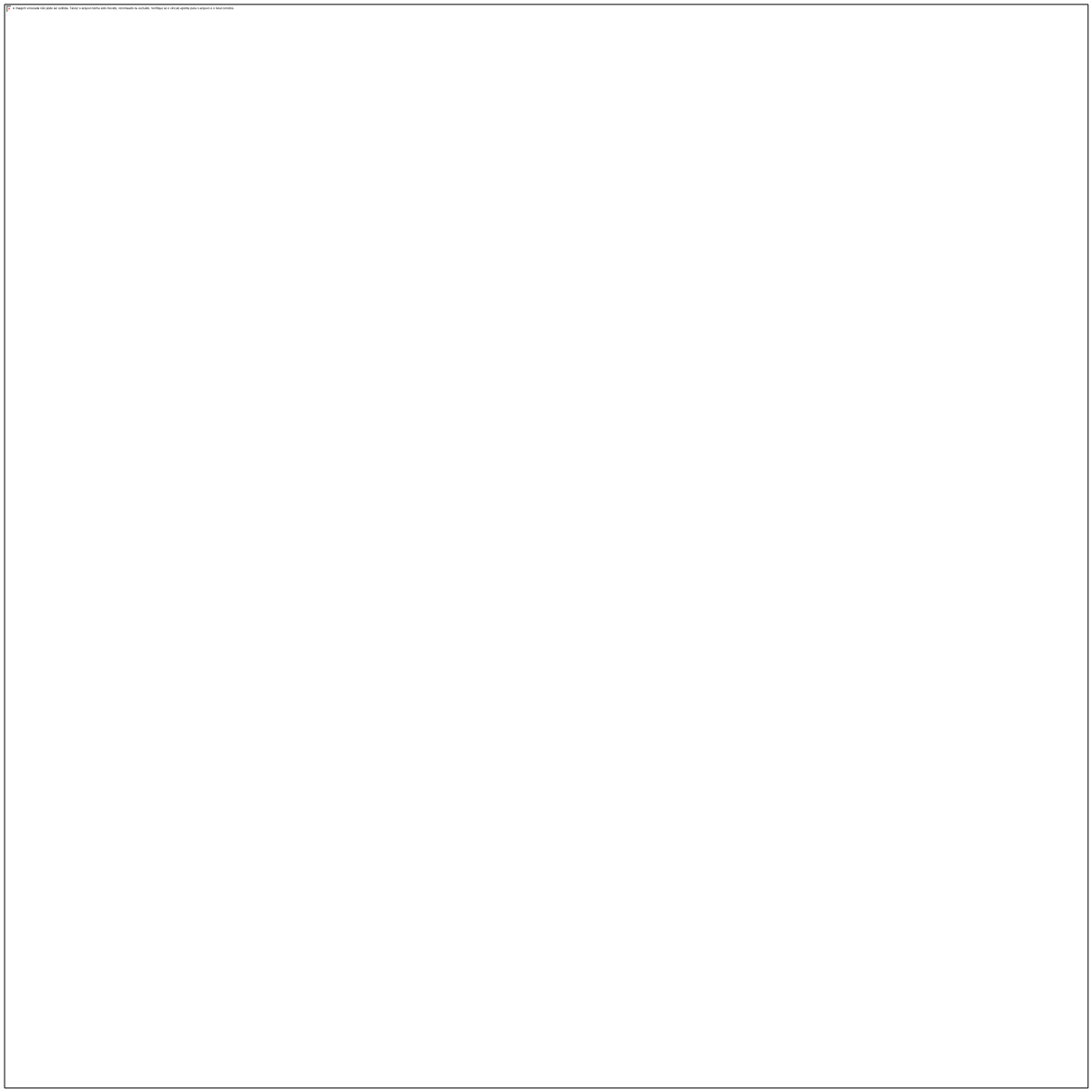


Figura 3 – Comparação entre o nível medido e o nível estimado pelo modelo para o tanque 2 no conjunto de teste.

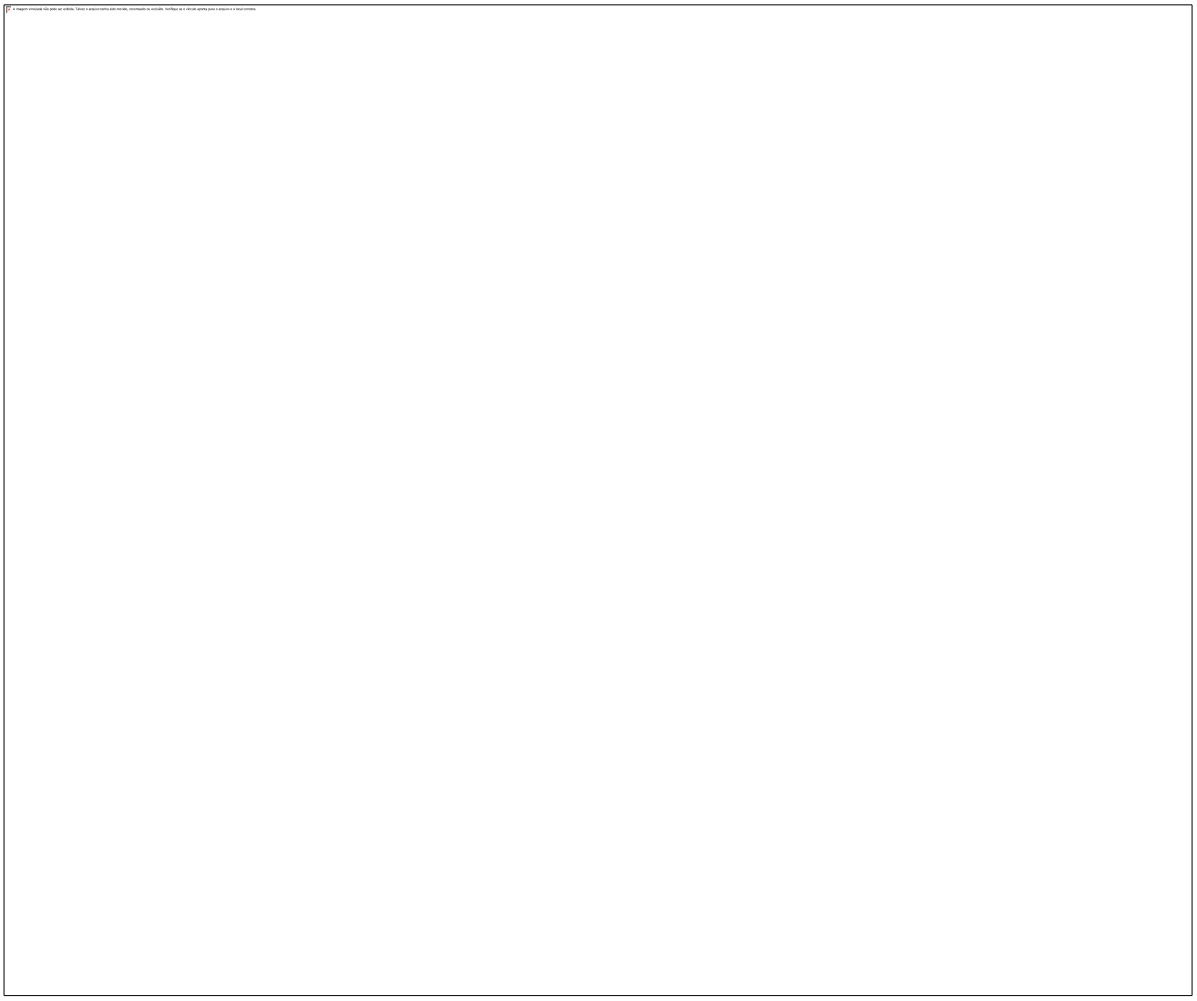


Figura 4 – Matriz de confusão normalizada do classificador para o conjunto de teste.

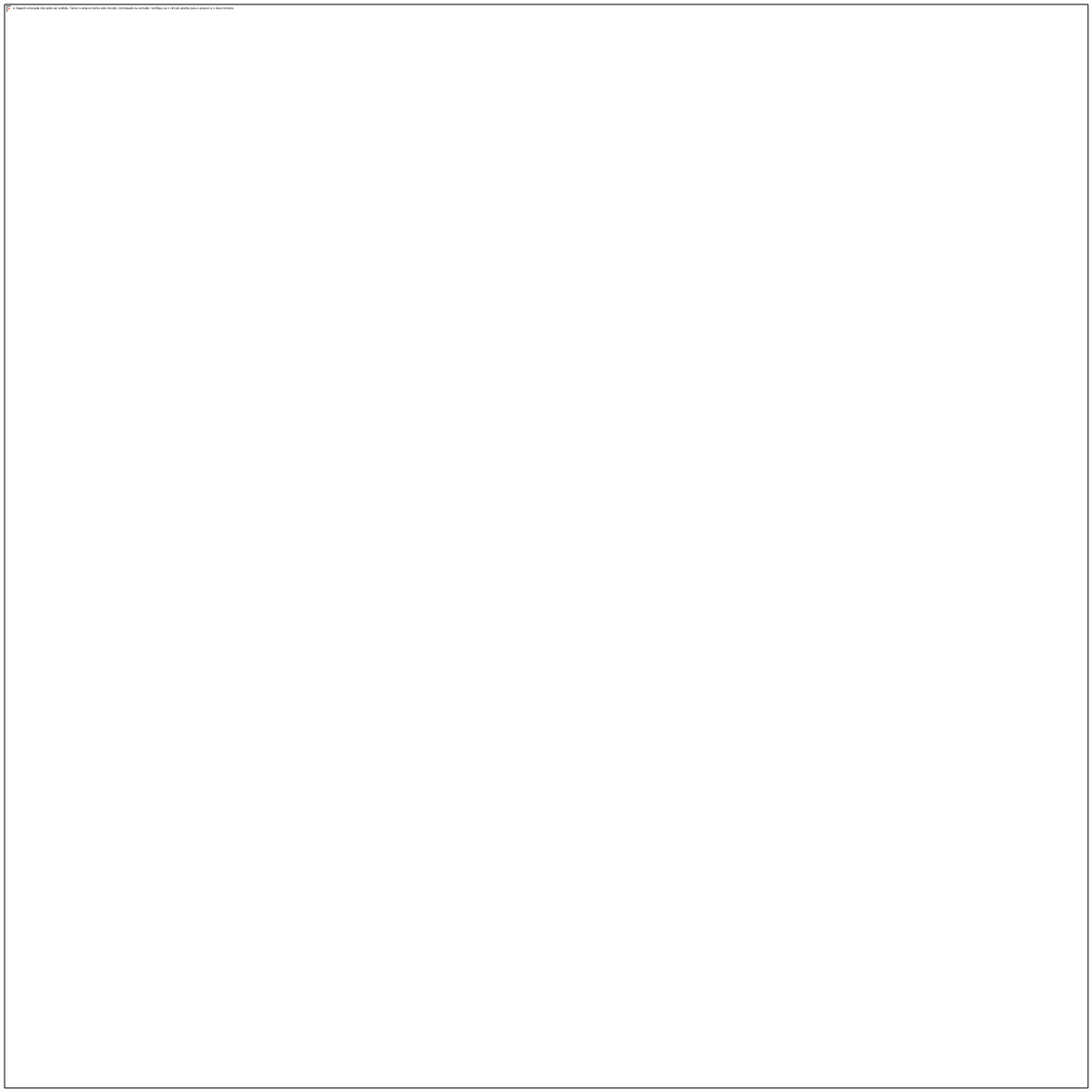


Figura 5 – Erro de predição do nível do tanque 1 no conjunto de teste.

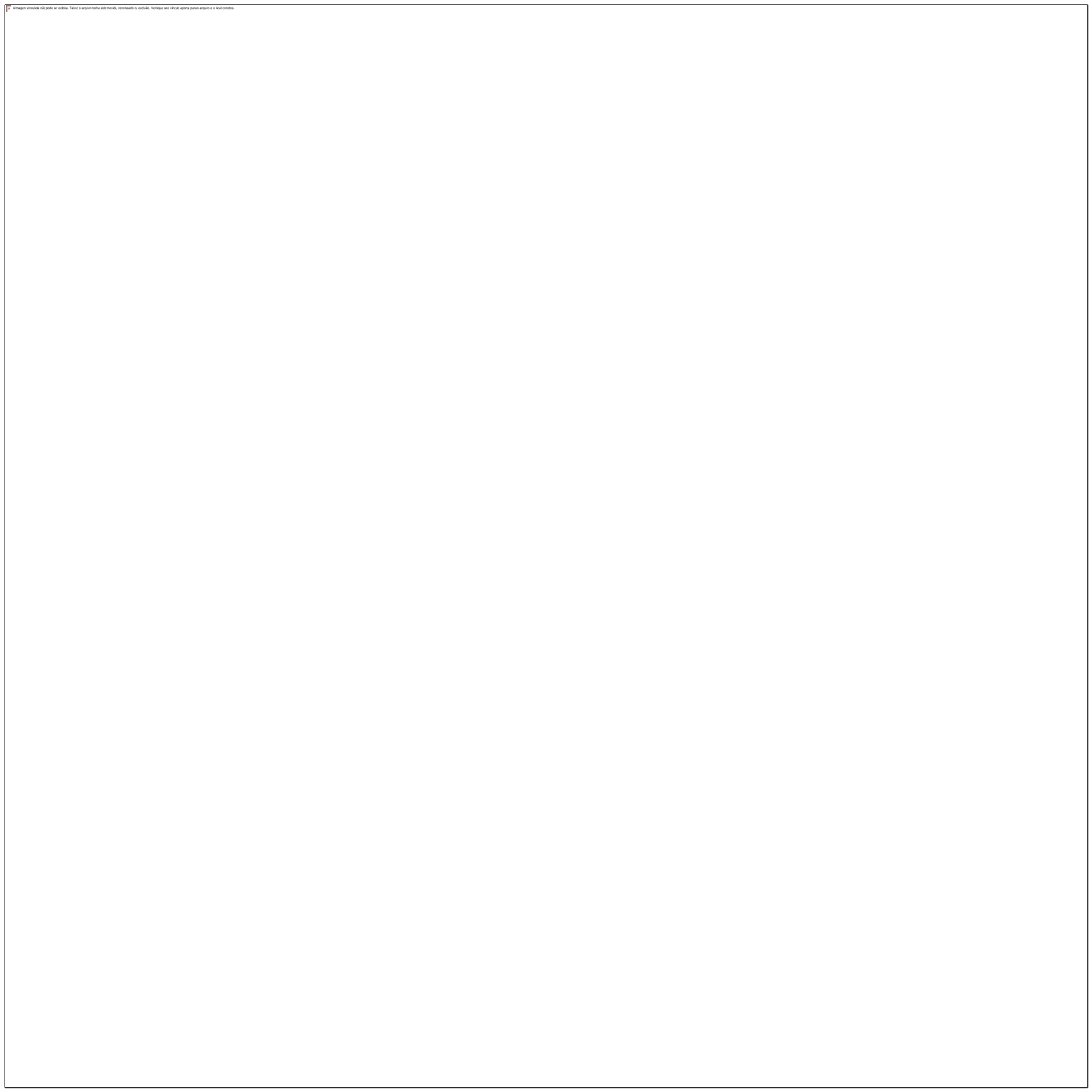


Figura 6 – Erro de predição do nível do tanque 2 no conjunto de teste.

CÓDIGO: ET1752

AUTOR: IRUZIKY ARAUJO DE MEDEIROS

ORIENTADOR: TIBERIO AZEVEDO PEREIRA

TÍTULO: Modelos de Redes Neurais Artificiais para aplicação multidisciplinar

Resumo

Modelos de linguagem são programas capazes de compreender perguntas e produzir textos, como ocorre em chatbots. Apesar dessa capacidade, podem usar informações desatualizadas ou responder sem apoio em fontes. A técnica RAG procura reduzir esses problemas. Antes de responder, o chatbot pesquisa documentos e recebe os trechos encontrados como material de consulta. Este trabalho não apresenta uma técnica nova. Ele reúne métodos publicados e compara as situações em que cada um funciona melhor. Foi escolhido o Open RAG Benchmark por quatro motivos: reúne 1.000 artigos e 3.045 perguntas; contém textos de tamanhos e áreas variadas; foi construído a partir de PDFs com texto, tabelas e imagens; e permite avaliar a busca sem executar uma LLM. Os PDFs originais são acessíveis por links, enquanto o conjunto também oferece versões convertidas em JSON. Isso permite estudar tanto a conversão quanto as demais etapas de RAG. Serão comparadas formas de dividir documentos, preparar perguntas e reorganizar resultados. A avaliação principal verificará se o sistema encontra a seção correta, usando as respostas já fornecidas pelo conjunto. Assim, o foco permanece no caminho percorrido pela informação, e não apenas na capacidade de uma LLM. Python, LangChain e OpenRouter serão usados nas etapas que precisarem de geração. Também serão registrados qualidade, tempo e custo.

Palavras-chave: RAG; LLM; recuperação de informação; chatbots

TITLE: Evaluation of retrieval-augmented generation techniques for chatbot applications

Abstract

Language models are programs able to understand questions and produce text, as in chatbots. Despite this ability, they may use outdated information or answer without support from sources. RAG addresses these problems by searching documents before answering and giving the retrieved passages to the chatbot. This work does not introduce a new technique. It compares published methods and the situations in which each works best. We selected the Open RAG Benchmark for four reasons: it contains 1,000 articles and 3,045 questions; its documents vary in size and subject; it was built from PDFs containing text, tables, and images; and retrieval can be evaluated without

running an LLM. Links provide access to the original PDFs, while converted JSON versions are also available. This supports experiments on both PDF conversion and the remaining RAG stages. We will compare ways of splitting documents, preparing questions, and reordering results. The main evaluation will check whether the correct section was found by using the labels already supplied with the dataset. The focus therefore remains on the information retrieval process rather than only on an LLM's ability. Python, LangChain, and OpenRouter will be used when generation is required. Quality, time, and cost will also be recorded.

Keywords: RAG; LLM; information retrieval; chatbots

Introdução

Um modelo de linguagem é um programa treinado com uma grande quantidade de textos para aprender relações entre palavras e ideias. Quando recebe uma pergunta, ele não consulta uma lista de respostas prontas. Ele calcula uma continuação provável para o texto. Os modelos de grande porte, conhecidos como LLMs, conseguem resumir, explicar conceitos e manter conversas. A arquitetura Transformer, apresentada por Vaswani et al. (2017), foi uma das bases desse avanço.

É possível executar um modelo de pesos abertos em servidor próprio ou acessá-lo por uma API, serviço que recebe a pergunta e devolve a resposta. Nesse caso, o custo costuma depender da quantidade de texto processada.

Esses modelos não possuem uma enciclopédia atualizada dentro deles. Por isso, podem não conhecer documentos privados, normas recentes ou artigos publicados depois do treinamento. Também podem produzir uma resposta convincente, mas incorreta. Esse erro é chamado de alucinação.

RAG é a sigla em inglês para geração aumentada por recuperação. A ideia se parece com permitir que uma pessoa consulte livros antes de responder a uma prova. Primeiro, o sistema pesquisa em uma coleção de documentos. Em seguida, entrega ao modelo a pergunta e os trechos encontrados. O modelo usa esse material para escrever a resposta (LEWIS et al., 2020). Assim, ele não depende apenas do que aprendeu durante o treinamento.

Documentos longos precisam ser divididos em partes menores, chamadas de chunks ou blocos. Uma divisão ruim pode separar uma explicação de sua conclusão. A pergunta também pode dificultar a busca. Por exemplo, “Como ele funciona?” pode ser reescrita como “Como o sistema RAG funciona?”.

Depois da primeira busca, outra etapa pode reler a pergunta e reorganizar os trechos encontrados. Essa etapa é chamada de reranking. Seu papel é colocar no início da lista os trechos que parecem mais úteis para responder à pergunta.

RAG facilita a atualização das fontes, o acesso a documentos privados e a apresentação dos textos que apoiam a resposta. Também pode ajudar modelos menores, rápidos e

baratos. Isso não significa que sempre alcançarão um modelo maior, pois ainda precisam compreender e relacionar as informações.

Este trabalho não apresenta uma técnica nova. O objetivo é reunir métodos publicados e testá-los nas mesmas condições. Serão comparadas formas de dividir documentos, preparar perguntas e reorganizar os resultados da busca. Também serão considerados a qualidade, o tempo necessário e o custo.

O Open RAG Benchmark foi escolhido pela quantidade e variedade. Ele reúne mil artigos de diferentes áreas e mais de três mil perguntas. Os documentos processados variam de cerca de 1 kB a 9,4 MB, com mediana próxima de 471 kB. Isso permite testar conteúdos curtos e documentos maiores.

O conjunto foi construído a partir de PDFs, formato comum que pode misturar texto, tabelas, imagens e fórmulas. O repositório fornece links para os PDFs originais e distribui o conteúdo convertido em JSON. Assim, é possível usar a versão pronta ou comparar OCR, extração de tabelas, tratamento de imagens e outras formas de conversão.

Por fim, o teste principal pode ser executado sem uma LLM geradora, pois o conjunto informa o documento e a seção corretos. É possível avaliar divisão, indexação, busca e ordenação com métodos locais e baratos. A LLM continua necessária para reescrever perguntas e produzir respostas finais, mas não precisa ser chamada em todas as etapas. Assim, mede-se o sistema de RAG sem confundir a busca com a capacidade de escrita de um modelo.

Metodologia

A pesquisa será feita como uma comparação controlada. Uma parte do sistema será modificada por vez, enquanto as demais permanecerão iguais. Isso ajuda a identificar qual mudança causou diferença no resultado. O programa será desenvolvido em Python. O LangChain será usado para ligar as etapas do chatbot, e o OpenRouter permitirá acessar diferentes modelos quando uma experiência precisar de geração de texto.

O Open RAG Benchmark foi criado a partir de 1.000 artigos científicos em PDF e possui 3.045 perguntas com respostas. O repositório fornece links para os PDFs, e o Hugging Face distribui versões em JSON com texto, tabelas e imagens. Há 400 documentos com respostas e 600 negativos difíceis, que tratam de assuntos possivelmente parecidos, mas não respondem à pergunta.

Primeiro será construído um sistema de referência. Nele, os documentos serão divididos em blocos com tamanho fixo. Uma pequena parte do texto será repetida entre blocos vizinhos para reduzir o risco de cortar uma explicação no meio. Cada bloco será convertido em uma lista de números que procura representar seu significado. Essa representação é chamada de embedding. A pergunta passa pelo mesmo processo. O

sistema compara os números e recupera os blocos mais próximos do sentido da pergunta.

Depois serão testadas três formas de divisão:

1. Janela deslizante: cria blocos de tamanho fixo com uma pequena repetição entre eles. É simples e rápida.
2. Divisão semântica: tenta identificar mudanças de assunto e separar o texto nesses pontos. Pode manter uma explicação inteira no mesmo bloco, mas depende da qualidade dessa identificação.
3. Divisão hierárquica: mantém blocos pequenos e resumos de partes maiores, como seções e subseções. Pode ajudar quando a resposta depende de várias partes do artigo. O método RAPTOR é uma referência para essa estratégia (SARTHI et al., 2024).

Também serão comparadas formas de preparar a pergunta antes da busca:

• usar a pergunta original;
• separar as palavras mais importantes;
• esclarecer referências vagas e perguntas ambíguas;
• criar várias versões da mesma pergunta e reunir os resultados;
• usar HyDE, técnica em que o modelo escreve um texto parecido com uma possível resposta e procura documentos semelhantes a ele (GAO et al., 2023).

O texto criado pelo HyDE pode conter erros. Por isso, ele servirá apenas como pista para a busca e não como fonte da resposta final.

Após a busca inicial, um reranker avaliará novamente os trechos encontrados. O processo pode ser entendido como duas seleções. A primeira é rápida e separa um grupo de candidatos. A segunda compara a pergunta com cada candidato de forma mais cuidadosa e escolhe os melhores (NOGUEIRA; CHO, 2019). O projeto atual ainda não possui essa segunda etapa.

O teste principal termina na recuperação. O arquivo qrels.json já indica o documento e a seção corretos para cada pergunta. Por isso, é possível verificar se o sistema encontrou a informação certa sem pedir que uma LLM escreva ou julgue a resposta. Essa escolha reduz custo e permite comparar diretamente as etapas de divisão, representação, busca e ordenação. Métodos como busca por palavras e modelos locais de representação também evitam cobrança por API.

Os testes verificarão quantas vezes o trecho correto foi encontrado e em que posição apareceu. Na área de recuperação de informação, algumas dessas medidas recebem os nomes Recall, MRR e nDCG. Os resultados serão apresentados com uma explicação em linguagem comum. Tempo e custo também serão registrados.

Em uma análise complementar, modelos menores e maiores responderão às mesmas perguntas, com e sem os documentos recuperados. Essa etapa e as reescritas de perguntas usarão LLMs e serão medidas separadamente. Assim, o resultado principal não dependerá de um único modelo gerador.

Resultados e Discussões

A pesquisa ainda está em andamento. Os números disponíveis neste momento descrevem o material usado no teste. Eles ainda não mostram a melhoria causada por cada técnica. Não foram encontrados no repositório arquivos com comparações entre as estratégias propostas. Por esse motivo, não são apresentados percentuais de ganho antes da execução dos experimentos.

O Open RAG Benchmark foi construído a partir de 1.000 artigos científicos em PDF. Desses, 400 contêm respostas e 600 são negativos difíceis. Eles verificam se o sistema rejeita textos que parecem relacionados, mas não resolvem a dúvida do usuário.

O conjunto possui 3.045 perguntas. Há 1.793 que exigem explicação ou reunião de informações e 1.252 com respostas diretas. A maior parte utiliza texto, mas algumas dependem de imagens e tabelas. Uma técnica pode funcionar bem para encontrar uma frase e ter dificuldade quando a resposta está espalhada.

Os mil documentos processados ocupam cerca de 742 MB. O menor tem 1 kB, a mediana é 471 kB e o maior chega a 9,4 MB. Há 110 arquivos abaixo de 100 kB, 644 entre 100 kB e 1 MB e 246 acima de 1 MB. Os valores se referem aos JSONs, que podem incluir imagens, e não aos PDFs. Ainda assim, mostram que o corpus não é uniforme.

O código atual baixa e processa os PDFs, extrai texto, tabelas e imagens, cria perguntas e remove exemplos de baixa qualidade. Também produz representações numéricas dos textos e seleciona documentos difíceis para a busca. Essas atividades preparam o conjunto de teste. Elas ainda não correspondem à comparação de um chatbot RAG completo.

Não foi encontrado um reranker no projeto. Atualmente, os trechos são ordenados pela proximidade entre a representação da pergunta e a representação de cada texto. O reranker planejado fará uma segunda análise dos candidatos. Portanto, ele é uma etapa que ainda será testada, não um recurso já usado no repositório.

Os estudos publicados indicam que nenhuma forma de dividir os documentos é melhor em todas as situações. A janela deslizante é simples e pode ser suficiente para perguntas sobre informações próximas. A divisão semântica tenta manter cada assunto no mesmo bloco. Ela pode ajudar em textos longos com mudanças claras de tema, mas um estudo recente com textos acadêmicos não encontrou vantagem constante sobre métodos simples (KREILEDER; REISINGER; FISCHER, 2026). A divisão hierárquica pode ajudar quando a resposta depende de partes diferentes do documento, embora exija mais processamento e possa perder detalhes ao resumir (SARTHI et al., 2024).

As formas de preparar a pergunta também tratam de problemas diferentes. Palavras-chave podem ajudar quando a pergunta contém nomes e termos técnicos. O esclarecimento é útil em frases como “Qual foi o resultado dele?”, nas quais não está claro a quem “dele” se refere. Criar várias versões aumenta a chance de encontrar a resposta, mas também pode trazer material desnecessário. HyDE aproxima uma pergunta curta da forma encontrada nos documentos. Se o texto imaginado estiver errado, porém, a busca pode seguir para o assunto incorreto (GAO et al., 2023).

O conjunto já fornece a localização correta para cada pergunta. Assim, a comparação principal não exige que uma LLM gere respostas nem que outra LLM dê notas. O que será medido diretamente é se o caminho de RAG encontrou e ordenou a evidência correta. Isso torna a execução mais barata e evita atribuir à busca um resultado que, na verdade, veio da capacidade do modelo gerador. Experimentos com reescrita de perguntas e geração de respostas serão apresentados como etapas adicionais, pois esses casos usam LLM.

Os experimentos deverão mostrar em quais perguntas cada método ajuda ou atrapalha. Qualquer ganho será apresentado junto do tempo e do custo adicionais. Uma técnica um pouco mais precisa, mas muito lenta, pode não ser adequada a um chatbot que precisa responder rapidamente.

Conclusão

A qualidade de um chatbot não depende apenas do modelo que escreve a resposta. Ela também depende dos documentos disponíveis, da forma como eles foram divididos, da pesquisa feita com a pergunta e da escolha dos trechos enviados ao modelo. RAG organiza essas etapas para que a resposta use informações externas e verificáveis.

Este trabalho reúne técnicas existentes e as compara nas mesmas condições. Não há proposta de um método novo. A intenção é produzir orientações práticas: quando uma divisão simples é suficiente, quando vale a pena preservar a estrutura do documento, quando uma pergunta precisa ser esclarecida e quando uma segunda ordenação dos resultados compensa seu custo.

O Open RAG Benchmark é adequado para o estudo pela quantidade e pela variedade. Ele reúne mil artigos, mais de três mil perguntas, tabelas, imagens e documentos difíceis de distinguir. Foi criado a partir de PDFs reais e permite voltar aos arquivos originais para testar formas de conversão. Ao mesmo tempo, oferece versões processadas para quem deseja começar diretamente pela busca. No entanto, o repositório atual contém a preparação do conjunto e não os resultados dos experimentos descritos. Também não possui um reranker. Os números apresentados até agora caracterizam o material de teste, mas não comprovam melhoria de nenhuma estratégia.

As próximas etapas serão executar cada comparação e medir se o sistema encontrou o trecho correto, quanto tempo levou e quanto custou. Essa avaliação pode ser realizada

sem uma LLM, pois o conjunto já informa a localização esperada. Isso mantém o foco nas etapas de RAG e reduz o custo da maior parte dos testes. Em uma etapa complementar, modelos maiores e menores serão comparados na produção das respostas. A hipótese é que boas fontes possam tornar um modelo menor mais competitivo, mas isso precisa ser verificado.

A conclusão final não deverá apontar uma solução para todos os casos. Blocos fixos podem ser rápidos e suficientes para perguntas diretas. A divisão por assunto pode preservar explicações. Uma organização por seções e resumos pode ajudar perguntas que usam partes diferentes do artigo. O reranker pode melhorar a ordem quando o trecho correto já aparece entre os candidatos. O resultado esperado é um guia com situações de uso, limitações e custos.

Referências

GAO, Luyu et al. Precise zero-shot dense retrieval without relevance labels. In: Proceedings of ACL 2023. 2023. Disponível em: <https://arxiv.org/abs/2212.10496>. Acesso em: 31 ago. 2026.

GÜNTHER, Michael et al. Late chunking: contextual chunk embeddings using long-context embedding models. 2024. Disponível em: <https://arxiv.org/abs/2409.04701>. Acesso em: 31 ago. 2026.

KREILEDER, Valentin J. J.; REISINGER, Johannes; FISCHER, Andreas. Evaluating chunking strategies for retrieval-augmented generation on academic texts. 2026. Disponível em: <https://arxiv.org/abs/2607.01852>. Acesso em: 31 ago. 2026.

LEWIS, Patrick et al. Retrieval-augmented generation for knowledge-intensive NLP tasks. In: Advances in Neural Information Processing Systems, v. 33, 2020. Disponível em: <https://arxiv.org/abs/2005.11401>. Acesso em: 31 ago. 2026.

MA, Xinbei et al. Query rewriting for retrieval-augmented large language models. In: Proceedings of EMNLP 2023. 2023. Disponível em: <https://arxiv.org/abs/2305.14283>. Acesso em: 31 ago. 2026.

NOGUEIRA, Rodrigo; CHO, Kyunghyun. Passage re-ranking with BERT. 2019. Disponível em: <https://arxiv.org/abs/1901.04085>. Acesso em: 31 ago. 2026.

NOGUEIRA, Rodrigo; JIANG, Zhiying; LIN, Jimmy. Document ranking with a pretrained sequence-to-sequence model. 2020. Disponível em: <https://arxiv.org/abs/2003.06713>. Acesso em: 31 ago. 2026.

SARTHI, Parth et al. RAPTOR: recursive abstractive processing for tree-organized retrieval. In: International Conference on Learning Representations. 2024. Disponível em: <https://arxiv.org/abs/2401.18059>. Acesso em: 31 ago. 2026.

VASWANI, Ashish et al. Attention is all you need. In: Advances in Neural Information Processing Systems, v. 30, 2017. Disponível em: <https://arxiv.org/abs/1706.03762>. Acesso em: 31 ago. 2026.

VECTARA. Open RAG Benchmark: a multimodal PDF dataset for comprehensive RAG evaluation. Disponível em: <https://github.com/vectara/open-rag-bench>. Acesso em: 31 ago. 2026.

VECTARA. Open RAGBench dataset. Hugging Face, 2025. Disponível em: <https://huggingface.co/datasets/vectara/open_ragbench>. Acesso em: 31 ago. 2026.

CÓDIGO: HS0605

AUTOR: LEONARDO XAVIER CRUZ FILHO

ORIENTADOR: LUCELIO DANTAS DE AQUINO

TÍTULO: O USO DO CHATGPT NAS PRÁTICAS DE LEITURA E ESCRITA NO ENSINO SUPERIOR: UMA REVISÃO SISTEMÁTICA DA LITERATURA

Resumo

A expansão da Inteligência Artificial Generativa (IAG), especialmente do ChatGPT, tem repercutido nas práticas de leitura e escrita no Ensino Superior. Este estudo, vinculado a um plano de Iniciação Científica no Bacharelado em Tecnologia da Informação (BTI), objetiva mapear e analisar usos e repercussões didático-pedagógicas do ChatGPT na produção científica recente, por meio de uma Revisão Sistemática da Literatura (RSL). A revisão seguiu Kitchenham e Charters (2007) e a estratégia PICOC, com buscas em seis bases: ACM Digital Library, IEEE Digital Library, Web of Science, SciELO, ScienceDirect e Scopus. Dos 160 registros identificados, 8 estudos empíricos atenderam aos critérios de elegibilidade. A avaliação metodológica, baseada em 15 critérios, classificou os estudos como de alta qualidade, com pontuações entre 10,5 e 15,0. A extração dos dados foi concluída, e a síntese qualitativa encontra-se em andamento, orientada pelas questões de pesquisa do protocolo. Os resultados deverão subsidiar o planejamento de futuras intervenções didáticas no contexto do BTI.

Palavras-chave: ChatGPT. Escrita acadêmica. Letramento crítico. Ensino Superior.

TITLE: THE USE OF CHATGPT IN READING AND WRITING PRACTICES IN HIGHER EDUCATION: A SYSTEMATIC LITERATURE REVIEW

Abstract

The expansion of Generative Artificial Intelligence (GenAI), particularly ChatGPT, has affected reading and writing practices in Higher Education. This study, linked to an Undergraduate Research project in the Bachelor's Degree in Information Technology (BTI), aims to map and analyze the didactic and pedagogical uses and implications of ChatGPT reported in recent scientific literature through a Systematic Literature Review (SLR). The review followed Kitchenham and Charters (2007) and the PICOC strategy, with searches conducted in six databases: ACM Digital Library, IEEE Digital Library,

Web of Science, SciELO, ScienceDirect, and Scopus. Of the 160 records identified, 8 empirical studies met the eligibility criteria. The methodological quality assessment, based on 15 criteria, classified all studies as high quality, with scores ranging from 10.5 to 15.0. Data extraction has been completed, and the qualitative synthesis is currently underway, guided by the research questions defined in the review protocol. The results are expected to support the planning of future didactic interventions in the BTI context.

Keywords: ChatGPT. Academic Writing. Critical Literacy. Higher Education.

Introdução

O avanço das Tecnologias Digitais de Informação e Comunicação (TDIC) e a expansão das ferramentas de Inteligência Artificial Generativa (IAG), impulsionada pelo lançamento do ChatGPT, têm gerado transformações profundas no Ensino Superior (Alves, 2023; Luckin et al., 2016). Diante disso, acredita-se que, no Bacharelado em Tecnologia da Informação (BTI), a familiaridade instrumental pode acelerar a apropriação espontânea e, por vezes, desorientada dessas ferramentas pelos estudantes. Essa realidade demanda uma reflexão sobre os impactos da IAG nas práticas de leitura e escrita acadêmica, fundamentais para a construção de conhecimento e do desenvolvimento intelectual no meio universitário.

O ChatGPT, ao mimetizar a linguagem natural com base em cálculos estatísticos de coocorrência de termos, oferece recursos úteis como a geração de esboços, síntese e revisão textual. Contudo, pesquisas fundamentais alertam que os Grandes Modelos de Linguagem (do inglês Large Language Model - LLM), agem essencialmente como “papagaios estocásticos” (Bender et al., 2021 apud Freire; Santos, 2023). Eles reproduzem padrões encontrados em massas de dados de treinamento, mas são incapazes de compreender de fato o significado ou as implicações do texto produzido. Dentre as limitações severas documentadas, destacam-se a inferioridade do repertório lexical em comparação com a escrita humana (Silva; Silva, 2025) e as recorrentes “alucinações” ou confabulações, nas quais o algoritmo gera conceitos incorretos e inventa referências inexistentes (Sampaio; Sabbatini; Limongi, 2024).

Frente a esses limites cognitivos da máquina, as práticas de escrita no Ensino Superior devem ser analisadas sob a ótica dos letramentos como práticas sociais de interação contextuais, situadas e indissociáveis das relações de poder e disputas discursivas, contrapondo-se à visão puramente tecnicista de leitura e escrita como habilidades neutras (Street, 2014). Nesse cenário, o letramento digital e crítico ganha centralidade ao exigir que os discentes não sejam usuários passivos da tecnologia, mas agentes ativos capazes de analisar e desconstruir discursos (Rojo; Moura, 2012).

Emergem, assim, discussões cruciais envolvendo autoria, agência humana, ética e mediação pedagógica. Sob a ótica da integridade científica, a IAG não pode ser listada como coautora de trabalhos científicos, visto que a autoria exige personalidade jurídica e responsabilidade moral. A mediação pedagógica deve apoiar-se no princípio "human-in-the-loop", assegurando que o controle final permaneça com o ser humano (Sampaio;

Sabbatini; Limongi, 2024). A tecnologia deve ser posicionada como assistente complementar de pesquisa, e não como substituta da cognição, promovendo a autonomia discente (Luckin et al., 2016).

Diante do dinamismo tecnológico e da aparente incipiência de regramentos específicos por agências de fomento brasileiras (Sampaio; Sabbatini; Limongi, 2024), mostra-se pertinente investigar sistematicamente a produção científica sobre o tema. Coerente com essa demanda, este trabalho vincula-se ao plano de trabalho "Mapeamento, análise e avaliação dos usos do ChatGPT nas práticas de leitura e escrita no BTI" (UFRN, 2026b), integrado ao projeto de pesquisa "O ChatGPT no ensino e aprendizagem de leitura e escrita em Tecnologia da Informação: reflexões, desafios e potencialidades". O objetivo específico é realizar uma Revisão Sistemática da Literatura (RSL) para identificar, mapear e avaliar os usos e repercussões didático-pedagógicas do ChatGPT documentados na produção científica recente, fornecendo subsídios para as etapas subsequentes da pesquisa-ação no BTI.

Metodologia

O delineamento metodológico deste trabalho constitui-se como uma Revisão Sistemática da Literatura (RSL), caracterizada como um estudo secundário focado em sintetizar evidências de estudos primários de maneira imparcial, rigorosa e replicável (Kitchenham; Charters, 2007; Dermeval; Coelho; Bittencourt, 2018). O processo estruturou-se em três fases sucessivas: planejamento, condução e relatório. Para apoiar o gerenciamento e a execução colaborativa de todo o protocolo, utilizou-se a plataforma digital e colaborativa baseada na web Parsifal.

Fase de Planejamento: objetivos e questões de pesquisa (QP)

O objetivo central definido no protocolo é analisar sistematicamente as evidências científicas sobre os usos do ChatGPT no ensino e na aprendizagem da leitura e da escrita no Ensino Superior, com especial atenção às práticas de produção, revisão e avaliação de textos argumentativos. Para responder a esse objetivo, estabeleceram-se cinco questões de pesquisa norteadoras:

- QP1: Quais atividades pedagógicas de leitura e escrita utilizam o ChatGPT?
- QP2: Como a ferramenta é empregada no planejamento, na produção, na revisão e na reescrita textual?
- QP3: Quais efeitos são identificados na argumentação, na qualidade textual, na autoria, na criticidade e na autonomia dos estudantes?
- QP4: Quais desafios éticos e pedagógicos são relatados?
- QP5: Quais práticas podem subsidiar a elaboração de uma sequência didática?

Estruturação do PICOC e String de Busca

O escopo de busca foi delimitado com base na estratégia PICOC (População, Intervenção, Comparação, Resultado/Outcome e Contexto), contemplando como população estudantes do ensino superior, de graduação e de pós-graduação; como intervenção, o ChatGPT e a inteligência artificial generativa; como resultado, aspectos relacionados à leitura, leitura crítica, escrita, escrita acadêmica e argumentativa, textos argumentativos, autoria e pensamento crítico; e, como contexto, o ensino superior, incluindo universidade, faculdade, graduação e pós-graduação. A dimensão de comparação não se aplica ao estudo. Os termos foram empregados em português e inglês para ampliar a abrangência da busca.

A partir da combinação de descritores e de operadores booleanos, estruturou-se a seguinte search string unificada: ("ChatGPT") AND ("argumentative texts" OR "argumentative writing" OR "escrita acadêmica" OR "leitura" OR "leitura crítica") AND ("ensino superior" OR "higher education"). A string foi configurada para rastreamento em seis importantes bases de dados científicas indexadas: ACM Digital Library, IEEE Digital Library, ISI Web of Science, SciELO, ScienceDirect e Scopus.

Fase de Condução

A triagem desenvolveu-se de forma sistemática e em duas etapas consecutivas para mitigar ruídos subjetivos. Na primeira etapa, realizou-se a leitura independente de títulos, palavras-chave e resumos (abstracts). Na segunda etapa, os artigos que não foram eliminados passaram pela análise por texto completo. A fim de garantir a confiabilidade e reduzir o viés de seleção, o processo foi conduzido de forma independente por mais de um pesquisador, resolvendo eventuais discordâncias e desajustes por meio de sessões de consenso.

Critérios de Inclusão (CI) e Exclusão (CE)

Critérios de Inclusão (CI):

1. Abordem leitura e escrita no Ensino Superior;
2. Desenvolvam intervenção pedagógica;
3. Investiguem o ChatGPT em práticas educacionais;
4. Tenham sido publicados nos idiomas português, inglês ou espanhol.

Critérios de Exclusão (CE):

1. Artigos científicos fechados (ou de acesso restrito);
2. Duplicatas e publicações sem o texto completo disponível;
3. Editoriais, notícias, resumos simples e textos de opinião;
4. Estudos publicados em período anterior ao ano de 2022;
5. Estudos focados apenas sobre detecção de textos produzidos por inteligência artificial;
6. Pesquisas focadas exclusivamente em programação ou geração de código de computador;
7. Outras revisões sistemáticas da literatura;
8. Ensaios teóricos ou estudos de natureza não empírica.

Avaliação de Qualidade

A avaliação de qualidade dos artigos selecionados é uma etapa crítica para determinar a validade interna, a força das evidências encontradas e a credibilidade das sínteses resultantes. Para este estudo, o protocolo estabeleceu um checklist de avaliação metodológica composto por 15 perguntas (Ver Anexo A - Quadro 1 - Critérios de avaliação da qualidade dos estudos).

Cada critério foi pontuado de maneira independente pelos pesquisadores, utilizando pesos específicos associados às respostas: Alta qualidade (peso 1.0), Qualidade moderada (peso 0.5) e Baixa qualidade (peso 0.0). O somatório total resulta em uma pontuação máxima de 15.0 pontos por estudo. Estabeleceu-se uma nota de corte de 10.0 para que o estudo primário fosse incluído na etapa de síntese de dados.

Extração e Síntese dos Dados

A etapa de extração dos dados foi concluída, e as informações obtidas encontram-se devidamente organizadas e tabuladas. No momento, está em andamento a etapa de síntese, construída a partir de uma leitura aprofundada e analítica dos dados extraídos, com o objetivo de responder às questões de pesquisa apresentadas neste protocolo. Desse modo, a apresentação e a discussão sistematizada dos resultados da RSL serão realizadas em momento posterior, após a conclusão dessa etapa de análise.

A seguir, por meio de gráficos e ilustrações, apresentamos os dados da RSL obtidos.

Resultados e Discussões

Considerando que a etapa de síntese qualitativa dos dados extraídos dos estudos selecionados ainda se encontra em desenvolvimento, esta seção apresenta e discute os resultados consolidados referentes às etapas de identificação, triagem, elegibilidade e avaliação da qualidade metodológica dos estudos que compõem o corpus da Revisão Sistemática da Literatura (RSL). Esses resultados permitem caracterizar a abrangência da estratégia de busca, o processo de refinamento do corpus e a consistência metodológica dos estudos selecionados, estabelecendo as bases para a posterior análise das questões de pesquisa definidas no protocolo.

Fluxo de seleção dos estudos

A estratégia de busca aplicada às seis bases de dados selecionadas resultou na identificação inicial de 160 registros. Conforme apresentado na Figura 1 (Anexo B - Figura 1 – Fluxo de seleção dos estudos segundo o protocolo PRISMA), a maior quantidade de estudos foi recuperada na ScienceDirect, com 79 registros, seguida pela ACM Digital Library, com 61. As demais bases apresentaram quantitativos consideravelmente menores: Scopus, com 12 registros; ISI Web of Science, com 5; SciELO, com 2; e IEEE Digital Library, com 1.

A concentração dos resultados em ScienceDirect e ACM Digital Library é expressiva: juntas, essas duas bases responderam por 140 dos 160 registros identificados, correspondendo a 87,5% do total recuperado. Esse resultado evidencia a importância dessas fontes para a cobertura inicial da temática investigada, embora o elevado número de registros recuperados não tenha implicado, necessariamente, maior quantidade de estudos elegíveis ao final do processo.

Na etapa de identificação foram removidos 16 registros duplicados, equivalentes a 10% do conjunto inicialmente recuperado. A Scopus concentrou o maior número de duplicatas, com oito registros, seguida pela ScienceDirect, com quatro, pela Web of Science, com três, e pela SciELO, com um. Após essa depuração, 144 registros únicos permaneceram para a etapa de triagem.

Para a etapa de triagem foi realizada a leitura de títulos, resumos e palavras-chave, o que levou à exclusão de 136 estudos, correspondentes a aproximadamente 94,4% dos registros submetidos à triagem. As exclusões decorreram da aplicação dos critérios previamente definidos no protocolo, incluindo estudos que não apresentavam caráter empírico, não investigavam práticas de leitura ou escrita no Ensino Superior ou abordavam a Inteligência Artificial sob perspectivas predominantemente técnicas, desvinculadas do foco pedagógico da revisão.

Como resultado desse processo, 8 estudos atenderam aos critérios de elegibilidade e foram mantidos no corpus da revisão, correspondendo a 5,6% dos 144 registros únicos triados e a 5% dos 160 registros inicialmente identificados. A redução observada entre a identificação e a inclusão final evidencia o grau de especificidade do recorte adotado na RSL, direcionado a estudos empíricos que articulam o uso do ChatGPT a práticas de leitura, escrita e produção acadêmica no Ensino Superior. Embora esse resultado não permita afirmar, de forma ampla, uma baixa produtividade científica sobre Inteligência Artificial Generativa nesse nível de ensino, ele sinaliza uma produção ainda restrita quando se consideram investigações empíricas voltadas especificamente às práticas de linguagem.

Tal cenário sugere a existência de uma lacuna que demanda maior investimento investigativo, sobretudo em estudos que ultrapassem abordagens centradas apenas na eficiência, no desempenho ou na adoção instrumental dessas tecnologias e incorporem uma perspectiva crítica sobre seus usos pedagógicos. Nesse sentido, mostram-se particularmente relevantes pesquisas que problematizem aspectos relacionados à autoria, autonomia, pensamento crítico, dependência tecnológica, vieses algorítmicos, transparência, ética e implicações da mediação da IAG nos processos de leitura e escrita acadêmica. A ampliação de estudos com esse enfoque poderá contribuir para uma compreensão mais consistente das condições, dos limites e das possibilidades de integração crítica da Inteligência Artificial Generativa às práticas formativas no Ensino Superior.

Distribuição dos estudos por base de dados e status da triagem

A Figura 2 (Anexo C - Figura 2 – Distribuição de artigos por base de dados e status da triagem) permite observar comparativamente a contribuição de cada base ao longo das etapas de seleção, distinguindo registros aceitos, rejeitados e duplicados.

A ScienceDirect apresentou não apenas o maior volume inicial de registros, mas também a maior contribuição para o corpus final. Dos 79 estudos recuperados nessa base, quatro foram identificados como duplicados e 69 foram excluídos após a triagem, resultando em 6 estudos incluídos. Assim, a ScienceDirect respondeu por 75% dos oito estudos que compõem o corpus final da revisão.

A ACM Digital Library apresentou comportamento distinto. Embora tenha fornecido 61 dos registros inicialmente identificados, apenas 1 estudo permaneceu após a aplicação dos critérios de seleção, enquanto 60 foram excluídos. Esse resultado sugere que a elevada recuperação inicial nessa base esteve associada a uma quantidade significativa de estudos situados em áreas próximas ao objeto da revisão, mas que não satisfaziam integralmente o recorte empírico e pedagógico definido no protocolo.

A Scopus, por sua vez, apresentou 12 registros iniciais, dos quais oito foram removidos por duplicidade. Entre os quatro registros únicos restantes, três foram excluídos e um estudo foi incluído no corpus final. As bases Web of Science, SciELO e IEEE Digital Library não contribuíram com estudos para a síntese qualitativa, uma vez que os registros recuperados foram eliminados como duplicados ou excluídos na etapa de triagem.

A distribuição observada reforça a importância da utilização de múltiplas bases em uma revisão sistemática. Embora apenas três fontes tenham contribuído diretamente para os oito estudos incluídos, a busca em diferentes indexadores amplia a cobertura da literatura e reduz o risco de que estudos potencialmente relevantes sejam omitidos. Além disso, a presença de duplicatas entre as bases evidencia a sobreposição de indexação existente entre diferentes plataformas, aspecto esperado em estratégias de busca abrangentes.

Os resultados também mostram que a quantidade de registros recuperados por uma base não deve ser interpretada isoladamente como indicador de sua relevância para o objeto investigado. A contribuição efetiva depende da correspondência entre o conteúdo dos estudos indexados e os critérios de elegibilidade definidos no protocolo. Nesse sentido, a ScienceDirect apresentou simultaneamente maior volume de recuperação e maior contribuição para o corpus, enquanto a ACM Digital Library, apesar da expressiva quantidade inicial de registros, apresentou baixa proporção de estudos elegíveis.

Avaliação da qualidade metodológica

Os oito estudos selecionados (ver Anexo D Quadro 2 - Informações básicas dos estudos selecionados) foram submetidos à avaliação da qualidade metodológica mediante a aplicação de um instrumento composto por 15 critérios, com pontuação máxima de 15,0 pontos. Conforme estabelecido no protocolo, 10,0 pontos constituíram a nota mínima para inclusão do estudo na etapa de síntese. Veja-se o

Todos os estudos alcançaram pontuação superior à nota de corte estabelecida, com valores entre 10,5 e 15,0 pontos (ver Anexo E - Figura 3 – Pontuação de avaliação da qualidade dos estudos selecionados). A maior pontuação foi obtida pelo estudo Leveraging ChatGPT in academic writing: ChatGPT enhances students' writing quality,

writing experience, and ownership (Bauer et al., 2026), que alcançou 15,0 pontos. Na sequência, destacam-se An AI-assisted critical thinking intervention to enhance undergraduate EFL learners' writing proficiency (Yin; Dou, 2025), com 14,0 pontos, e You've Got AI Friend in Me: LLMs as Collaborative Learning Partners (Oppenheimer; Cash; Connell Pensky, 2025), com 13,5 pontos.

Os demais estudos apresentaram as seguintes pontuações: The role of ChatGPT and Grammarly in promoting emotion regulation, psychological well-being, motivation, and academic writing in Chinese college students: a self-determination theory perspective (Wang; Wang; Xu, 2025), com 12,5; Exploring the Effects of Generative AI Assistance on Writing Self-efficacy (Song et al., 2026), com 12,0; The Use of ChatGPT in Source-Based Writing Tasks (Tarchi et al., 2025), com 12,0; Exploring generative AI assisted-EFL academic writing: an activity theory perspective (Jia; Lu, 2026), com 11,5; e, Exploring AI integration within an eclectic pedagogy to enhance university students' genre writing skills across writing aspects (Robillos, 2025), com 10,5 pontos. A média do conjunto foi de aproximadamente 12,6 pontos em 15, e seis dos oito estudos, equivalentes a 75% do corpus, obtiveram pelo menos 12 pontos. Esses resultados indicam que o processo de seleção resultou em um conjunto de estudos que atende ao patamar de qualidade metodológica estabelecido no protocolo.

A variação entre 10,5 e 15,0 pontos, entretanto, indica diferentes níveis de atendimento aos critérios avaliados. Assim, embora todos os artigos tenham ultrapassado a nota mínima definida para inclusão, não se deve assumir homogeneidade absoluta quanto à qualidade de seus delineamentos ou à completude de seus relatos metodológicos. A pontuação agregada permite afirmar que os estudos apresentam condições metodológicas suficientes para integrar a síntese, enquanto eventuais fragilidades específicas deverão ser consideradas durante a interpretação dos achados de cada estudo. Esse aspecto é particularmente relevante para a etapa subsequente da revisão.

A avaliação da qualidade não constitui apenas um procedimento classificatório, mas fornece um parâmetro para a leitura crítica das evidências encontradas. Dessa forma, resultados provenientes de estudos com diferentes níveis de detalhamento metodológico poderão ser analisados considerando-se também a robustez das evidências que os sustentam.

Conclusão

Esta RSL permitiu mapear os usos e as repercussões didático-pedagógicas do ChatGPT em práticas de leitura e escrita acadêmica no Ensino Superior. Os 8 estudos empíricos selecionados apresentaram qualidade metodológica suficiente para integrar a síntese, mas o reduzido número de trabalhos elegíveis indica que esse campo ainda se encontra em consolidação, sobretudo quando se consideram investigações voltadas à escrita acadêmica, à autoria, ao pensamento crítico e à mediação pedagógica da Inteligência Artificial Generativa (IAG). Os resultados também evidenciam concentração das

publicações em determinadas bases e contextos, reforçando a necessidade de ampliar a diversidade de pesquisas, abordagens teóricas e cenários educacionais investigados.

Nesse sentido, destaca-se que não foram identificados, no corpus, estudos de pesquisadores brasileiros com aplicações desenvolvidas no contexto do Ensino Superior brasileiro. Esse resultado deve ser interpretado com cautela, uma vez que pode estar relacionado às bases de dados contempladas na estratégia de busca e aos critérios definidos no protocolo. Ainda assim, a ausência de estudos brasileiros entre os trabalhos elegíveis sinaliza uma lacuna relevante e aponta para a necessidade de pesquisas empíricas que investiguem, de modo situado, como a IAG vem sendo incorporada às práticas de leitura e escrita acadêmica nas instituições de Ensino Superior do país.

As evidências reunidas indicam ainda que o uso da IAG não deve ser analisado apenas sob a perspectiva de eficiência, produtividade ou melhoria do desempenho textual. Também é necessário problematizar seus efeitos sobre autonomia, autoria, criticidade e construção do conhecimento. Nessa direção, Moreira, Oliveira e Holanda (2024) alertam para os riscos de uma apropriação tecnocêntrica da IA, capaz de favorecer dependência tecnológica e enfraquecer o pensamento crítico quando a ferramenta é tratada como solução automática para atividades intelectuais complexas.

Essa perspectiva converge com Santos, Ribeiro e Aquino (2025), ao ressaltarem que a IA não deve ser compreendida como tecnologia neutra ou autossuficiente. Seu uso educacional exige sujeitos capazes de interpretar, questionar e avaliar criticamente as respostas produzidas pelos sistemas generativos. No contexto da escrita acadêmica, isso implica compreender o ChatGPT como recurso de apoio e mediação, e não como substituto da leitura, da argumentação, da autoria e da elaboração intelectual do estudante.

Como desdobramento deste trabalho, será concluída a análise qualitativa dos dados já extraídos, com o objetivo de responder às questões de pesquisa previstas no protocolo da RSL. Essa etapa aprofundará a compreensão sobre os usos do ChatGPT e fornecerá subsídios para futuras intervenções didático-pedagógicas no Ensino Superior, especialmente no contexto brasileiro, comprometidas com o uso ético, crítico e responsável da IA e com a preservação da autonomia intelectual e da formação acadêmica.

Referências

ALVES, Lynn (org.). Inteligência artificial e educação: refletindo sobre os desafios contemporâneos. Salvador: EDUFBA; Feira de Santana: UEFS Editora, 2023.

BAUER, Nicolas; FÜTTERER, Tim; BRUCKER, Birgit; GERJETS, Peter. Leveraging ChatGPT in academic writing: ChatGPT enhances students' writing quality, writing experience, and ownership. Computers and Education Open, v. 10, art. 100351, 2026. Disponível em: https://doi.org/10.1016/j.caeo.2026.100351. Acesso em: 7 ago. 2026.

DERMEVAL, Diego; COELHO, Jorge A. P. de M.; BITTENCOURT, Ig I. Mapeamento sistemático e revisão sistemática da literatura em Informática na Educação. In: JAQUES, Patrícia Augustin; SIQUEIRA, Sean; BITTENCOURT, Ig; PIMENTEL, Mariano (org.). Metodologia de pesquisa científica em Informática na Educação: abordagem quantitativa. Porto Alegre: SBC, 2020. v. 2, cap. 3, p. 1-26. Disponível em: https://ceie.sbc.org.br/metodologia/index.php/livro-2/. Acesso em: 21 ago. 2026.

FREIRE, Wendel; SANTOS, Edméa. Inteligência artificial generativa e os saberes científicos. In: ALVES, Lynn (org.). Inteligência artificial e educação: refletindo sobre os desafios contemporâneos. Salvador: EDUFBA; Feira de Santana: UEFS Editora, 2023. p. 123-135.

JIA, Chengyuan; LU, Yao. Exploring generative AI assisted-EFL academic writing: an activity theory perspective. Journal of English for Academic Purposes, v. 81, art. 101672, 2026. Disponível em: https://doi.org/10.1016/j.jeap.2026.101672. Acesso em: 7 ago. 2026.

KITCHENHAM, Barbara; CHARTERS, Stuart. Guidelines for performing Systematic Literature Reviews in Software Engineering. Keele: Keele University; Durham: Durham University, 2007. Technical Report EBSE 2007-001.

LUCKIN, Rose; HOLMES, Wayne; GRIFFITHS, Mark; FORCIER, Laurie B. Intelligence unleashed: an argument for AI in education. London: Pearson, 2016.

MOREIRA, Ailton Pinheiro; OLIVEIRA, Diego Bandeira de; HOLANDA, Wiron de Araújo. Inteligência artificial e letramentos críticos: o discurso tecnocêntrico em memes. Revista Linguagem em Foco, Fortaleza, v. 16, n. 2, p. 70-88, 2024. DOI: 10.46230/lef.v16i2.13090. Acesso em: 12 jun. 2026.

OPPENHEIMER, Daniel M.; CASH, Trent N.; CONNELL PENSKY, Allison E. You've Got AI Friend in Me: LLMs as collaborative learning partners. International Journal of Artificial Intelligence in Education, v. 35, p. 3896-3921, 2025. Disponível em: https://doi.org/10.1007/s40593-025-00521-6. Acesso em: 7 ago. 2026.

ROBILLOS, Roderick Julian. Exploring AI integration within an eclectic pedagogy to enhance university students' genre writing skills across writing aspects. Social Sciences & Humanities Open, v. 12, art. 101886, 2025. Disponível em: https://doi.org/10.1016/j.ssaho.2025.101886. Acesso em: 7 ago. 2026.

ROJO, Roxane; MOURA, Eduardo (org.). Multiletramentos na escola. São Paulo: Parábola Editorial, 2012.

SAMPAIO, Rafael Cardoso; SABBATINI, Marcelo; LIMONGI, Ricardo. Diretrizes para o uso ético e responsável da Inteligência Artificial Generativa: um guia prático para pesquisadores. São Paulo: Editora Intercom, 2024.

SANTOS, Alline Silva Medeiros dos; RIBEIRO, Janaina Raso; AQUINO, Lucélio Dantas de. Inteligência artificial na educação: uma análise crítica da obra de Lynn

Alves. Revista de Estudos de Cultura, São Cristóvão, v. 11, n. 27, p. 303-308, jan./jun. 2025. DOI: 10.32748/revec.v11i27.22690. Acesso em: 16 jul. 2026.

SILVA, Franklin Oliveira; SILVA, Francisca Silveline Pereira da. Inteligência artificial e modalidades argumentativas em textos gerados pelo ChatGPT. Revista de Letras, Fortaleza, v. 2, n. 43, p. 83-103, 2025. Disponível em: https://periodicos.ufc.br/revletras/article/view/94364. Acesso em: 21 ago. 2026.

SONG, Yejoon; KIM, Bandi; KWON, Yeju; PARK, Sung Jun. Exploring the effects of Generative AI Assistance on Writing Self-efficacy. In: CHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS – EXTENDED ABSTRACTS (CHI EA '26), 2026, Barcelona. Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems. New York: ACM, 2026. 7 p. Disponível em: https://doi.org/10.1145/3772363.3798610. Acesso em: 7 ago. 2026.

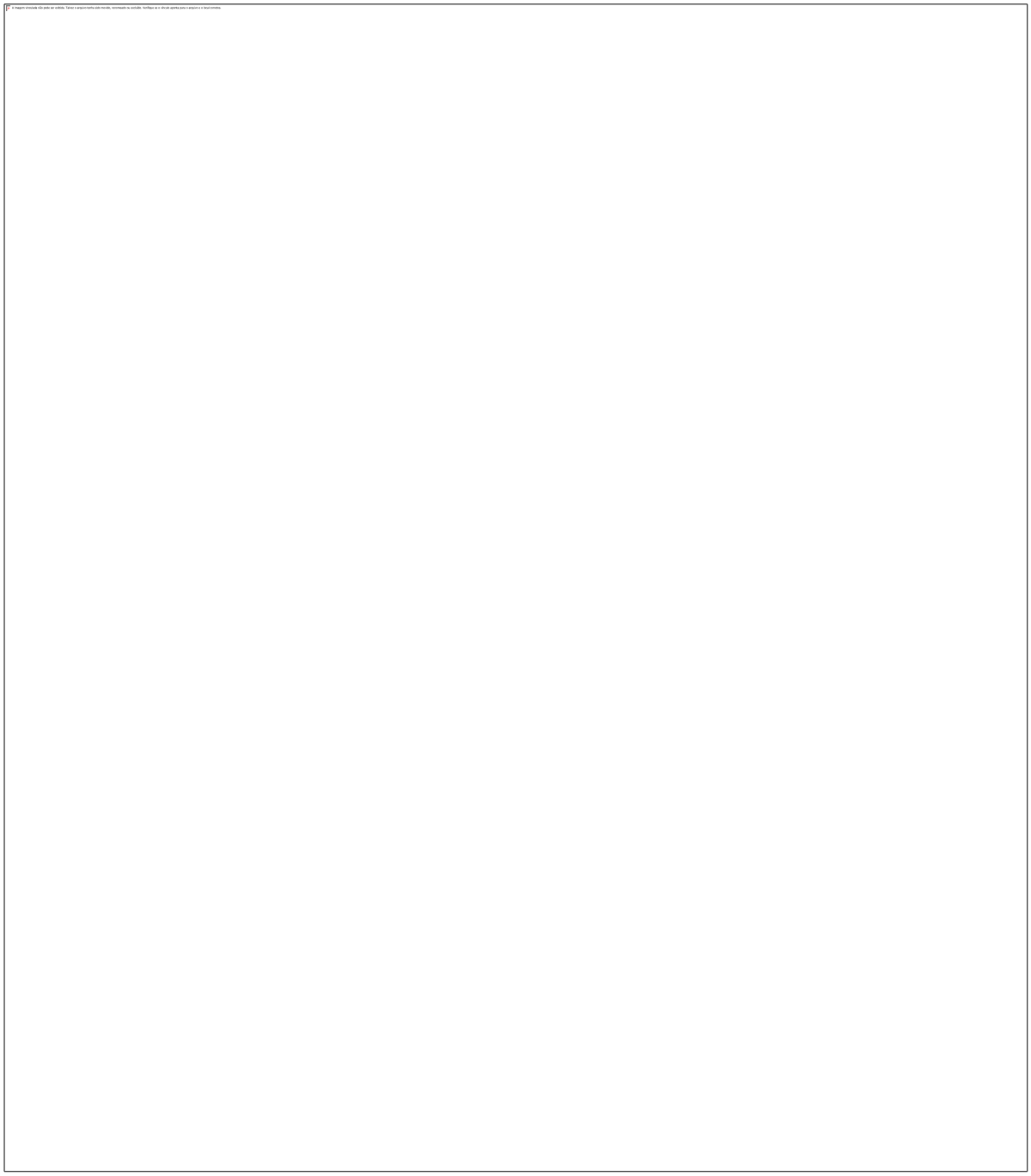
STREET, Brian. Letramentos sociais: abordagens críticas do letramento no desenvolvimento, na etnografia e na educação. Tradução de Marcos Bagno. São Paulo: Parábola Editorial, 2014.

TARCHI, Christian; ZAPPOLI, Alessandra; CASADO LEDESMA, Lidia; WENNÅS BRANTE, Eva. The Use of ChatGPT in Source-Based Writing Tasks. International Journal of Artificial Intelligence in Education, v. 35, p. 858-878, 2025. Disponível em: https://doi.org/10.1007/s40593-024-00413-1. Acesso em: 7 ago. 2026.

WANG, Caili; WANG, BingCheng; XU, Duoshu. The role of ChatGPT and Grammarly in promoting emotion regulation, psychological well-being, motivation, and academic writing in Chinese college students: a self-determination theory perspective. Learning and Motivation, v. 90, art. 102131, 2025. Disponível em: https://doi.org/10.1016/j.lmot.2025.102131. Acesso em: 7 ago. 2026.

YIN, Xue; DOU, Kun. An AI-assisted critical thinking intervention to enhance undergraduate EFL learners' writing proficiency. Studies in Educational Evaluation, v. 86, art. 101480, 2025. Disponível em: https://doi.org/10.1016/j.stueduc.2025.101480. Acesso em: 7 ago. 2026.

Anexos



Quadro 1 - Critérios de avaliação da qualidade dos estudos

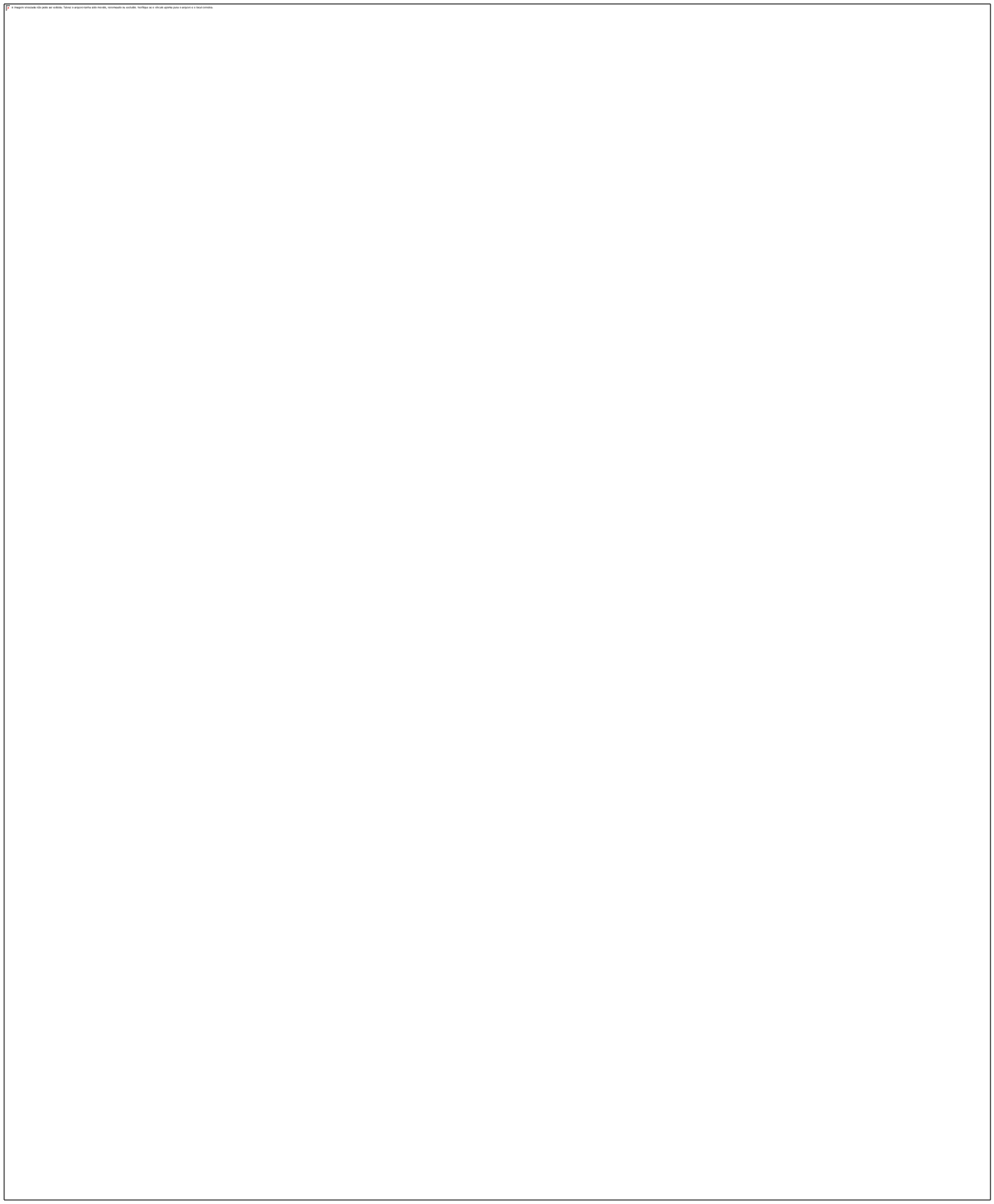


Figura 1 - Fluxo de seleção dos estudos segundo o protocolo PRISMA

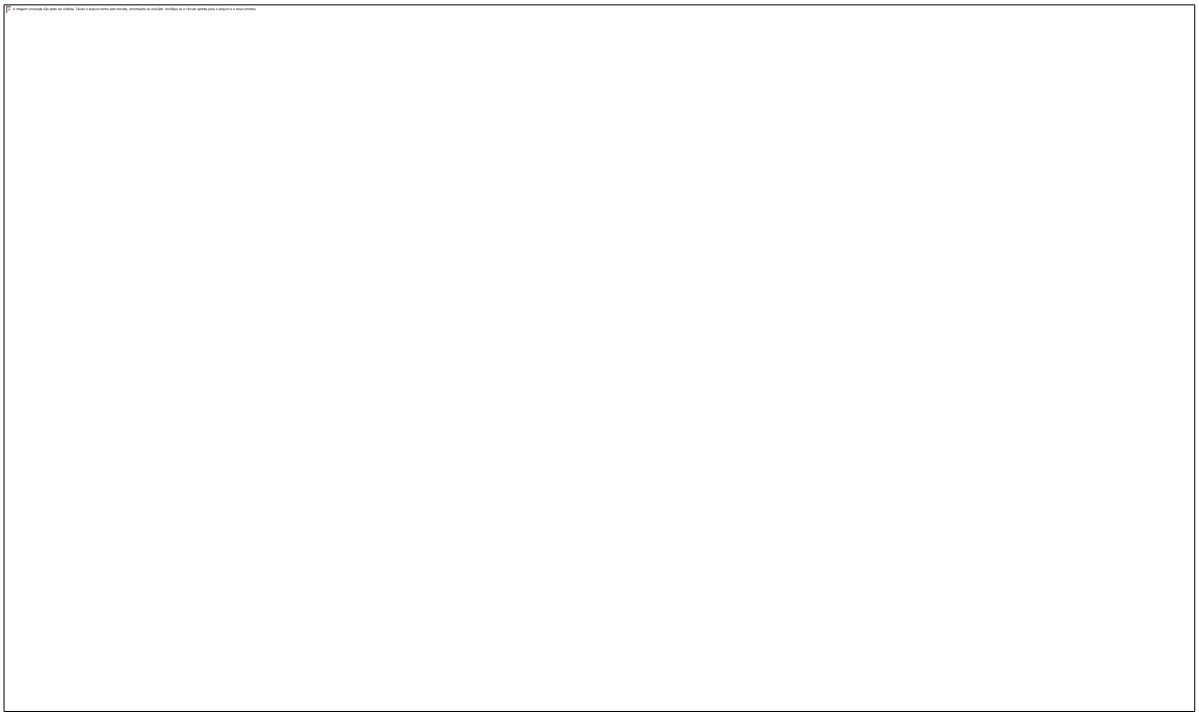


Figura 2 – Distribuição de artigos por base de dados e status da triagem



Quadro 2 – Informações básicas dos estudos selecionados

CÓDIGO: HS0759

AUTOR: DÊNIS ROCHA DA SILVA

ORIENTADOR: DENNYS LEITE MAIA

TÍTULO: Planejamento e implementação de desafios introdutórios no Objeto de Aprendizagem "Planning and implementation of introductory challenges in the Learning Object

Resumo

Este artigo descreve a concepção e desenvolvimento de desafios introdutórios do Objeto de Aprendizagem (OA) “A Saga do Caju e do Código”, voltado ao ensino do pensamento algébrico e pensamento computacional nos anos iniciais do Ensino Fundamental. O trabalho descreve a integração de dinâmicas interativas de experimentação prática que antecedem as questões avaliativas formais, com o objetivo de promover a aprendizagem ativa e reduzir a resolução por mera tentativa e erro. Para a viabilização do recurso, utilizou-se Inteligência Artificial Generativa para a co-criação do roteiro e dos elementos visuais ambientados no Cajueiro de Pirangi, além do motor de jogos GDevelop para a implementação técnica. São apresentados os desafios projetados para as três fases do OA, analisando como a manipulação física e lógica de elementos (como classificação, ordenação e traçado de caminhos) introduz conceitos de padrões e algoritmos de forma intuitiva. Conclui-se que o design intencional de dinâmicas prévias de experimentação consolida uma alternativa viável para estruturar ferramentas educacionais alinhadas à perspectiva construcionista, sem abdicar da valorização cultural local.

Palavras-chave: Objeto de Aprendizagem, Pensamento Computacional, Matemática

TITLE: Planning and implementation of introductory challenges in the Learning Object "A Saga do Caju e do Código".

Abstract

This article describes the design and development of introductory challenges for the Learning Object (LO)"A Saga do Caju e do Código", aimed at teaching algebraic and computational thinking in the early years of Elementary School. The work details the integration of interactive, hands-on experimentation dynamics that precede formal assessment questions, with the goal of promoting active learning and reducing problem-solving by mere trial and error. To enable the resource, Generative Artificial Intelligence was utilized as a co-creation tool for both the script and the visual elements set in the context of the Pirangi Cashew Tree, alongside the GDevelop game engine for technical implementation. The challenges designed for the three phases of the LO are presented, analyzing how the physical and logical manipulation of elements (such as classification, ordering, and path tracing) intuitively introduces concepts of patterns and algorithms. It is concluded that the intentional design of prior experimentation dynamics establishes a viable alternative for structuring educational tools aligned with a constructionist perspective, while preserving local cultural representation.

Keywords: Learning Objects, Computational Thinking, Mathematics

Introdução

O advento da Base Nacional Comum Curricular (BNCC) Computação trouxe uma demanda a mais para as escolas brasileiras, ensinar Computação. Entretanto, há no país um enorme déficit quanto a licenciados em Informática, como apontado por Santos (2025), o que torna inviável o estabelecimento da disciplina de Computação em uma vasta porção das escolas da rede pública. Para sanar isso, uma das formas é implementar esses conhecimentos de forma transversal, interdisciplinar com outras matérias já consolidadas no currículo regular, como Matemática, Ciências e entre outras, o que é corroborado por Oliveira e Maia (2026).

Adicionalmente, a ascensão da Inteligência Artificial Generativa (IAGen) possibilita ampliar horizontes no âmbito da educação. Silva et al. (2025) mostram o resultado de oficinas que formavam professores para criar objetos de aprendizagem (OAs) voltados à Matemática, usando essa ferramenta para contornar o obstáculo da proficiência em programação. Rebouças, Maia e Scaico (2021) definem OA como um recurso digital, acessível pela internet e cujo uso pode ser adequado ou adaptado para promover o ensino de um tema específico.

Sob essa ótica, foi pensado o OA intitulado “A Saga do Caju e do Código” reunindo desafios e perguntas voltadas para o desenvolvimento do pensamento algébrico, operações aritméticas, aliadas interdisciplinarmente com o pensamento computacional, que é um dos três eixos estipulados pela BNCC Computação.

Adicionalmente, buscando tornar a experiência discente engajadora também no âmbito cultural, visando proporcionar um senso de pertencimento (Castro; Almeida, 2026), foi inserido como cenário do OA o Cajueiro de Pirangi, que é um notório ponto turístico situado no município de Parnamirim-RN. Cabe ressaltar que a concepção inicial desse recurso foi feita por duas doutoras da área de Educação Matemática, e pretendem

aplicá-lo nas suas pesquisas com estudantes dos anos iniciais do Ensino Fundamental (EF) do citado município potiguar.

O roteiro desenvolvido compreende quatro personagens na história, as professoras Luna e Laura, além do orientador Eco e o programador Byte. As professoras têm a tarefa de dialogar e apresentar a maior parte das perguntas ao usuário, enquanto que o Orientador é responsável pelas mensagens de reforço positivo, em caso de acerto, e o programador trás dicas em caso de erro. Em suma, o OA compreende um fluxo de perguntas, com opções de resposta, e um sistema de Eco-pontos que acumula de acordo com o êxito do discente.

Entretanto, esse formato de quiz, não é algo inovador, e cai em padrões típicos de softwareseducacionais (SE) instrucionistas, visto que recai em mecânicas repetitivas e que só engajam o estudante momentaneamente (Almeida, 2000). Surgiu então o desafio de tentar deixá-lo mais alinhado à abordagem Construcionista (Papert, 2008), o que pode ser alcançado ao permitir o desenvolvimento do conhecimento por meio das ações dos estudantes, sendo elas físicas ou mentais (Almeida, 2000).

Para tal, buscou-se inserir pequenos desafios no diálogo inicial, proporcionando interações que reúnem elementos dos pilares do pensamento computacional, especialmente o reconhecimento de padrões e algoritmos, explicitados por Brackmann (2017), antes de efetuar a sistematização desses conceitos aliados com matemática, através das questões previstas originalmente. Além disso, Wing (2006) defende que o pensamento computacional está entrelaçado ao pensamento matemático, uma vez que o primeiro deriva do segundo. Logo, essas duas áreas do conhecimento se complementam.

Metodologia

Esse trabalho visa descrever o processo de inclusão de elementos construcionistas no OA “A Saga do Caju e do Código”, sem causar prejuízos ao contexto inicial, centrado no Cajueiro. Visando cumprir tal objetivo, a metodologia foi estruturada em três etapas: concepção pedagógica e curricular, prototipação e co-criação mediada por IAGen, e implementação técnica das mecânicas ativas de aprendizagem.

A primeira etapa consistiu em utilizar a curadoria e mapeamento das habilidades da Base Nacional Comum Curricular (BNCC) de Matemática (unidade temática Álgebra) e de Computação para os anos iniciais do Ensino Fundamental (1º ao 5º ano), que já estavam previstas no roteiro original, mas buscando amenizar o aspecto instrucionista e repetitivo de quizzes tradicionais. Com foco nisso, adotou-se a abordagem Construcionista, por meio da construção de dinâmicas interativas que permitam a experiência desses conceitos, sem apresentá-los formalmente, e utilizando elementos pertinentes ao cenário, como cajus, castanhas, e o próprio cajueiro.

A segunda etapa envolveu o design instrucional do roteiro introdutório, feito com auxílio da ferramenta ChatGPT. Em que foram elaborados prompts detalhados para estruturar a narrativa, além de encaminhar como anexo os documentos dos storyboards e do roteiro das perguntas, elaborados pela equipe de concepção do OA. No recurso, cada uma das fases recebe um roteiro de diálogos próprio, contendo a apresentação dos personagens e destacando o cenário do jogo, e, posteriormente é feito o encaminhamento para o desafio, contextualizando a situação-problema em que o usuário-aprendiz deverá resolver. Adicionalmente, houve a preocupação em aumentar a dificuldade dos desafios progressivamente, pensando na elevação da faixa etária escolar nas duas últimas fases.

A terceira etapa compreendeu a materialização técnica das dinâmicas lúdicas por meio da plataforma de criação de jogos GDevelop, contornando a necessidade de escrita de código complexo. Nesse estágio, foram programadas as interações essenciais descritas nas telas operacionais:

- Fase 1 (1º e 2º anos do EF): Desenvolvimento do sistema de colheita por arraste e solte (drag-and-drop) de cajus vermelhos e amarelos para seus cestos correspondentes;
- Fase 2 (3º e 4º anos do EF): Implementação da ordenação lógica de caixas de castanhas dispostas irregularmente;
- Fase 3 (4º e 5º anos do EF): Construção da mecânica de traçado vetorial de caminhos, na qual o estudante desenha a rota que os nutrientes percorrem do tronco até as folhas.

Resultados e Discussões

Cabe ressaltar, que a criação desses desafios ocorreu em paralelo ao desenvolvimento do “A Saga do Caju e do Código” e a inspiração para eles foi o próprio roteiro do OA, seguindo a sequência de fases em que o recurso educacional digital está estruturado. Os parágrafos seguintes detalham sobre quais perguntas nortearam os desafios propostos.

Para a fase 1, usou-se como referência a proposta da primeira pergunta, que pede que o estudante aponte qual o critério que está sendo usado para separar os cajus, que seria por cor. Nesse sentido, foi pensado uma dinâmica de coleta de cajus, em que há os frutos amarelos e vermelhos dispostos e é preciso arrastá-los para a caixa de cor correspondente, o que estimula o usuário a aplicar o reconhecimento de padrões, a partir do critério cor, para resolver o desafio, e, posteriormente, testa com uma pergunta se ele de fato compreendeu.

Posteriormente, para a fase 2, seguiu-se a linha do primeiro quesito, em que pedia pro usuário prever quantas castanhas haveria em uma sequência de caixas, em que a quantidade de crescia sempre quatro em comparação a anterior. Sob essa ótica, a proposta foi de apresentar as caixas com castanhas alinhadas irregularmente e pedir que o usuário-aprendiz as ordene de forma crescente, desde uma caixa vazia, passando por caixas com 4, 8, 12, até chegar em 16 unidades. Essa proposta é um pouco mais robusta,

pois elenca elementos de algoritmo como ordenação, junto ao reconhecimento do padrão de crescimento.

Por fim, para a fase 3, tomou-se como referência a questão 6, que traça semelhanças entre um cajueiro e uma estrutura de redes, em que um nutriente que vem do tronco da árvore precisa caminhar pelos galhos adentrando em cada caju, e chegando finalmente nas folhas. Em consonância a isso, elaborou-se um desafio de dois níveis para desenhar o traçado do galho, desde o tronco, passando pelos cajus visíveis, até chegar nas folhas. No primeiro nível o desenho do galho é linear, enquanto que no segundo o galho tem uma bifurcação e é preciso desenhar os dois caminhos possíveis, requisitando dos discentes que façam reconhecimento de padrões para enxergar o caminho, a princípio, invisível, e também que demarquem o caminho correto do galho, o que é um algoritmo.

Em suma, as três propostas aliam-se com as bases construcionistas discutidas por Valente (1999), a qual as interações viabilizam o ciclo de descrição-execução-reflexão-depuração de forma pedagógica e sem um feedback hostil. Assim, o erro, despido de seu caráter de fracasso ou de "resposta incorreta" sumária, é transformado em um dado para reflexão e adequação de hipóteses pelo próprio discente, reduzindo a dependência da mera tentativa e erro às cegas. Outrossim, o quadro 1 detalha melhor as questões usadas para nortear os desafios, elencando o número das questões, os diálogos de apoio, às perguntas e as respostas corretas para cada um deles.

As figuras de 1 a 8 contemplam as telas de cada desafio em estágios diferentes, as imagens à esquerda expõem sempre o desafio em seu estado inicial, ou seja, como ele é apresentado ao usuário-aprendiz, enquanto que as da direita exibe-os após algumas manipulações, de maneira a exemplificar o funcionamento.

Conclusão

O desenvolvimento deste trabalho permitiu constatar que a incorporação de mini-desafios práticos que precedem as perguntas formais constitui uma alternativa de design essencial para superar o modelo meramente instrucionista e repetitivo dos tradicionais softwares educativos em formato de quiz. À luz do construcionismo de Papert, a mecânica proposta ressignifica a tela do jogo ao transformá-la em um espaço de manipulação de "objetos para pensar".

Elementos virtuais interativos, como os cajus móveis (Figuras 1 e 2), os cestos de castanhas a serem ordenados (Figuras 3 e 4) e as trilhas vetoriais de galhos a serem desenhadas (Figuras 5 a 8), deixam de ser meras ilustrações para funcionarem como representações externas dos próprios objetos de conhecimento da Álgebra e da Computação. Em vez de expor o estudante a um ambiente de recepção passiva de teorias prontas, o recurso pedagógico convida-o à manipulação ativa de conceitos lógicos e matemáticos, propiciando que a aprendizagem abstrata ocorra a partir de ações sobre o ambiente digital.

Paralelamente, o projeto evidenciou o valor da IAGen como parceira no processo de co-criação didática de roteiros, artes e elementos midiáticos, pavimentando o caminho para que educadores e pesquisadores superem barreiras técnicas de design e programação, podendo construir seus próprios recursos autorais, incorporando elementos culturalmente relevantes.

Conclui-se que "A Saga do Caju e do Código" demonstra que é possível articular as habilidades de Álgebra e Computação da BNCC, numa perspectiva Construcionista, de forma a valorizar a identidade cultural potiguar, promovendo um cenário de aprendizagem significativa, ativa e proporcionando ao discente o lugar que melhor lhe cabe no processo de aprendizagem: o de protagonista.

Referências

ALMEIDA, Maria Elizabeth Bianconcini. Informática e formação de professores. SEED/MEC, 2000.

BRACKMANN, Christian Puhlmann. Desenvolvimento do pensamento computacional através de atividades desplugadas na educação básica. 2017.

BRASIL. Ministério da Educação. Base Nacional Comum Curricular Computação (Complemento à BNCC). Brasília: MEC. 2022.

CASTRO, J. B. de; ALMEIDA, L. C. de. Equidade e justiça social nos processos de ensino e aprendizagem da matemática. In: SANTANA, E. R. dos S.; CASTRO, J. B. de; ALMEIDA, L. C. de (org.). Equidade na Educação Matemática. Curitiba: CRV, 2026. p. 39-68. (Coleção Teoria e Prática na Educação Matemática, v. 1)

OLIVEIRA, Hery TF; MAIA, Dennys Leite. Digital Culture in the BNCC: A Transversal Mapping of Skills for Middle School. In: Society for Information Technology & Teacher Education International Conference. Association for the Advancement of Computing in Education (AACE), 2026. p. 1732-1741.

PAPERT, S. (2008). A Máquina das Crianças: repensando a Escola na Era da

Informática. Porto Alegre: Artes Médicas.

REBOUÇAS, Ayla Dantas; MAIA, Dennys Leite; SCAICO, Pasqueline Dantas. Objetos de Aprendizagem: da definição ao desenvolvimento, passando pela sala de aula. Informática na Educação: ambientes de aprendizagem, objetos de aprendizagem e empreendedorismo. Porto Alegre: Sociedade Brasileira de Computação, 2021.

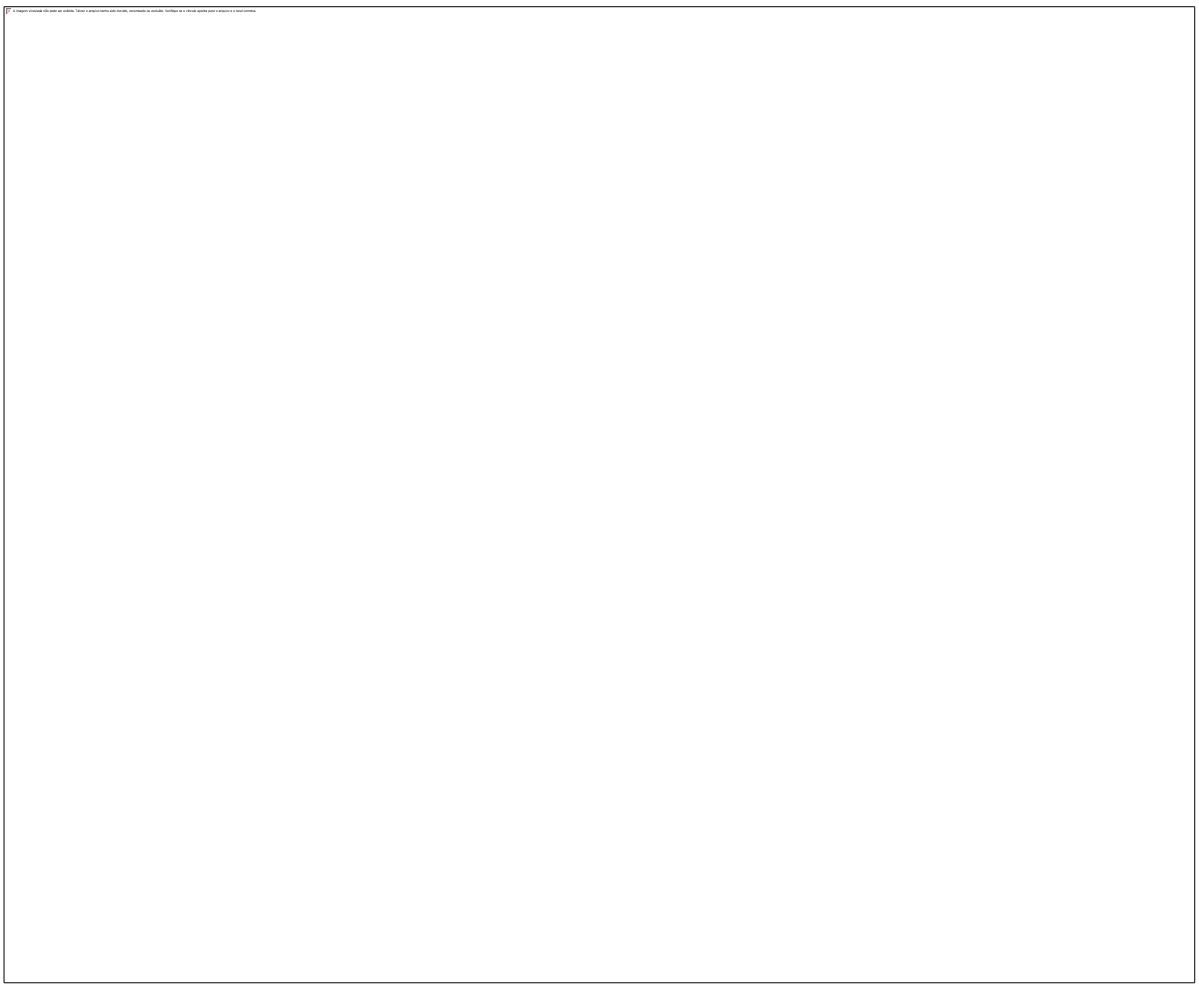
SANTOS, R. Educação Digital No Brasil: Conflitos E Contradições Da Bncc Computação. Revista Políticas Públicas & Cidades, v. 14, n. 6, p. e2406-e2406, 2025.

SILVA, Dênis R.; SILVA, Ana Carolina C.; SILVA, Igor Rauan S.; MAIA, Dennys L.; OLIVEIRA, Amanda Maria D.. Uso de Inteligência Artificial Generativa para criação de Objetos de Aprendizagem para a Matemática. In: CONGRESSO SOBRE TECNOLOGIAS NA EDUCAÇÃO (CTRL+E), 9. , 2025, Fortaleza/CE. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, 2025. p. 357-367.

VALENTE, Jose. O computador na sociedade do conhecimento. Campinas: Nied, 1999.

WING, Jeannette M. Pensamento Computacional. Communications of the ACM, v. 49, n. 3, p. 33-35, 2006. (Tradução de Ana Figueiredo, Revista de Educação Matemática da APM, 2022)

Anexos



Quadro 1 - Reúne os dados relativos às questões usadas como inspiração para os desafios.

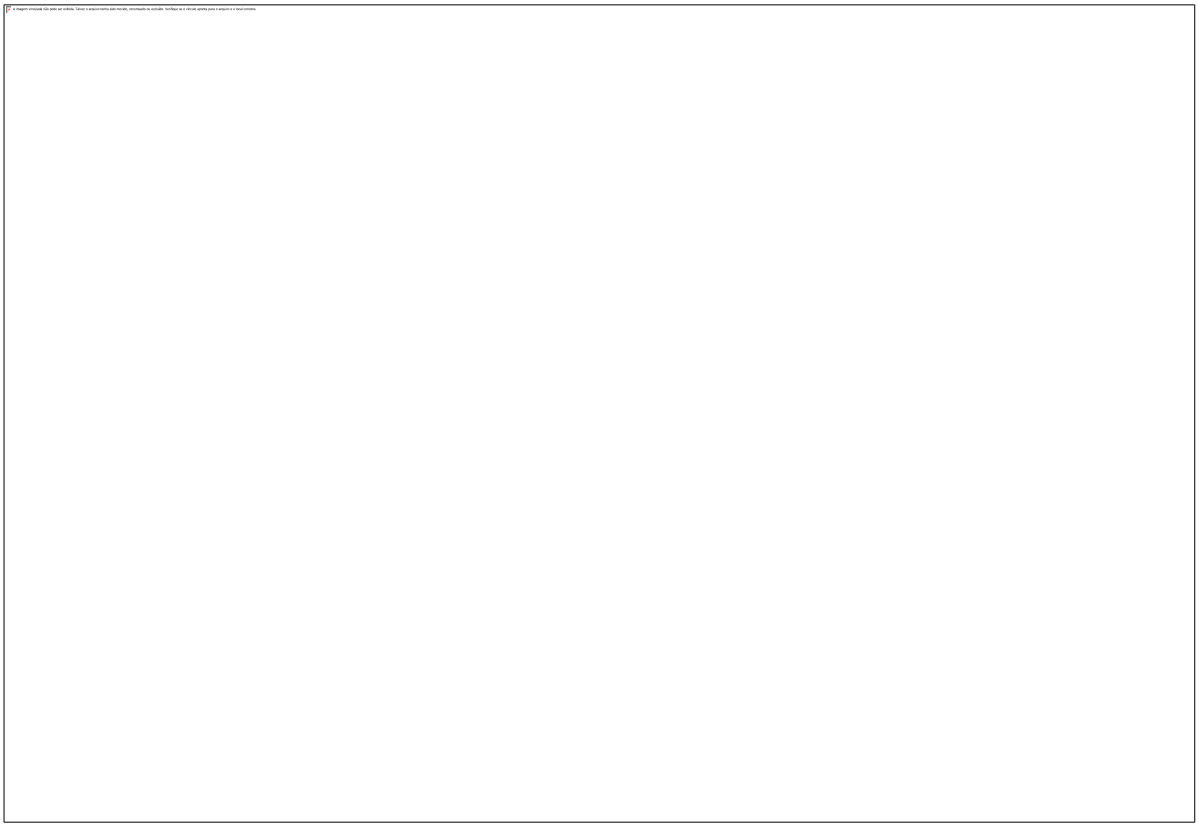


Figura 1 - Desafio 1

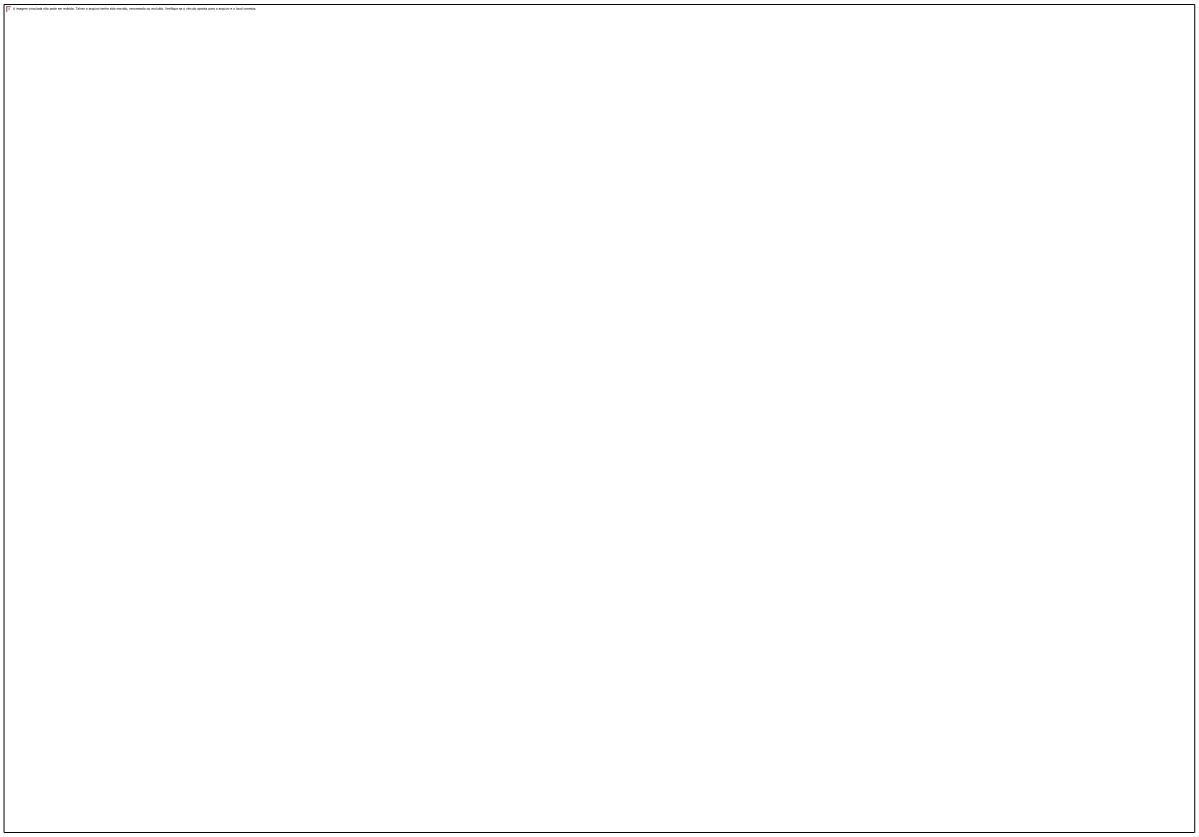


Figura 2 - Desafio 1 com alguns cajus colocados nas caixas

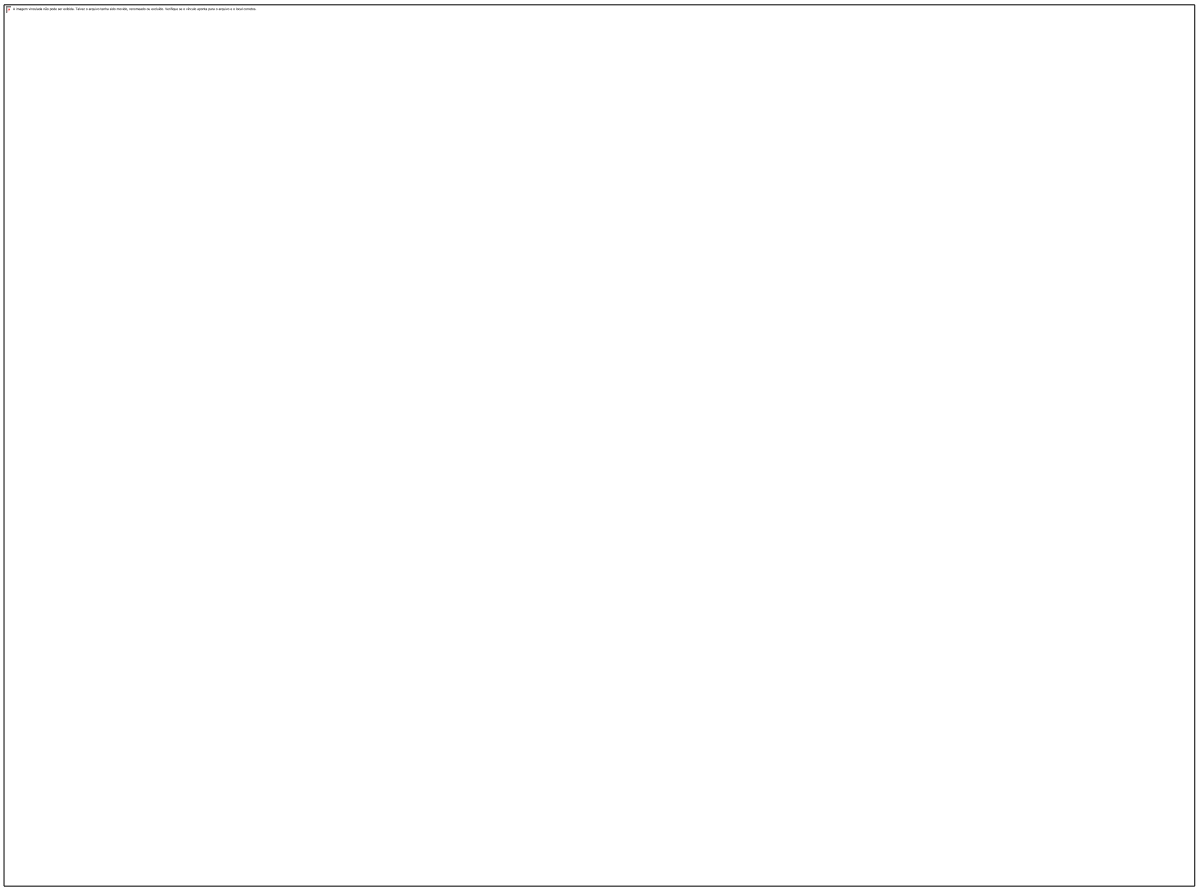


Figura 3 - Desafio 2

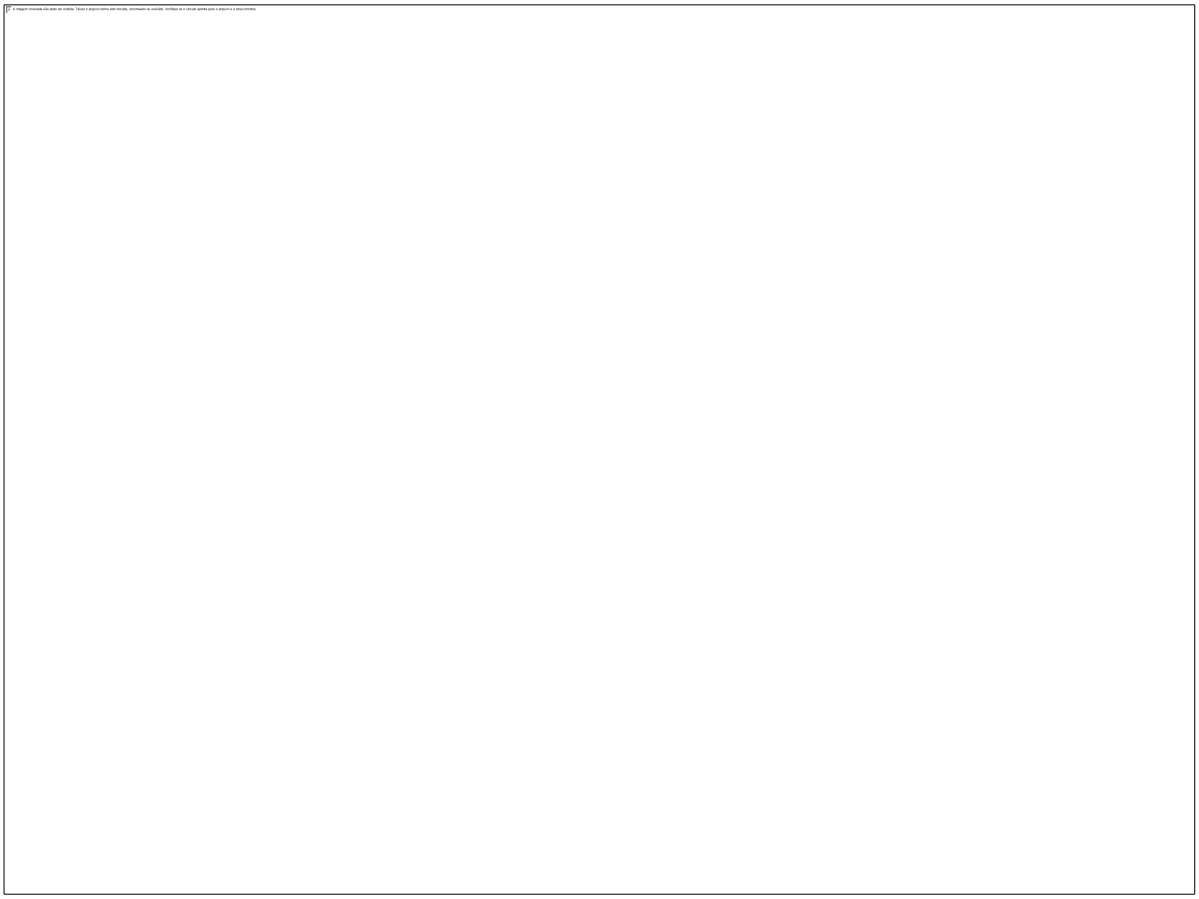


Figura 4 - Desafio 2 com algumas caixas ordenadas.



Figura 5 - Primeiro nível do desafio 3

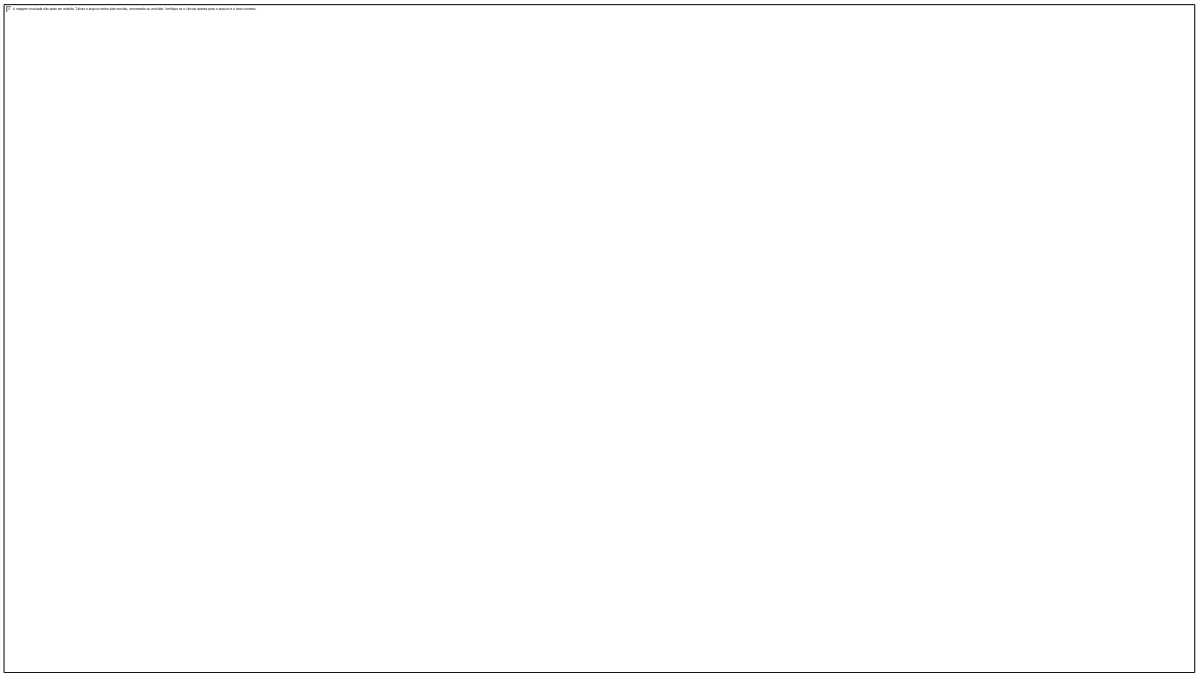


Figura 6 - Primeiro nível completo

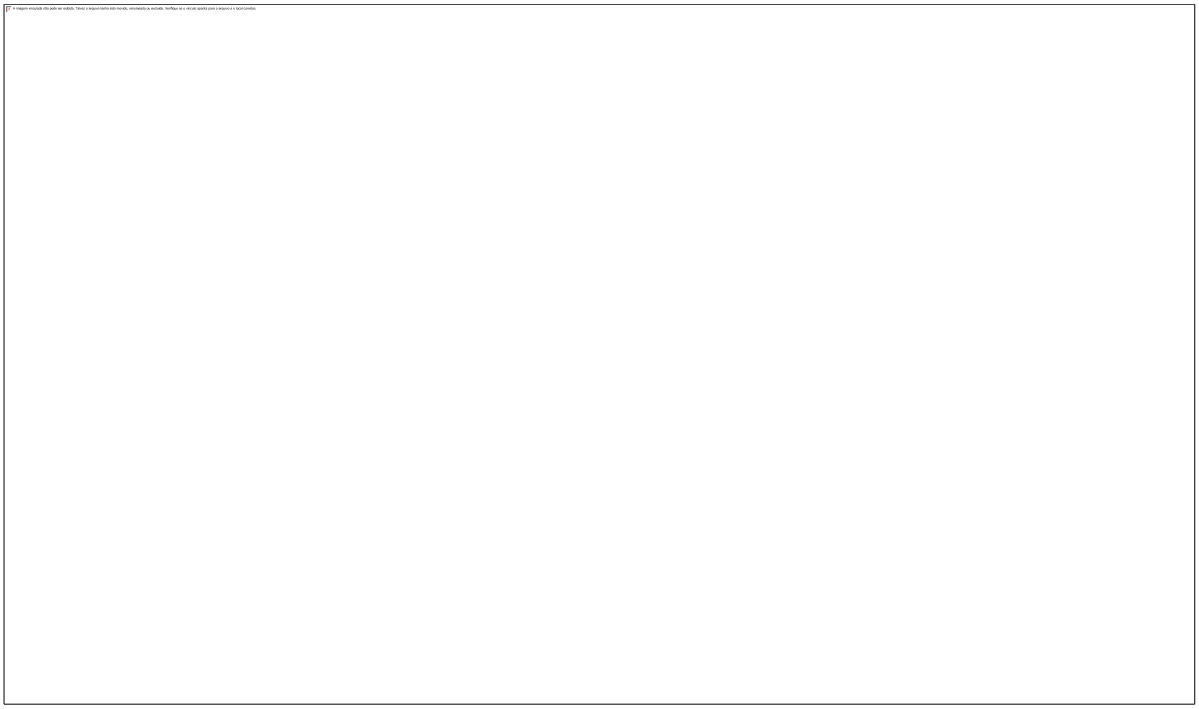


Figura 7 - Segundo nível do desafio 3

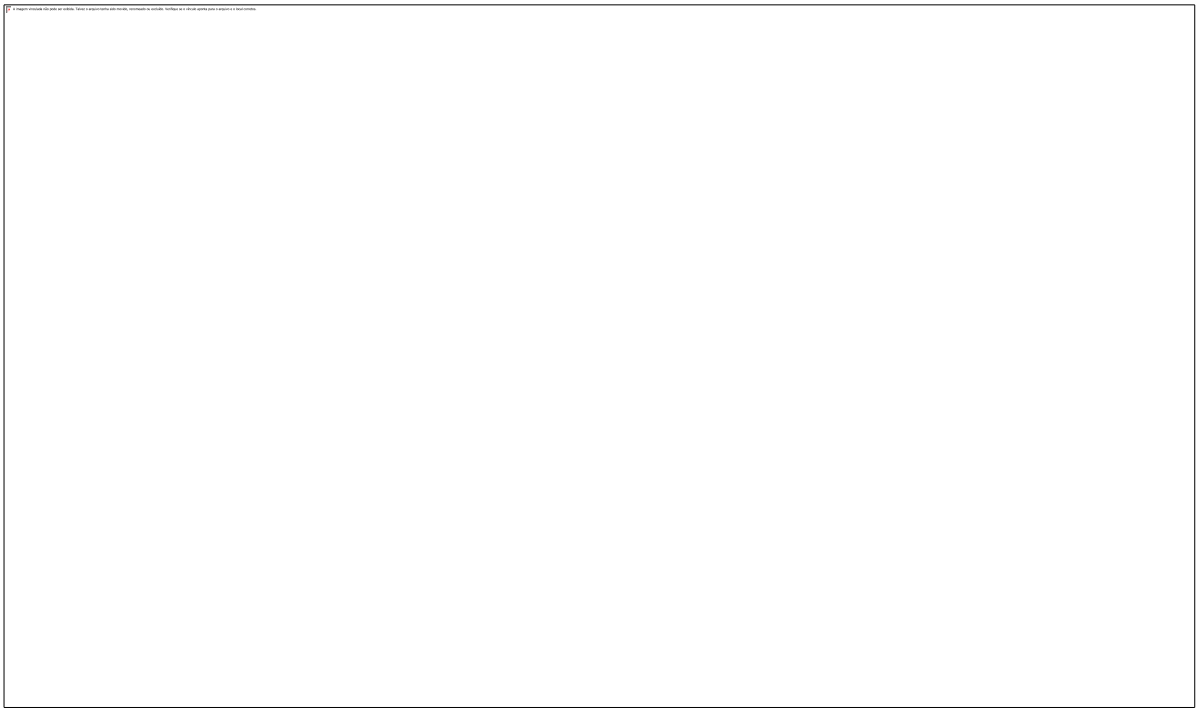


Figura 8 - Segundo nível do desafio 3 completo

CÓDIGO: HS0786

AUTOR: SAMUEL ANDERSON MACHADO LOPES

ORIENTADOR: DENNYS LEITE MAIA

TÍTULO: Fortalecendo a Educação STEAM no RN: Mapeamento de Plataformas Digitais e o Impacto da Rede STEAM Potiguar

Resumo

Este trabalho trata do mapeamento de ambientes e recursos educacionais digitais relacionados à abordagem STEAM. O estudo surge da necessidade de mapear soluções tecnológicas e pedagógicas para apoio à docência, buscando fortalecer a Educação STEAM no Rio Grande do Norte, e tem como objetivo identificar iniciativas voltadas à disseminação da abordagem entre professores da Educação Básica e articulá-las ao Observatório Rede STEAM Potiguar. Realizou-se uma pesquisa na internet de plataformas digitais, analisando suas principais características, modalidades e tipos de materiais disponíveis, com dados organizados em planilha eletrônica. Foram identificadas seis plataformas (AVAMEC, APICE FEBRACE, Aprenda Mais, Moodle – IFRS, Escola Virtual Gov e IFES – MOOC), com predominância de cursos 100% online e assíncronos. Como principais achados, constatou-se uma concentração de recursos em plataformas como AVAMEC e IFES – MOOC, que oferecem múltiplos cursos e trilhas voltadas à robótica, cultura maker e metodologias ativas, enquanto outras plataformas apresentaram ausência de catálogos específicos. Esses achados oportunizaram conhecer um panorama da oferta formativa digital e evidenciaram a dispersão de conteúdos na web. Por fim, conclui-se que o Observatório Rede STEAM Potiguar se consolida como uma solução estratégica essencial, atuando como um repositório e mapa interativo capaz de sanar essa dispersão, expandindo e qualificando a abordagem STEAM no RN.

Palavras-chave: Abordagem STEAM; Recursos educacionais digitais; Formação docente;

TITLE: Strengthening STEAM Education in RN: Mapping Digital Platforms and the Impact of the Potiguar STEAM Network

Abstract

This work addresses the mapping of digital educational environments and resources related to the STEAM approach. The study stems from the need to map technological and pedagogical solutions to support teaching—aiming to strengthen STEAM education

in Rio Grande do Norte—and seeks to identify initiatives focused on disseminating this approach among K-12 teachers and linking them to the *Observatório Rede STEAM Potiguar* (Potiguar STEAM Network Observatory). An online survey of digital platforms was conducted, analyzing their key features, formats, and types of available materials, with data organized in a spreadsheet. Six platforms were identified (AVAMEC, APICE FEBRACE, Aprenda Mais, Moodle – IFRS, Escola Virtual Gov, and IFES – MOOC), with a predominance of 100% online, asynchronous courses. Key findings revealed a concentration of resources on platforms such as AVAMEC and IFES – MOOC, which offer multiple courses and learning paths focused on robotics, maker culture, and active methodologies, whereas other platforms lacked specific catalogs. These findings provided an overview of the digital training landscape and highlighted the fragmentation of content across the web. Finally, the study concludes that the *Observatório Rede STEAM Potiguar* stands as an essential strategic solution, acting as a repository and interactive map capable of addressing this fragmentation, thereby expanding and enhancing the quality of the STEAM approach in Rio Grande do Norte.

Keywords: STEAM approach; Digital educational resources; Teacher training;

Introdução

Plataformas educacionais digitais são ambientes disponibilizados na web e voltados para a aprendizagem de conteúdos específicos e a formação docente. Esses ambientes podem favorecer o acesso a recursos educacionais, o trabalho colaborativo e o planejamento pedagógico, contribuindo para práticas docentes mais dinâmicas e para o desenvolvimento profissional dos professores Oliveira et al. (2018). A utilização desses ambientes no ensino se configura como uma estratégia capaz de articular competências técnicas e conhecimentos pedagógicos, possibilitando experiências educativas mais dinâmicas e significativas no ambiente educacional.

Nesse sentido, abre-se uma oportunidade profícua para divulgação e disseminação de práticas investigativas e criativas na Educação Básica, como a abordagem STEAM. Como destacam Maia et al. (2024), a adoção dessa abordagem pode contribuir para a diversificação das práticas pedagógicas ao integrar diferentes áreas do conhecimento de forma inter e transdisciplinar.

De acordo com Mendes et al. (2025), a avaliação e o mapeamento da abordagem STEAM na educação básica observam aspectos de fundamentação pedagógica e pertinência do conteúdo, garantindo que as plataformas não apenas funcionem de maneira eficiente, mas também esteja alinhado a objetivos educacionais claros. A avaliação exploratória, acompanhada da análise em plataformas reais, permite compreender em que medida os recursos favorecem a aprendizagem significativa, além de identificar fragilidades no acesso a esses materiais.

Nesse ínterim se insere o presente artigo, que tem como objetivo identificar iniciativas voltadas à disseminação da abordagem STEAM entre professores da Educação Básica, mapeando ambientes e recursos educacionais digitais e articulando-os ao Observatório Rede STEAM Potiguar (Cabral-Filho et al., 2025).

Metodologia

A motivação para a produção deste mapeamento decorreu da necessidade de identificar recursos educacionais voltados para a abordagem STEAM e da constatação de que o Rio Grande do Norte carecia de estratégias para disseminar essas práticas (Maia et al, 2024). Nesse contexto, a equipe de desenvolvimento e pesquisa propôs um levantamento focado em plataformas de formação. A relação entre a inovação pedagógica e a tecnologia encontra nos cursos virtuais um espaço em que ambas se complementam.

O levantamento foi desenvolvido utilizando planilhas eletrônicas para organizar os dados e pesquisas diretas nas plataformas governamentais e institucionais. A busca foi realizada mediante a exploração individual de cada plataforma, pesquisando áreas e conteúdos relacionados à educação STEAM, formação docente, aprendizagem investigativa e criativa, resolução de problemas e interdisciplinaridade. Foram consideradas plataformas abertas e gratuitas, com cursos ou recursos voltados a professores da Educação Básica ou ao público em geral. Como parte do processo de desenvolvimento e avaliação, foi elaborada uma matriz de análise com categorias como: Instituição, Público, Certificado, Modalidade, Recursos e Ano de criação. Esse procedimento segue recomendações de Marconi e Lakatos (1992) sobre a utilização de pesquisa documental e formulários de catalogação, possibilitando a coleta de informações detalhadas sobre as plataformas educacionais.

A amostra foi selecionada de forma intencional, considerando o alcance nacional e regional das plataformas. Os dados coletados foram analisados por meio de estatística descritiva básica, utilizando planilhas eletrônicas para compilar, organizar e comparar os recursos encontrados.

Resultados e Discussões

O mapeamento permitiu a análise de uma amostra de seis plataformas educacionais, possibilitando uma visão inicial da disponibilidade de recursos digitais voltados à formação e ao apoio à prática docente, conforme sintetizado na Quadro 1. Há uma distribuição de público, em que plataformas como a AVAMEC e APICE FEBRACE focam fortemente em Professores e Estudantes da Educação Básica, enquanto o Aprenda Mais, Moodle - IFRS, Escola Virtual Gov e IFES - MOOC apresentam também cursos destinados ao público em geral e servidores públicos. Conforme evidenciado na mesma tabela, a modalidade predominante em todas elas foi 100% online (assíncrona), com emissão de certificados e uso de recursos como videoaulas,

fóruns, materiais de leitura e questionários, o que demonstra uma estruturação voltada para a flexibilidade do professor.

Em relação à compreensão da oferta do conteúdo acerca da abordagem STEAM, os dados detalhados na Tabela 2 revelaram que a plataforma AVAMEC (Brasil, 2019) possui em seu catálogo um leque de opções formativas acerca da temática (incluindo cursos como Práticas Pedagógicas Inovadoras com Abordagem STEAM e STEAM integrados à BNCC). A plataforma APICE FEBRACE (2013) também se destacou com conteúdos específicos voltados para a área, assim como o IFES - MOOC, que apresentou uma forte vertente voltada à Cultura maker e Robótica, temáticas vinculadas à abordagem STEAM (Maia, Carvalho e Appelt, 2021). Por outro lado, nas plataformas Moodle - IFRS (2017), Escola Virtual Gov e Aprenda Mais (2021), não foram encontrados cursos específicos sobre a abordagem STEAM no período analisado, sendo os resultados classificados como "NOT FOUND" (não encontrado). Ainda assim, essas plataformas foram consideradas na pesquisa por serem ambientes públicos e institucionais de formação que oferecem cursos com certificação, permitindo identificar possíveis lacunas na oferta de conteúdos relacionados ao STEAM.

A análise dos dados revelou que a oferta de materiais relacionados à abordagem STEAM está concentrada em quatro plataformas, enquanto duas não registraram cursos específicos sobre a temática. Conforme ilustra o Gráfico 1, a plataforma AVAMEC lidera a oferta com seis recursos específicos voltados para a abordagem STEAM e metodologias ativas voltadas para o desenvolvimento de práticas investigativas e criativas. Em seguida, destacam-se o IFES - MOOC, com cinco cursos voltados à Cultura Maker e Robótica, e a APICE FEBRACE, que apresenta três cursos. A Escola Virtual Gov contribui com dois cursos transversais correlatos à área (focados em Design Thinking e resolução de problemas). Em contrapartida, as demais plataformas (Aprenda Mais e Moodle - IFRS) não registraram catálogos específicos direcionados a essa temática.

A avaliação subjetiva sobre a aplicabilidade desses dados revelou que o Observatório Rede STEAM Potiguar (Cabral Filho et al., 2025) será uma solução estratégica. A dinâmica do Observatório foi avaliada como essencial, pois atua como um repositório de recursos locais e um mapa interativo, sanando o problema da dispersão identificado no mapeamento.

Conclusão

O mapeamento realizado indicou que o uso de plataformas educacionais digitais contribui significativamente para o desenvolvimento profissional dos professores, ao mesmo tempo em que desperta a necessidade de organização e curadoria desses materiais. Os resultados da avaliação evidenciam que a maioria das plataformas consultadas (AVAMEC, APICE FEBRACE) disponibiliza recursos voltados à abordagem STEAM, utilizando a dinâmica da modalidade online assíncrona, reforçando a importância da integração entre diferentes instituições na criação de recursos

educacionais. Além disso, a ausência de conteúdos específicos em algumas plataformas (Moodle-IFRS, Escola Virtual Gov) evidencia o potencial de expansão dessa temática.

Apesar dos resultados positivos na identificação de cursos, a pesquisa apontou a dispersão como um obstáculo. Entre as soluções propostas, destaca-se o Observatório Rede STEAM Potiguar como uma ferramenta centralizadora. Essas constatações fornecem subsídios importantes para o refinamento de políticas de formação, garantindo que atue com

Referências

Araujo, C. L., Santos, Q. P., Ribeiro, H. M. L., Freitas, E. B., & Coutinho, D. J. G. (2025). Evasão escolar: causas e impactos da evasão escolar no Brasil e no mundo. Revista Ibero-Americana De Humanidades, Ciências E Educação, 11(1), 1945-1965.

Brasil. Ministério da Educação. (2018). Base Nacional Comum Curricular. MEC.

CABRAL FILHO, Emanuel Kywal Pinto ; OLIVEIRA, Amanda Maria Domingos de ; SILVA, Ana Carolina Costa ; SILVA, Dênis Rocha da ; MONTEIRO, Keven William Pereira ; SILVA, Lucas Toshio Nascimento da ; LOPES, Samuel Anderson Machado ; MAIA, Dennys Leite ; SOARES, Raiza de Araújo Domingos . Observatório Rede STEAM Potiguar. In: Ctrl+E 2025, 2025, Fortaleza. Anais do Ctrl+E 2025. Porto Alegre: SBC, 2025. p. 555-559.

Fullan, M., & Langworthy, M. (2014). A rich seam: How new pedagogies find deep learning. Pearson.

Lourenço, R. S. P., et al. (2025). Stop the Pigeon! A STEAM Project on socio-environmental sustainability awareness in a northeastern brazilian school. In Anais do Society for Information Technology & Teacher Education International Conference. Association for the Advancement of Computing in Education (AACE).

Maia, D., et al. (2024a). Abordagem STEAM no Rio Grande do Norte: Uma análise longitudinal a partir de experiências publicadas na FEBRACE. Revista Brasileira de Ensino de Ciência e Tecnologia (RBECT), 17, 1-20.

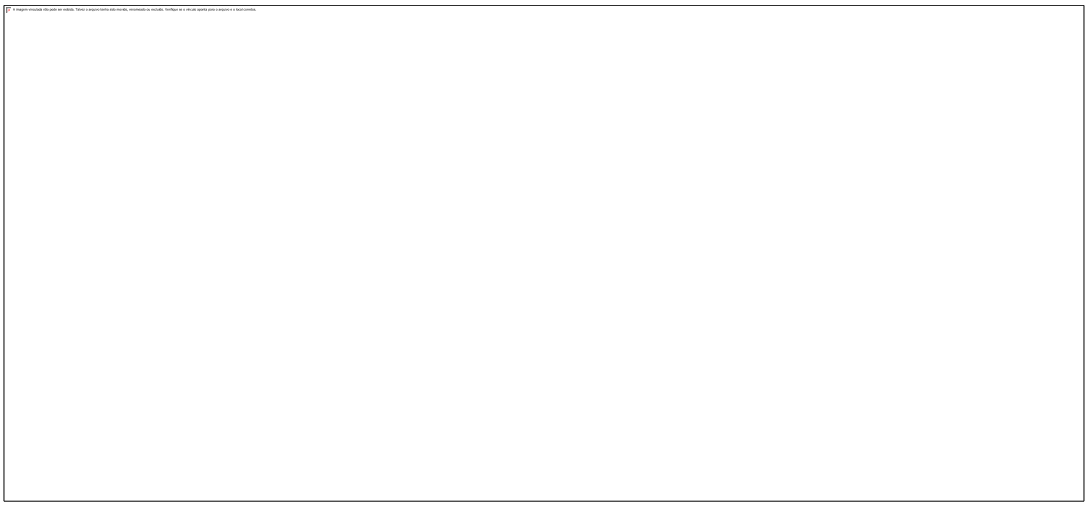
Maia, D., et al. (2024b). A abordagem STEAM como proposta pedagógica interdisciplinar para aprendizagem matemática. Revista Ensino Em Debate (REDE), 2, e2024016.

Marconi, Marina de Andrade; LAKATOS, Eva Maria. Fundamentos de metodologia científica. 7. ed. São Paulo: Atlas, 1992.

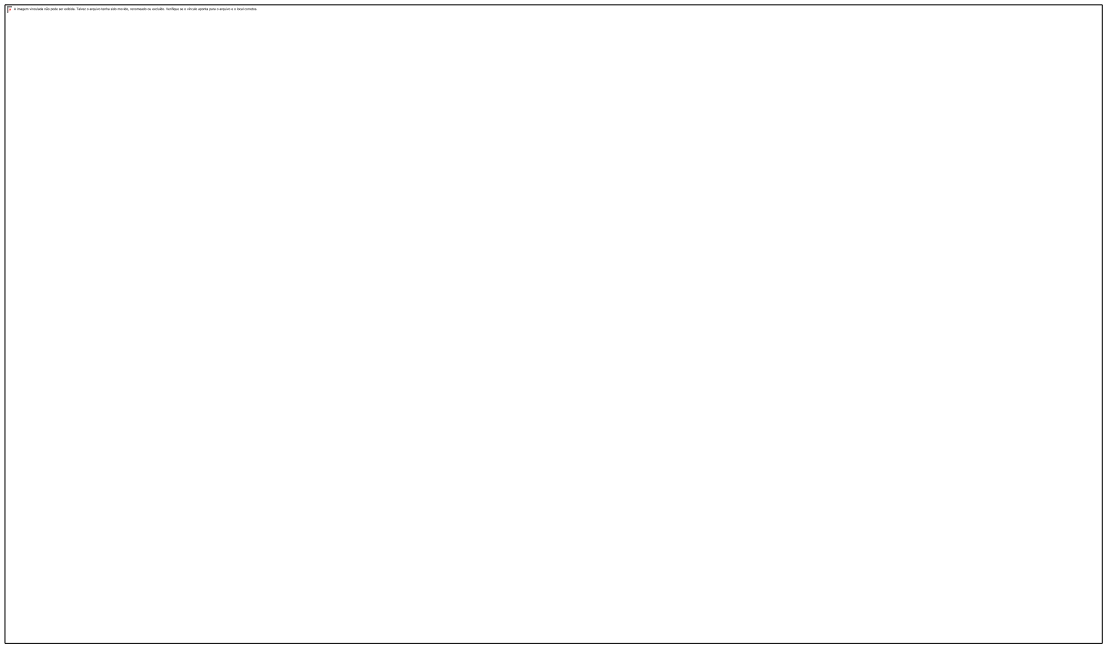
Mendes, A. A., et al. (2025). Mapeamento da abordagem STEAM na educação básica: analisando impactos e desafios atuais. Ensino de Ciências e Tecnologia em Revista - ENCITEC, 15(2), 305-319.

Oliveira, A. M. D. de, et al. (2018). OBAMA: Uma Plataforma para Inovação e Colaboração na Prática Docente. Anais dos Workshops do VII Congresso Brasileiro de Informática na Educação (WCBIE), 156–163.

Anexos



Quadro 1: Caracterização Geral das Plataformas.



Quadro 2: Oferta de Recursos STEAM.

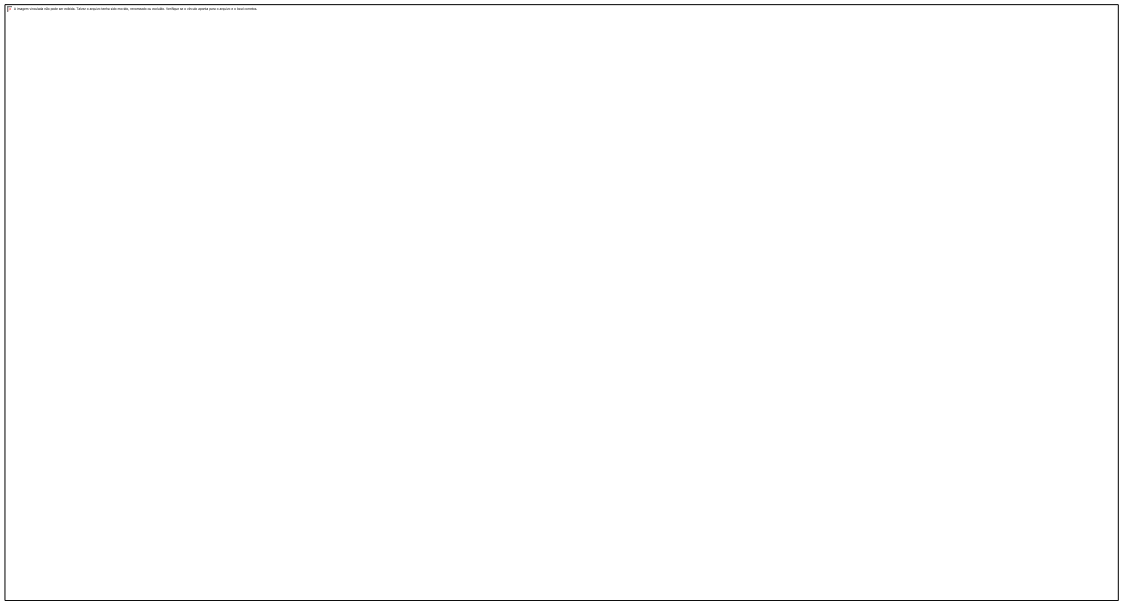


Gráfico 1: Quantidade de Recursos STEAM nas Plataformas Mapeadas

CÓDIGO: HS0791

AUTOR: RADLA OLIVEIRA ARAÚJO DOS SANTOS

COAUTOR: CESAR RENNO COSTA

ORIENTADOR: RENAN CIPRIANO MOIOLI

TÍTULO: Desenvolvimento e Análise de Software de Rastreamento Ocular no Estudo do Processamento de Expressões Idiomáticas

Resumo

O processamento de expressões idiomáticas apresenta desafios tanto para a psicolinguística quanto para o Processamento de Linguagem Natural (PLN). Falantes não nativos (L2) frequentemente adotam uma abordagem literal inicial, o que acarreta um custo cognitivo mensurável. Este trabalho foi desenvolvido em apoio à pesquisa de Eduardo Santos, vinculada à iniciativa internacional MIA, com o objetivo de construir e validar um dataset de rastreamento ocular. A amostra abrangeu estudantes universitários falantes nativos de português brasileiro, distribuídos pelos níveis de proficiência em inglês (de A1 a C2). O escopo de apoio metodológico focou na condução dos experimentos com o Tobii Pro Spark (60 Hz) e o software PsychoPy, além do auxílio no pipeline de processamento para extrair macroeventos cognitivos, como fixações e regressões, a partir das coordenadas espaciais. Os resultados indicam uma forte correlação inversa entre a proficiência linguística e o número de regressões oculares. Leitores de nível intermediário (B1) exibiram picos de regressão (média de 1,776), sinalizando alto esforço cognitivo, ao passo que usuários fluentes (C2) demonstraram eficiência e acesso lexical direto (média de 987). O dataset estruturado fornecerá embasamento cognitivo para avaliar e aprimorar modelos artificiais de compreensão figurativa.

Palavras-chave: Rastreamento Ocular; Expressões Idiomáticas; Psicolinguística; PLN

TITLE: Development and Analysis of Eye-Tracking Software for the Study of Idiomatic Expression Processing

Abstract

The processing of idiomatic expressions presents challenges for both psycholinguistics and Natural Language Processing (NLP). Non-native speakers (L2) often adopt an initial literal approach, generating a measurable cognitive cost. This work was developed to support Eduardo Santos's research, linked to the international MIA initiative, aiming to build and validate an eye-tracking dataset. The sample included university students who are native speakers of Brazilian Portuguese, distributed across English proficiency levels (A1 to C2). The methodological support focused on conducting experiments with the Tobii Pro Spark (60 Hz) and PsychoPy, as well as assisting with the processing pipeline to extract macro-cognitive events, such as fixations and regressions, from spatial coordinates. The results indicate a strong inverse correlation between language proficiency and the number of ocular regressions. Intermediate-level readers (B1) exhibited peaks in regressions (mean of 1,776), signaling high cognitive effort, while fluent users (C2) demonstrated efficiency and direct lexical access (mean of 987). The structured dataset will provide cognitive grounding to evaluate and improve artificial models in figurative comprehension.

Keywords: Eye-tracking; Idiomaticity; Psycholinguistics; NLP; Dataset

Introdução

A compreensão de expressões de múltiplas palavras (MWEs), especialmente as expressões idiomáticas, é um tópico desafiador tanto para a psicologia cognitiva quanto para o desenvolvimento de modelos de processamento de linguagem natural (PLN). Expressões como "spill the beans" contêm uma carga semântica não composicional, em que o significado global não pode ser deduzido unicamente pela soma de suas partes. Enquanto falantes nativos acessam esses significados diretamente, falantes de uma segunda língua (L2) enfrentam custos cognitivos elevados e frequentemente recorrem à tradução literal antes de recorrer à interpretação figurativa. Esta dinâmica é descrita pela Hipótese de Ressonância de Saliência Literal, que prevê um viés inicial para a decodificação "literal-primeiro" em L2.

Atualmente, os grandes modelos de linguagem (LLMs) enfrentam dificuldades semelhantes ao processar linguagem figurada, atuando como "caixas-pretas". A iniciativa internacional Modeling Idiomaticity in Human and Artificial Language Processing (MIA) e o protocolo AdMIRe (Advancing Multimodal Idiomaticity Representation) visam preencher essa lacuna integrando sinais de processamento humano aos sistemas computacionais. O presente relatório descreve as atividades de apoio prestadas ao aluno Eduardo Santos na construção e no processamento de um banco de dados de rastreamento ocular (eye tracking) para quantificar o esforço cognitivo de aprendizes de língua inglesa. A medição contínua da leitura permite isolar o acesso lexical (duração das fixações) da integração semântica tardia (sacadas regressivas), fornecendo métricas empíricas fundamentais para avaliar modelos de IA.

Metodologia

O estudo seguiu um delineamento de pares correspondentes, submetendo os participantes à leitura de frases em língua inglesa contendo Expressões Potencialmente Idiomáticas (PIEs) inseridas tanto em contextos figurativos quanto literais. O conjunto de estímulos foi adaptado de corpora como o MAGPIE. O público-alvo consistiu em estudantes universitários (IMD/UFRN), nativos de português brasileiro (L1), distribuídos por todos os níveis de proficiência do Quadro Europeu Comum de Referência para Línguas (CEFR), de A1 a C2.

A coleta e a apresentação de estímulos, conduzidas em laboratório controlado, utilizaram o software PsychoPy, de código aberto. O rastreamento visual foi realizado com o hardware Tobii Pro Spark, operando a uma taxa de amostragem de 60 Hz. Como o equipamento de 60 Hz exporta as coordenadas de olhar brutas sem heurísticas nativas de macroeventos, o apoio à pesquisa incluiu o tratamento computacional dos dados telemétricos. Um pipeline open-source foi utilizado junto à biblioteca PyGaze para mapeamento visual, aplicando o algoritmo Identification by Dispersion Threshold (I-DT) para identificação de fixações. Além disso, como parte do processamento analítico dos dados brutos, adotou-se uma heurística personalizada capaz de mapear sacadas regressivas (movimentos oculares para a esquerda no texto) diretamente a partir das sequências de fixação, contornando as limitações de ruído do equipamento. O dataset final foi organizado em uma árvore de diretórios hierárquica, contendo os dados brutos e os arquivos de métricas sumarizadas.

Resultados e Discussões

Os dados processados validam o pipeline computacional empregado e confirmam a premissa central da pesquisa: há uma correlação inversa substancial entre o nível de proficiência (CEFR) e a ocorrência de comportamentos oculares indicativos de carga cognitiva, especificamente as regressões oculares. A análise preliminar dos indicadores de regressão extraídos evidenciou que a coorte de nível B1 (intermediário) apresentou o maior nível de heterogeneidade e esforço, com uma média de 1.776 regressões por participante e um pico extremo de 4.046 regressões. Este perfil denota uma exigência cognitiva aguda, em que o leitor adota a decodificação literal, encontra falhas de coerência semântica e necessita retornar o olhar para reavaliar os termos, o que se alinha à Hipótese da Saliência Literal.

O comportamento apresenta uma transição no nível B2 (média de 1.439 regressões) e avança progressivamente até à estabilidade no grupo avançado C1 (média de 1.103 regressões). Por fim, o grupo de nível C2 demonstrou acesso lexical eficiente, registrando a menor média (987 regressões) e a menor variância. Os aprendizes de nível C2 apresentaram comportamentos oculares que sugerem leitura fluente e direta, estabelecendo um limite de estabilidade metodológica que pode ser aproveitado como "sinal de supervisão cognitiva" no treinamento de modelos computacionais avaliados pelo projeto MIA.

Conclusão

O presente trabalho de suporte à iniciação científica comprovou a viabilidade técnica de empregar hardware de rastreamento ocular de nível de entrada (60 Hz) para a captura de eventos macro-cognitivos relevantes à psicolinguística e ao PLN. O auxílio nas etapas de aquisição laboratorial e de refinamento algorítmico do pipeline de dados viabilizou a extração robusta de regressões oculares.

A estruturação deste dataset quantifica de forma transparente o atrito cognitivo característico de falantes não nativos, demonstrando que o domínio da linguagem figurada mitiga progressivamente o viés "literal-primeiro", transicionando do processamento errático e custoso (nível B1) para a estabilidade de fluência (nível C2). Com os dados disponibilizados em repositórios abertos, a infraestrutura construída se integra ao ecossistema da iniciativa MIA, viabilizando métricas cognitivas humanas concretas para ancorar o desenvolvimento e o alinhamento semântico da inteligência artificial.

Referências

CIEŚLICKA, A. Literal salience in on-line processing of idiomatic expressions by second language learners. Second Language Research, v. 22, n. 2, p. 115-144, 2006.

HAAGSMA, H.; BOS, J.; NISSIM, M. MAGPIE: A Large Corpus of Potentially Idiomatic Expressions. ACL Anthology, p. 2792-287, maio 2020.

RAYNER, K. Eye movements and attention in reading, scene perception, and visual search. Quarterly Journal of Experimental Psychology, v. 62, n. 8, p. 1457-1506, ago. 2009.

SIYANOVA-CHANTURIA, A.; CONKLIN, K.; SCHMITT, N. Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. Studies in Second Language Acquisition, 33(2):323-346, 2011.

VILLAVICENCIO, A. MIA: Modeling Idiomaticity in Human and Artificial Language Processing. UK Research and Innovation (UKRI) Grant, 2020.

CÓDIGO: HS0797

AUTOR: LUCAS TOSHIO NASCIMENTO DA SILVA

COAUTOR: TAYNARA BEATRIZ SALES DA SILVA

ORIENTADOR: DENNYS LEITE MAIA

TÍTULO: MATERIAIS DIDÁTICOS PARA DOCENTES SOBRE ABORDAGEM E ESPAÇOS STEAM

Resumo

A abordagem STEAM tem se consolidado como uma estratégia educacional transformadora, exigindo, contudo, a capacitação contínua dos professores para a utilização de laboratórios e espaços de prototipação. Este trabalho apresenta o desenvolvimento de materiais didáticos audiovisuais voltados para a formação de docentes da Educação Básica, vinculados ao projeto eSTEAMulAR. A pesquisa focou na roteirização e produção de quatro vídeos instrucionais contemplando a operação segura e pedagógica de uma furadeira de bancada, uma serra tico-tico de bancada, uma impressora 3D e uma cortadora CNC laser. A metodologia adotada envolveu a curadoria de conteúdos existentes e a apropriação técnica dos equipamentos, culminando em recursos multimidiáticos que articulam o funcionamento prático com o design instrucional. O desenvolvimento desse material visa facilitar a compreensão do manuseio das ferramentas, promovendo maior autonomia docente para a criação de sequências de ensino baseadas em problemas reais e alinhadas aos Objetivos de Desenvolvimento Sustentável.

Palavras-chave: Formação Docente. Educação STEAM. Tecnologias Educacionais.

TITLE: Teaching materials for instructors on STEAM approaches and spaces.

Abstract

The STEAM approach has consolidated itself as a transformative educational strategy, requiring, however, the continuous training of teachers for the use of laboratories and prototyping spaces. This work presents the development of audiovisual didactic materials aimed at the training of Basic Education teachers, linked to the eSTEAMulAR project. The research focused on the scripting and production of four instructional videos covering the safe and pedagogical operation of a drill press, a scroll saw, a 3D printer, and a CNC laser cutter. The adopted methodology involved the curation of existing content and the technical appropriation of the equipment, culminating in

multimedia resources that articulate practical operation with instructional design. The development of this material aims to facilitate the understanding of tool handling, promoting greater teacher autonomy for the creation of teaching sequences based on real problems and aligned with the Sustainable Development Goals.

Keywords: Teacher Training. STEAM Education. Educational Technologies.

Introdução

A educação contemporânea exige constantes adaptações para engajar os estudantes e conectá-los aos desafios do século XXI. Nesse cenário, a abordagem STEAM (Ciências, Tecnologia, Engenharia, Artes e Matemática) consolida-se como um caminho promissor para a transformação do ensino. Conforme definem Maia e Soares (2025) em sua proposta voltada ao Sul Global, a STEAM atua como uma inovação educacional, configurando-se como uma prática pedagógica investigativa e criativa, de caráter interdisciplinar e com potencial para alcançar a transdisciplinaridade, centrada na exploração de problemas do mundo real, especialmente aqueles oriundos do contexto dos estudantes.

Contudo, a efetivação dessa proposta inovadora no cenário da Educação Básica brasileira esbarra em desafios estruturais e formativos. Observa-se que as escolas frequentemente carecem de infraestrutura adequada, somando-se à urgência de formação continuada para os professores. Muitos educadores ainda desconhecem as práticas relacionadas ao STEAM ou sentem-se inseguros para incorporar equipamentos de laboratório em seus planejamentos didáticos diários. Consequentemente, a simples disponibilização de máquinas não garante a mudança pedagógica; é imprescindível que os docentes desenvolvam a autonomia e a segurança técnica necessárias para articular o uso dessas ferramentas com as metodologias ativas.

Para suprir essa lacuna, o projeto de pesquisa eSTEAMulAR propõe intervenções diretas na qualificação docente por meio de um ecossistema de formações. O Instituto Metrópole Digital concebeu o eSTEAMuLab, um ambiente equipado com diversas ferramentas para fomentar ações investigativas. Nesse espaço, são ofertadas formações imersivas que abrangem desde a fundamentação da abordagem STEAM e temas contemporâneos, como o uso de Inteligências Artificiais Generativas, até oficinas totalmente práticas focadas na operação de equipamentos, como impressoras 3D e ferramentas de marcenaria. O propósito é romper as barreiras do desconhecimento tecnológico, permitindo que os professores experimentem recursos que transitam desde a manipulação manual até a fabricação digital, que os encorajem a implementar e ampliar práticas pedagógicas com abordagem STEAM.

Desse modo, o presente trabalho relata o desenvolvimento de materiais didáticos audiovisuais criados para atuar como recurso de apoio nessas oficinas de formação e também em professores, em qualquer lugar do mundo, que desejem conhecer melhor o funcionamento dos equipamentos de espaços STEAM. O escopo da pesquisa abrangeu a roteirização e a produção de quatro vídeos com instruções básicas para o manuseio e a operação segura de uma furadeira de bancada, uma serra tico-tico de bancada, uma impressora 3D e uma cortadora CNC laser. Ao fornecer esse suporte multimidiático

prático, o objetivo é garantir que os docentes tenham uma fonte de consulta rápida e confiável para complementar as atividades presenciais. Munidos desse embasamento técnico inicial, espera-se que os educadores sintam-se preparados para criar soluções e sequências de ensino alinhadas aos Objetivos de Desenvolvimento Sustentável, consolidando efetivamente o STEAM em suas realidades.

Metodologia

A pesquisa caracteriza-se como um estudo de natureza aplicada e exploratória, estruturado em etapas interconectadas para garantir a eficácia técnica e pedagógica dos materiais desenvolvidos. A fase inicial consistiu na curadoria de manuais e conteúdos instrucionais existentes, permitindo identificar as principais lacunas na documentação técnica disponível. Observou-se que grande parte dos materiais adota uma linguagem excessivamente voltada para a Engenharia, distanciando professores da Educação Básica que não possuem formação prévia em fabricação digital ou marcenaria.

A partir desse diagnóstico, a etapa de apropriação técnica direcionou seu foco para quatro equipamentos específicos do eSTEAMuLab: a furadeira de bancada, a serra tico-tico de bancada, a impressora 3D e a cortadora CNC laser. A escolha dessas máquinas baseou-se na premissa de que, embora muitas ferramentas manuais e elétricas de um espaço maker dispensem tutoriais complexos, estes quatro equipamentos específicos costumam gerar medo e insegurança em usuários leigos. Isso inclusive foi identificado ao longo das oficinas realizadas no Esteamulab, em que muitos participantes, apesar de curiosos pelo uso dos equipamentos, não se sentiam aptos para explorá-los. A imersão prática da equipe visou mapear as etapas operacionais, os protocolos de segurança e as configurações básicas de cada máquina, com o intuito de traduzir a complexidade técnica em instruções acessíveis.

Para a estruturação pedagógica e elaboração dos roteiros, optou-se por uma abordagem direta e instrucional focada no suporte imediato, refletindo as reais necessidades do ambiente de prototipação. O planejamento concentrou-se na produção de vídeos curtos, objetivos e estritamente visuais. Para garantir essa dinamicidade e a aplicação "just-in-time" durante as oficinas presenciais, a estratégia de acesso ao material foi construída por meio de códigos QR (QR Codes). Esses códigos são elaborados para serem fixados fisicamente nas estações de trabalho, próximos a cada máquina, permitindo que o docente escaneie a etiqueta com um smartphone e seja direcionado instantaneamente para a plataforma YouTube, acessando instruções simples sobre o manuseio seguro do equipamento correspondente.

A etapa final compreendeu a produção audiovisual propriamente dita. A gravação e a edição enfatizaram o detalhamento visual das interfaces, o posicionamento correto das mãos, o uso de equipamentos de proteção individual (EPIs) e as etapas sequenciais de acionamento e parada das máquinas. A finalização dessa fase culminou com o upload dos quatro vídeos para a internet, consolidando o desenvolvimento do material de apoio e deixando a infraestrutura instrucional pronta para ser utilizada nas oficinas de capacitação do projeto.

Resultados e Discussões

O desenvolvimento prático da pesquisa culminou na produção, finalização e publicação de quatro vídeos instrucionais na plataforma YouTube, estruturando um arcabouço de suporte tecnológico fundamental para as dinâmicas do eSTEAMuLab. A principal entrega deste trabalho materializa-se na arquitetura de um sistema de acesso sob demanda à informação, concretizado pela integração de códigos QR físicos fixados diretamente nas estações de trabalho do laboratório. Essa configuração espacial e digital assegura que a orientação técnica chegue ao educador de forma autônoma e exatamente no instante de sua necessidade, configurando um modelo de instrução de suporte imediato. Ao optar pelo formato audiovisual de curta duração, o material atende à urgência da prática em oficina, entregando a diretriz de funcionamento sem sobrecarregar cognitivamente o usuário que já lida com a novidade e a tensão do equipamento.

A concepção do conteúdo audiovisual foi intencionalmente direcionada para desconstruir a insegurança inerente ao primeiro contato com maquinários robustos. Para as ferramentas de operação mecânica, especificamente a furadeira e a serra tico-tico de bancada, a narrativa visual concentrou-se na mitigação do receio natural frente a equipamentos de perfuração e corte. Os roteiros enfatizaram exaustivamente a utilização correta dos equipamentos de proteção individual (EPIs) e o posicionamento ergonômico e seguro das mãos durante o manuseio das peças. Dessa forma, a mídia não se restringe a ensinar o acionamento da máquina, mas estabelece uma cultura de segurança visualmente assimilável, encorajando o docente a superar a hesitação inicial por meio da espelhagem da postura correta observada na tela.

No que tange aos equipamentos de fabricação digital, o desafio instrucional pautou-se pela decodificação da complexidade técnica em etapas visuais intuitivas. Os materiais referentes à impressora 3D e à cortadora CNC laser foram desenhados para desmistificar procedimentos que frequentemente afastam o público leigo, como as nuances do preparo da mesa de impressão e os rigorosos protocolos de segurança óptica exigidos pelo feixe de laser. A linguagem adotada nos vídeos priorizou a demonstração clara das interfaces de controle e do comportamento esperado das máquinas ao longo do processo. Essa estratégia transformou conceitos abstratos de fabricação digital em sequências de ações lógicas e replicáveis, reduzindo significativamente a ansiedade operacional frente a tecnologias avançadas.

A discussão destes resultados evidencia que a facilitação do acesso técnico atua como um catalisador direto para a inovação pedagógica. Ao eliminar a barreira do medo por meio de uma instrução clara e posicionada estrategicamente no ambiente físico, o docente conquista uma autonomia tecnológica basilar. É precisamente essa independência operacional que permite ao professor deslocar sua carga de atenção, deixando de focar excessivamente na mecânica da ferramenta para concentrar seus esforços no planejamento estratégico da aula. Sem a preocupação paralisante com a operação básica, o educador encontra espaço mental para desenhar sequências didáticas investigativas.

Consequentemente, a infraestrutura de apoio desenvolvida alinha-se de forma coesa à premissa do projeto eSTEAMulAR e às definições de Maia e Soares sobre a inovação educacional. A apropriação segura do espaço e das ferramentas possibilita que a abordagem STEAM transcenda o plano discursivo e se materialize em atividades práticas voltadas à resolução de problemas do mundo real. Os equipamentos do laboratório, que em um primeiro momento representavam potenciais obstáculos técnicos causadores de insegurança, passam a ser plenamente assimilados como recursos pedagógicos integrados, essenciais para o fomento da criatividade, do pensamento crítico e da interdisciplinaridade no cotidiano da Educação Básica.

Conclusão

A presente pesquisa cumpriu seu objetivo ao estruturar e disponibilizar um conjunto de materiais didáticos multimidiáticos voltados para a capacitação docente no contexto da abordagem STEAM. A produção dos quatro vídeos instrucionais e a sua subsequente integração ao espaço físico do eSTEAMuLab, viabilizada pelo uso de códigos QR fixados nas estações de trabalho, consolidaram uma solução de suporte tecnológico ágil e sob demanda. Essa arquitetura instrucional revelou-se estratégica para contornar a barreira do desconhecimento técnico, entregando a diretriz de funcionamento precisa no exato instante em que o educador interage com o maquinário.

A desmistificação do manuseio da furadeira de bancada, da serra tico-tico, da impressora 3D e da cortadora CNC laser atua de forma direta na superação da insegurança inerente ao primeiro contato com o laboratório de prototipação. Ao prover um material de apoio visualmente acessível e ancorado em protocolos rígidos de segurança, o trabalho desempenha um papel fundamental na construção da autonomia tecnológica do professor. Esta independência operacional figura como a condição basilar que permite ao educador transcender a simples manipulação mecânica da ferramenta, encorajando-o a focar sua energia criativa no planejamento de metodologias investigativas que respondam aos desafios reais do ambiente escolar.

Como desdobramentos futuros, o projeto estabelece a aplicação sistemática desta infraestrutura de apoio durante a condução das oficinas presenciais de formação. O contato direto e contínuo com os docentes da Educação Básica configurará um cenário real de validação, permitindo avaliar a usabilidade dos vídeos e coletar percepções fundamentais para o aprimoramento iterativo do material. Além disso, projeta-se a expansão deste acervo instrucional para contemplar os demais equipamentos analógicos e digitais do laboratório, garantindo que o espaço se consolide continuamente como um polo seguro e propulsor de inovação pedagógica por meio da abordagem STEAM.

Referências

MAIA, Dennys L.; SOARES, Raíza. A conceptual framework for STEAM Education: a proposal from and for the Global South. 2025.

MAIA, Dennys L. et al. Abordagem STEAM na Educação básica brasileira: uma revisão de literatura. Revista Tecnologia e Sociedade, v. 17, n. 49, p. 68-88, 2021.

D'AMBRÓSIO, Ubiratan. Sobre las propuestas curriculares STEM y STEAM y el Programa de Etnomatemática. Revista Paradigma, v. 41, p. 151-167, jun. 2020.

Anexos



qr codes



qr codes



qr codes



qr codes

CÓDIGO: HS1029

AUTOR: TAYNARA BEATRIZ SALES DA SILVA

COAUTOR: LUCAS TOSHIO NASCIMENTO DA SILVA

ORIENTADOR: DENNYS LEITE MAIA

TÍTULO: PRODUÇÃO DE MATERIAL DIDÁTICO DIGITAL PARA DOCENTES SOBRE ABORDAGEM STEAM

Resumo

Este trabalho apresenta o desenvolvimento inicial de um plano de trabalho vinculado ao projeto eSTEAMulAR, desenvolvido no Instituto Metrópole Digital da Universidade Federal do Rio Grande do Norte, cujo foco é a produção de materiais didáticos digitais destinados à formação docente para a abordagem STEAM (Ciência, Tecnologia, Engenharia, Artes e Matemática) na Educação Básica. O estudo parte da necessidade de ampliar o acesso de professores a recursos pedagógicos que contribuam para a compreensão e utilização de práticas interdisciplinares e tecnologias educacionais. O processo de desenvolvimento dos materiais é orientado pelo modelo ADDIE, contemplando as etapas de análise, planejamento, desenvolvimento, implementação e avaliação. No período analisado, foram realizadas atividades de estudo e acompanhamento da equipe de pesquisa, participação em oficinas de formação com professores e estudantes de licenciatura, além da aproximação com equipamentos disponíveis no eSTEAMuLab, como impressora 3D, cortadora CNC, serra tico-tico de bancada e furadeira de bancada. Também foram realizados registros audiovisuais de demonstrações relacionadas ao uso desses equipamentos, como etapa inicial para a produção de materiais tutoriais. Os resultados preliminares indicam que a participação em atividades formativas e a exploração prática dos recursos tecnológicos contribuem para a compreensão das demandas relacionadas à produção de materiais digitais para formação docente em STEAM.

Palavras-chave: STEAM, Material Didático, Formação Docente, Tecnologia educacional.

TITLE: Production of Digital Educational Material for Teachers on the STEAM Approach

Abstract

This paper presents the initial development of a work plan linked to the eSTEAMulAR project—carried out at the Digital Metropolis Institute of the Federal University of Rio Grande do Norte—which focuses on producing digital educational materials for teacher training in the STEAM (Science, Technology, Engineering, Arts, and Mathematics) approach within K-12 education. The study stems from the need to expand teachers' access to pedagogical resources that foster an understanding and application of interdisciplinary practices and educational technologies. The material development process follows the ADDIE model, encompassing the stages of analysis, design, development, implementation, and evaluation. During the period under review, activities included research team study and monitoring, participation in training workshops with teachers and pre-service teachers, and hands-on engagement with equipment available at the eSTEAMuLab, such as a 3D printer, CNC cutter, benchtop jigsaw, and bench drill. Audiovisual recordings of demonstrations regarding the use of this equipment were also produced as an initial step toward creating tutorial materials. Preliminary results indicate that participating in training activities and practically exploring technological resources contribute to understanding the requirements for producing digital materials for STEAM teacher training.

Keywords: STEAM, Teaching Materials, Teacher Training, Educational Technology.

Introdução

A educação contemporânea tem sido marcada pela crescente presença das tecnologias digitais e pela necessidade de desenvolver práticas pedagógicas capazes de integrar diferentes áreas do conhecimento. Nesse cenário, a abordagem STEAM, acrônimo para Science, Technology, Engineering, Arts and Mathematics, destaca-se por propor uma perspectiva interdisciplinar de ensino e aprendizagem, favorecendo processos de investigação, criatividade, resolução de problemas e construção de soluções para situações contextualizadas (Maia et al., 2021).

No contexto da Educação Básica brasileira, a abordagem STEAM apresenta possibilidades para aproximar conhecimentos científicos, tecnológicos e artísticos das experiências dos estudantes. Entretanto, sua implementação enfrenta desafios relacionados à formação docente, à organização das práticas pedagógicas e à disponibilidade de recursos didáticos adequados. Revisões e levantamentos realizados por pesquisadores do grupo (Maia, Carvalho e Appelt, 2021; Soares; Maia, 2023) evidenciam que as experiências STEAM desenvolvidas no Brasil ainda são pouco numerosas e apresentam concentração regional e disciplinar, indicando a necessidade de ampliar e diversificar essas práticas (Maia; Carvalho; Appelt, 2021). No contexto do Rio Grande do Norte, pesquisas do grupo também têm investigado experiências desenvolvidas em iniciativas como a Feira Brasileira de Ciências e Engenharia (FEBRACE), destacando a presença de práticas investigativas e criativas relacionadas à abordagem STEAM (Maia et al., 2024).

A formação docente constitui, portanto, um aspecto fundamental para a implementação de práticas pedagógicas que integrem tecnologias e diferentes áreas do conhecimento. A presença de equipamentos e recursos tecnológicos no ambiente escolar, por si só, não

garante sua utilização pedagógica. É necessário que os professores desenvolvam conhecimentos que possibilitem articular os recursos tecnológicos aos objetivos de aprendizagem, aos conhecimentos específicos de suas áreas e às estratégias pedagógicas adotadas. Nessa perspectiva, a integração entre conhecimentos tecnológicos, pedagógicos e de conteúdo constitui um elemento importante para o planejamento de experiências educacionais mediadas por tecnologias (Mishra et al., 2023).

Além da formação dos professores, outro desafio está relacionado à disponibilidade de materiais didáticos que possam apoiar processos formativos. Recursos digitais podem ampliar as possibilidades de acesso a conteúdos, procedimentos, tutoriais e exemplos de práticas pedagógicas, especialmente quando organizados de maneira acessível, interativa e contextualizada. Dessa forma, materiais digitais podem funcionar como recursos de apoio à formação docente, permitindo que os professores consultem conteúdos e procedimentos de acordo com suas necessidades.

Nesse contexto, o eSTEAMuLab, espaço de formação e experimentação vinculado ao Instituto Metrópole Digital, constitui um ambiente relevante para o desenvolvimento de práticas relacionadas à abordagem STEAM. O espaço dispõe de equipamentos e recursos tecnológicos, como impressora 3D, cortadora CNC e kits de eletrônica, que podem ser explorados em atividades de investigação, criação, prototipagem e resolução de problemas.

A partir desse contexto, evidencia-se a necessidade de disponibilizar recursos digitais que possam apoiar professores no processo de aproximação com equipamentos e tecnologias relacionados à abordagem STEAM. Nesse sentido, o presente trabalho concentra-se na produção de materiais didáticos digitais, com ênfase em vídeos tutoriais sobre o funcionamento e a utilização de equipamentos disponíveis no eSTEAMuLab, destinados a apoiar processos de formação docente.

Metodologia

O trabalho foi desenvolvido no contexto do projeto eSTEAMulAR e do eSTEAMuLab, no Instituto Metrópole Digital da Universidade Federal do Rio Grande do Norte (IMD/UFRN), tendo como foco a produção de materiais didáticos digitais destinados ao apoio à formação docente em STEAM. Para a organização do processo de desenvolvimento dos materiais, adotou-se como referência o modelo ADDIE (Filatro, 2008), composto pelas etapas de Analysis (Análise), Design (Planejamento), Development (Desenvolvimento), Implementation (Implementação) e Evaluation (Avaliação). Neste trabalho, foram consideradas, até o momento, as etapas de análise, design e desenvolvimento, ou seja, a etapa de concepção (Filatro, 2008), de acordo com o estágio de execução das atividades.

Análise: A etapa de análise compreende a aproximação com o contexto formativo e tecnológico no qual os materiais serão desenvolvidos. Nesse período, foram acompanhadas atividades de formação realizadas com professores e estudantes de licenciatura, nas quais foi possível auxiliar na aplicação das atividades e identificar

demandas dos participantes sobre a abordagem STEAM e recursos disponíveis no Esteamulab. Também foram acompanhadas reuniões semanais da equipe de pesquisa e atividades de grupo de estudos, possibilitando a aproximação com os fundamentos teóricos e metodológicos do projeto.

Além disso, foi realizada uma aproximação inicial com os equipamentos disponíveis no eSTEAMuLab, incluindo impressora 3D, cortadora CNC, serra tico-tico de bancada e furadeira de bancada. Essa etapa permitiu conhecer mais sobre as possibilidades de utilização dos equipamentos em atividades relacionadas à abordagem STEAM e identificar conteúdos que poderão ser contemplados nos materiais digitais.

Design: A etapa de design correspondeu ao planejamento dos materiais didáticos digitais a serem produzidos, considerando o público docente e as características do contexto formativo do eSTEAMuLab. A partir da aproximação realizada na etapa de análise, definiu-se a utilização de vídeos em formato de tutorial como recurso para apresentar os equipamentos disponíveis no espaço e apoiar a familiarização dos docentes com suas características e possibilidades de utilização.

Durante o planejamento, foram considerados como público-alvo professores da Educação Básica participantes das ações de formação desenvolvidas pelo projeto. Dessa forma, buscou-se utilizar uma linguagem clara e objetiva, priorizando a demonstração prática dos equipamentos e a apresentação de informações relacionadas ao seu funcionamento e à sua utilização.

Foram selecionados para a produção dos materiais quatro equipamentos disponíveis no eSTEAMuLab: impressora 3D, cortadora CNC, serra tico-tico de bancada e furadeira de bancada. A escolha desses equipamentos esteve relacionada às experiências de formação acompanhadas pela bolsista e à possibilidade de utilizar os vídeos como recursos complementares às atividades desenvolvidas no espaço.

A organização dos materiais foi pensada de forma a apresentar um equipamento por vez, permitindo concentrar as informações e as demonstrações práticas em cada recurso tecnológico. Nesse sentido, os vídeos foram planejados para privilegiar a visualização dos equipamentos durante seu funcionamento, associando a explicação oral à demonstração prática dos procedimentos. Buscou-se, assim, produzir materiais objetivos e de fácil consulta, que pudessem ser utilizados como apoio às atividades de formação docente.

Desenvolvimento: A etapa de desenvolvimento correspondeu à produção dos materiais audiovisuais planejados na etapa anterior. Foram realizados registros em vídeo de demonstrações relacionadas ao funcionamento e à utilização dos equipamentos disponíveis no eSTEAMuLab, contemplando a impressora 3D, a cortadora CNC, a serra tico-tico de bancada e a furadeira de bancada.

As gravações foram realizadas individualmente para cada equipamento. Durante os registros, um integrante da equipe realizou a explicação e a demonstração prática dos equipamentos, enquanto a bolsista participou do processo realizando a captação das imagens em vídeo. Essa participação possibilitou registrar os procedimentos demonstrados e organizar os registros para posterior edição.

Após as gravações, os materiais audiovisuais foram editados e o processo resultou na produção de vídeos em formato de tutorial, destinados a apresentar as características, o funcionamento e os procedimentos relacionados à utilização dos equipamentos.

Os vídeos produzidos foram disponibilizados no canal @esteamulab, na plataforma YouTube, constituindo recursos digitais que podem ser utilizados como materiais complementares nas ações de formação docente desenvolvidas no âmbito do projeto eSTEAMulAR. Os tutoriais produzidos são:

Cortadora CNC a laser:https://www.youtube.com/shorts/MFVk-fxjjGo

Impressora 3D:https://www.youtube.com/shorts/L-KJ0PIpfUM

Serra tico-tico de bancada:https://www.youtube.com/shorts/6dJm-YEAH8g

Furadeira de bancada:https://www.youtube.com/shorts/9txxbchvXOY

Resultados e Discussões

Os resultados apresentados correspondem às atividades desenvolvidas no período inicial de execução do plano de trabalho. Considerando que a pesquisa se encontra em andamento, os resultados são apresentados como preliminares e estão relacionados principalmente à participação em atividades formativas, à aproximação com as tecnologias disponíveis no eSTEAMuLab e às etapas iniciais de produção dos recursos digitais.

A participação nas oficinas de formação realizadas com professores e estudantes de licenciatura constituiu uma das primeiras experiências da pesquisadora no desenvolvimento do plano de trabalho. Embora sua atuação não tenha envolvido a condução das oficinas, a participação como auxiliar possibilitou acompanhar a organização das atividades e observar a interação dos participantes com as propostas formativas e os recursos tecnológicos utilizados. Durante essas experiências, foram observadas dificuldades recorrentes relacionadas à familiarização com os equipamentos, incluindo a compreensão de suas funções e formas de funcionamento, o manuseio, a identificação dos procedimentos necessários para sua utilização e a compreensão de comandos e configurações. Também foram observadas dificuldades mais específicas relacionadas à operação da impressora 3D e da cortadora CNC, bem como à compreensão dos cuidados e procedimentos de segurança necessários durante o uso dos equipamentos.

Essas observações contribuíram para a etapa de análise do processo de desenvolvimento dos materiais, uma vez que permitiram identificar aspectos que poderiam ser considerados na elaboração dos recursos digitais. A partir dessa experiência, verificou-se a importância de disponibilizar materiais que apresentassem os equipamentos de maneira visual e objetiva, contemplando suas funções, características, formas de

utilização e procedimentos básicos de operação. Essa percepção orientou a etapa de design e contribuiu para a definição dos equipamentos como foco dos materiais audiovisuais produzidos.

Nesse sentido, a escolha dos vídeos em formato de tutorial mostrou-se adequada às necessidades observadas nas oficinas, pois possibilita associar explicações à demonstração prática dos equipamentos. A produção dos materiais passou, assim, a priorizar recursos que pudessem auxiliar na familiarização dos docentes com as tecnologias disponíveis no eSTEAMuLab e servir como apoio às atividades formativas desenvolvidas pelo projeto.

Essa experiência foi relevante para a compreensão de que a integração de tecnologias às práticas pedagógicas demanda não apenas a disponibilidade dos equipamentos, mas também orientação e conhecimentos que permitam ao professor relacioná-los aos objetivos educacionais. Essa perspectiva dialoga com a discussão sobre a integração entre conhecimentos tecnológicos, pedagógicos e de conteúdo na formação docente.

A participação nas reuniões semanais da equipe e nas atividades de grupo de estudos contribuiu para a aproximação da bolsista com os fundamentos teóricos e metodológicos relacionados ao projeto. Nesse processo, o estudo de Rossi, Diogo e Mello (2025) contribuiu para compreender diferentes possibilidades de utilização da abordagem STEAM no Ensino Fundamental, especialmente em propostas que favorecem a aprendizagem ativa, o desenvolvimento de habilidades e competências e a participação dos estudantes. Essa perspectiva permitiu relacionar os fundamentos estudados às experiências acompanhadas nas atividades formativas do projeto, nas quais o contato com equipamentos e tecnologias foi realizado por meio de demonstrações e atividades práticas. A partir dessa relação entre os referenciais estudados e as experiências vivenciadas, compreendeu-se que os materiais digitais produzidos poderiam contribuir não apenas para apresentar o funcionamento dos equipamentos, mas também para aproximar os docentes de recursos tecnológicos que podem ser utilizados em propostas práticas e contextualizadas de ensino.

Outra atividade desenvolvida no período inicial foi a aproximação com os equipamentos disponíveis no eSTEAMuLab. O acompanhamento de demonstrações relacionadas à impressora 3D, à cortadora CNC, à serra tico-tico de bancada e à furadeira de bancada contribuiu para compreender suas características, possibilidades de utilização e procedimentos básicos de operação. Esse contato direto com os equipamentos constituiu um dos insumos para a produção dos vídeos tutoriais, uma vez que permitiu identificar quais informações e procedimentos precisavam ser apresentados nos materiais.

Esse contato direto com os equipamentos constituiu um importante insumo para a produção dos materiais didáticos digitais, pois permitiu conhecer os recursos que seriam posteriormente apresentados nos vídeos tutoriais e identificar quais informações e procedimentos precisavam ser registrados durante as demonstrações. Dessa forma, a familiarização com os equipamentos contribuiu para orientar a produção dos registros audiovisuais, possibilitando que os materiais contemplassem aspectos relacionados ao funcionamento, à utilização e aos cuidados necessários durante o uso dos recursos tecnológicos disponíveis no eSTEAMuLab.

Como parte das atividades de desenvolvimento, foram realizados registros audiovisuais das demonstrações dos equipamentos disponíveis no eSTEAMuLab. As gravações foram realizadas durante demonstrações conduzidas por um integrante da equipe, enquanto a pesquisadora participou da captação das imagens. Cada equipamento foi registrado individualmente, contemplando a impressora 3D, a cortadora CNC, a serra tico-tico de bancada e a furadeira de bancada. Após as gravações, os materiais audiovisuais foram editados, resultando na produção de vídeos em formato de tutorial, posteriormente disponibilizados no YouTube.

O processo de produção dos vídeos envolveu diferentes etapas, desde a preparação e realização das demonstrações até a captação e edição dos registros. A participação nesse processo possibilitou compreender aspectos relacionados à produção de recursos audiovisuais para fins educacionais, especialmente a importância de registrar os procedimentos de forma clara e de organizar as informações de modo que possam ser compreendidas pelo público ao qual os materiais se destinam.

Como forma de documentar o processo de elaboração dos tutoriais, foram realizados registros fotográficos e audiovisuais ao longo das atividades. Parte desses registros é apresentada neste trabalho em diferentes momentos da descrição do processo, contribuindo para a documentação das etapas de produção e das demonstrações realizadas.

As experiências realizadas até o momento indicam que a produção de materiais digitais para formação docente não constitui apenas uma atividade técnica de elaboração de conteúdos. Trata-se de um processo que envolve compreensão do público, seleção de conteúdos, domínio progressivo das tecnologias e articulação entre conhecimentos pedagógicos e tecnológicos. Nesse sentido, a utilização do modelo ADDIE mostra-se pertinente por organizar o desenvolvimento dos recursos em etapas articuladas, permitindo que a produção seja progressivamente planejada, desenvolvida, aplicada e avaliada.

Conclusão

O desenvolvimento inicial do plano de trabalho possibilitou uma aproximação com os fundamentos da abordagem STEAM, com processos de formação docente e com tecnologias utilizadas no eSTEAMuLab. A participação em oficinas, reuniões da equipe e atividades de grupo de estudos contribuiu para a construção de conhecimentos relacionados ao projeto e para a compreensão das demandas envolvidas na produção de materiais didáticos digitais.

A exploração inicial da impressora 3D, da cortadora CNC, da serra tico-tico de bancada e da furadeira de bancada, juntamente com a realização de registros audiovisuais de demonstrações, constituiu uma etapa inicial para a produção de recursos digitais voltados à formação docente.

Os resultados obtidos até o momento são preliminares, uma vez que as etapas de desenvolvimento, implementação e avaliação dos materiais ainda estão em andamento. Espera-se que a continuidade do trabalho possibilite o aprimoramento dos recursos produzidos e sua utilização em processos formativos, contribuindo para ampliar as possibilidades de acesso a conteúdos e práticas relacionadas à abordagem STEAM.

Referências

CHIU, T. K. F. et al. Teacher professional development for STEAM education: A systematic review. Computers & Education, v. 208, p. 104950, 2024.

D'AMBRÓSIO, U. Educação para uma sociedade em transição. 2. ed. Natal: Editora da UFRN, 2020.

FILATRO, Andrea. Design instrucional contextualizado: educação e tecnologia. 2. ed. São Paulo: Senac, 2008.

MAIA, Dennys Leite; CARVALHO, Rodolfo Araújo de; APPELT, Veridiana Kelin. Abordagem STEAM na educação básica brasileira: uma revisão de literatura. Revista Tecnologia e Sociedade, Curitiba, v. 17, n. 49, p. 68-87, 2021.

MAIA, Dennys Leite et al. Experiências com a abordagem STEAM na Educação Básica: análise de práticas pedagógicas. Revista Tecnologias na Educação, v. 35, p. 1-15, 2021.

MAIA, Dennys Leite et al. Investigação e criação na Feira Brasileira de Ciências e Engenharia (FEBRACE): contribuições da abordagem STEAM. In: CONGRESSO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 2024. Anais [...]. Porto Alegre: SBC, 2024. p. 1-10.

MISHRA, Punya et al. TPACK in the age of AI: Rethinking technological pedagogical content knowledge. Journal of Digital Learning in Teacher Education, v. 39, n. 3, p. 140-151, 2023.

PAPERT, Seymour. Mindstorms: Children, computers, and powerful ideas. New York: Basic Books, 1980.

ROSSI, Mayara; DIOGO, Rodrigo Claudino; MELLO, Geison Jader. Investigações sobre o uso da abordagem STEAM na prática escolar. Ensaio Pesquisa em Educação em Ciências, Belo Horizonte, v. 27, e53892, 2025.

SKOWRONEK, M. et al. Maker education and digital fabrication in teacher training: Challenges and perspectives. Educational Technology Research and Development, v. 70, p. 815-835, 2022.

Anexos



Figura 1 – Impressora 3D durante atividade de demonstração no eSTEAMuLab. Fonte: Acervo da pesquisa (2026).



Figura 2 – Registro da gravação de tutorial sobre a serra tico-tico de bancada no eSTEAMuLab.



Figura 3 – Cortadora CNC durante demonstração no eSTEAMuLab. Fonte: Acervo da pesquisa (2026).

CÓDIGO: HS1383

AUTOR: JOÃO PAULO DE MEDEIROS NETO

ORIENTADOR: DENNYS LEITE MAIA

TÍTULO: Refatoração de Objetos de Aprendizagem Gerados por IA nas oficinas do Inovática: Engenharia de Software na Plataforma OBAMA.

Resumo

Este artigo trata do processo de refatoração técnica de Objetos de Aprendizagem (OAs) gerados por professores da Educação Básica de Luís Gomes e Natal (RN) com suporte de Inteligência Artificial Generativa (IA) nas oficinas do projeto Inovática. O trabalho é fruto das ações de desenvolvimento da Plataforma OBAMA e tem como objetivo corrigir, otimizar e homologar protótipos de OAs focados na abordagem STEAM para disponibilização pública. Para este trabalho, realizou-se a análise de código-fonte dos recursos gerados via IAGen, aplicando metodologias de engenharia de software web para corrigir falhas de lógica, sintaxe e responsividade. Foram identificadas falhas estruturais recorrentes nos códigos (HTML5, CSS3, JavaScript) gerados por ferramentas como ChatGPT, incluindo validações incorretas em OAs e erros factuais de domínio (alucinações). O estudo oportunizou conhecer que, embora a IA acelere a prototipagem por docentes, a mediação técnica de um desenvolvedor é imprescindível para transformar esses protótipos em softwares educativos abertos, funcionais, seguros e acessíveis para a Educação Básica.

Palavras-chave: Objetos de Aprendizagem. Inteligência Artificial. Engenharia Web.

TITLE: Refactoring AI-Generated Learning Objects in Inovática Workshops: Software Engineering on the OBAMA Platform.

Abstract

This article discusses the process of technically refactoring Learning Objects (LOs) created by elementary and secondary school teachers in Luís Gomes and Natal (RN) with the support of Generative Artificial Intelligence (AI) during the Inovática project's workshops. This work is the result of development efforts for the OBAMA Platform and aims to correct, optimize, and validate LO prototypes focused on the STEAM approach for public release. For this work, we analyzed the source code of the resources

generated via IAGen, applying web software engineering methodologies to correct flaws in logic, syntax, and responsiveness. Recurring structural flaws were identified in the code (HTML5, CSS3, JavaScript) generated by tools such as ChatGPT, including incorrect validations in educational applications and factual domain errors (hallucinations). The study revealed that, although AI accelerates prototyping by educators, the technical guidance of a developer is essential to transform these prototypes into open, functional, secure, and accessible educational software for K-12 education.

Keywords: Learning Objects. Artificial Intelligence. Engineering Web.

Introdução

A integração da abordagem STEAM (Ciência, Tecnologia, Engenharia, Artes e Matemática) com a Inteligência Artificial (IA) tem se consolidado como uma demanda urgente na Educação Básica brasileira (BACICH; HOLANDA, 2020; VICARI; BRACKMANN; MIZUSAKI, 2023). O projeto Inovática investiga essa formação docente, enquanto a Plataforma OBAMA atua como o principal referatório de Objetos de Aprendizagem (OAs) interativos para materializar essa integração em sala de aula (SILVA et al., s.d.; OLIVEIRA et al., 2018).

Ferramentas de IA generativa (como ChatGPT e DeepSeek) foram utilizadas por professores em formação para prototipar recursos educacionais, sem o conhecimento técnico de programação. Entretanto, observou-se que os códigos gerados por essas ferramentas frequentemente apresentavam problemas de implementação e funcionamento, ainda que muitos fossem executáveis. OAs prototipados exclusivamente via prompting vinham com lógicas incompletas, incompatibilidade com navegadores móveis, e, por vezes, erros conceituais no próprio conteúdo didático embarcado no código-fonte.

Diante do desafio de compreender até que ponto as novas tecnologias garantem que as intenções pedagógicas dos docentes se convertam em softwares educacionais robustos, acessíveis e no padrão web (W3C), este trabalho objetiva relatar o processo de desenvolvimento, refatoração e homologação técnica de Objetos de Aprendizagem interativos gerados em oficinas de IAGen para a Plataforma OBAMA.

Metodologia

O trabalho foi conduzido em um ambiente de desenvolvimento local baseado em sistema operacional Linux (Mint), utilizando metodologias ágeis de engenharia de software iterativa. O ciclo de vida da estrutura da refatoração foi estruturado em quatro fases principais:

A Fase 1 (Levantamento e Diagnóstico Técnico) consistiu na recepção de sete protótipos de OAs gerados pelos professores com suporte de IAGen disponibilizado à equipe para avaliação e tratamento antes de sua publicação na Plataforma OBAMA. Por intermédio do ambiente de desenvolvimento integrado (IDE) Visual Studio Code, realizou-se uma auditoria no código-fonte (arquivos estáticos) para mapear falhas de execução, memory leaks, erros de escopo de variáveis e incompatibilidades multiplataforma.

Na Fase 2 (Refatoração e Desenvolvimento), aplicou-se a recodificação dos artefatos. Utilizando a tríade HTML5, CSS3 e JavaScript, a lógica interna dos OAs foi reescrita. Foram implementados algoritmos de validação de input de usuários, modularização de funções e estilização responsiva.

A Fase 3 (Correção de Domínio e Validação) envolveu a revisão das bases de dados embutidas nos OAs. Como as IAs generativas estão sujeitas a "alucinações", procedeu-se com a validação rigorosa dos gabaritos contidos nos arrays e arquivos JSON consumidos pelo front-end, corrigindo dados factuais incorretos inseridos pela IA antes da renderização.

Por fim, a Fase 4 (Homologação e Disseminação) compreendeu a realização de testes de usabilidade cross-browser (desktop e mobile) e a integração final dos OAs ao banco de dados da Plataforma OBAMA para uso nas escolas.

Resultados e Discussões

Durante a etapa de diagnóstico, identificou-se um padrão de falha semântica em todos os códigos-fonte gerados inicialmente pela IA. Observou-se a presença generalizada de aspas duplicadas ou quebradas e trechos de código que corromperam a estrutura HTML inicial. Para solucionar essa limitação, foi necessário o desenvolvimento e execução de um script em Python para realizar a leitura, limpeza e reescrita em nível de arquivo de todos os documentos HTML, garantindo que os OAs ficassem prontos para uso.

Ao longo do ciclo iterativo de desenvolvimento, os apontamentos documentados nos protótipos analisados distribuíram-se por múltiplas camadas arquiteturais, abrangendo desde a interface gráfica (front-end) até a precisão conceitual e a lógica de negócios. No OA "Jogo da Velha Matemático", diagnosticaram-se inconsistências no registro de eventos de toque e na validação das condições de vitória. A intervenção exigiu a reescrita modular das funções de verificação em JavaScript, ajustes nos event listeners e a separação estrita entre a camada de apresentação e as regras de negócio. Analogamente, o OA "Influencer da Matemática" apresentou vulnerabilidades no

tratamento de inputs, falhando ao processar variáveis de ponto flutuante e dados nulos, o que possibilitava o atingimento de estados de vitória com pontuações negativas. A refatoração englobou a implementação de estruturas de controle rigorosas, validação de entradas e a readequação algorítmica da lógica de pontuação.

No que concerne à integração de recursos multimídia, protótipos como "Caverna do Dragão" e "Missão Geo Exploradores" exibiram anomalias no referenciamento de ativos. A geração automatizada incluiu hiperlinks externos inacessíveis ou inexistentes, resultando na ausência de renderização das ilustrações didáticas e na exibição restrita de atributos alternativos (alt). A mitigação desse problema consistiu na substituição sistemática das referências externas pela importação estática de imagens armazenadas localmente no diretório do projeto, garantindo a integridade visual da aplicação.

A auditoria e refatoração dos OAs gerados por IA revelaram padrões claros nas limitações atuais da geração de código automatizada para fins educacionais. Durante a análise técnica, constatou-se que a IA generativa é altamente eficaz em montar o "esqueleto" do Objeto de Aprendizagem (estruturas HTML básicas e layouts iniciais), mas falha substancialmente na gestão de estado e na lógica de aplicações interativas.

Como os sete OAs correspondem ao conjunto de protótipos analisados, o estudo não pretende estimar a frequência de problemas em todos os OAs produzidos com IAG. O que se pode afirmar é que, nesse ciclo específico, os problemas documentados se repetiram em diferentes camadas do software, indo da interface e dos recursos multimídia à lógica interna e à correção do conteúdo.

Por fim, a refatoração demonstrou um ganho de performance. A otimização dos arquivos estáticos, a remoção de requisições redundantes geradas pela IA e a adaptação do CSS reduziram o tempo de carregamento dos OAs, um fator crítico para a acessibilidade em escolas públicas com infraestrutura de rede limitada.

Conclusão

O presente trabalho evidenciou que a adoção da IA Generativa na formação docente possui imenso potencial para a ideação criativa, mas não dispensa a necessidade do desenvolvimento e da engenharia de software. A transformação de um prompt em um Objeto de Aprendizagem funcional, responsivo e pedagogicamente correto exige processos rigorosos de refatoração de código, correção de bugs de lógica e validação humana.

Conclui-se que o trabalho técnico de desenvolvimento da Plataforma OBAMA — corrigindo rotinas em JavaScript, adaptando interfaces e sanando erros conceituais (alucinações) injetados pelas máquinas nos gabaritos — é o que efetivamente garante a disponibilização de recursos educacionais abertos (REA) de qualidade. A atuação do desenvolvedor consolida a ponte entre a inovação pensada pelo professor nas oficinas do Inovática e a realidade estrutural e pedagógica das escolas públicas potiguares.

Referências

VICARI, Rosa Maria; BRACKMANN, Christian Puhlmann; MIZUSAKI, Lorraine. Inteligência Artificial na Educação Básica. São Paulo: Novatec Editora, 2023.

SILVA, Ana Carolina Costa et al. Projeto Inovática: relato de experiência com a criação de objetos de aprendizagem com Inteligência Artificial Generativa buscando potencializar a autonomia docente. [S.d.].

CÓDIGO: HS1421

AUTOR: ANA CAROLINA COSTA SILVA

ORIENTADOR: DENNYS LEITE MAIA

TÍTULO: FORMAÇÃO CONTINUADA EM AMBIENTES VIRTUAIS: ANÁLISE DE PLATAFORMAS EDUCACIONAIS E SUAS POTENCIALIDADES PARA PRÁTICAS INVESTIGATIVAS E CRIATIVAS

Resumo

Este artigo trata da oferta de cursos de formação continuada para professores em plataformas educacionais digitais. O trabalho parte da necessidade de compreender como esses espaços contribuem para a inovação docente e a transformação das práticas em sala de aula, e tem como objetivo analisar plataformas educacionais que oferecem cursos de formação continuada, com foco no desenvolvimento de práticas investigativas e criativas de ensino. Realizou-se uma busca por essas plataformas no Google utilizando descritores específicos, seguida do mapeamento e análise das propostas formativas. Foram selecionadas seis plataformas com cursos gratuitos relacionados à Educação. Os resultados evidenciaram a oferta de formações sobre metodologias ativas, mas também a predominância de videoaulas, leituras e questionários. Embora as propostas apresentem potencial para práticas investigativas e criativas, ainda são limitadas as oportunidades de planejamento e experimentação docente. Conclui-se que as plataformas ampliam o acesso à formação continuada, mas podem fortalecer seu potencial por meio de experiências mais participativas e reflexivas.

Palavras-chave: Formação continuada. Plataformas educacionais. Autonomia docente.

TITLE: Continuing Education in Virtual Environments: Analysis of Educational Platforms and Their Potential for Investigative and Creative Practices

Abstract

This article addresses the provision of continuing education courses for teachers on digital educational platforms. The study stems from the need to understand how these spaces contribute to teaching innovation and the transformation of classroom practices; it aims to analyze educational platforms offering continuing education courses, focusing on the development of inquiry-based and creative teaching practices. A search for such platforms was conducted on Google using specific keywords, followed by a mapping and analysis of the training programs. Six platforms offering free education-related

courses were selected. The results revealed the availability of training on active learning methodologies, alongside a predominance of video lectures, readings, and quizzes. Although the programs show potential for fostering inquiry-based and creative practices, opportunities for teacher planning and experimentation remain limited. The study concludes that while these platforms expand access to continuing education, they could enhance their impact by incorporating more participatory and reflective experiences.

Keywords: Continuing education. Educational platforms. Teacher autonomy.

Introdução

A formação continuada consolida-se como um pilar indispensável no aprimoramento constante dos saberes e competências que regem a prática docente. De acordo com Libâneo (2014), diante das transformações sociais, culturais e tecnológicas, o trabalho do professor não deve restringir-se à mera transmissão passiva de conteúdos. Dessa forma, torna-se necessário promover práticas de ensino que favoreçam a participação ativa dos estudantes no processo de aprendizagem, possibilitando experiências mais significativas e contextualizadas.

Nesse sentido, práticas investigativas e criativas de ensino surgem como importantes possibilidades para repensar os processos de aprendizagem Segundo Maia et al. (2024), essas experiências permitem que os estudantes analisem situações do mundo real, relacionando-as aos conhecimentos estudados em sala de aula, bem como buscar informações e orientar a aprendizagem a partir de um processo investigativo. Ao propor situações em que os estudantes questionam, exploram e constroem respostas é possível fomentar o protagonismo dos estudantes no processo de aprendizagem.

Assim, a partir de processos formativos os professores podem ampliar seus conhecimentos e refletir sobre os desafios vivenciados em seus contextos de atuação. Mais do que oferecer metodologias ou técnicas a serem aplicadas em sala de aula, esses momentos podem se constituir como espaços de reflexão crítica do fazer docente. Assim, a formação continuada favorece o fortalecimento da identidade profissional, uma vez que permite que o professor se reconheça como sujeito ativo de sua formação.

Nesse contexto, com o avanço das tecnologias digitais, novas possibilidades também têm sido incorporadas aos processos de formação docente. As plataformas educacionais virtuais passaram a ampliar o acesso a cursos, materiais e diferentes experiências formativas. Entretanto, a simples oferta de cursos em ambientes digitais não garante, por si só, contribuições efetivas para a transformação da prática pedagógica.

Considerando esse cenário, levanta-se a seguinte questão: diante da crescente disponibilidade de plataformas educacionais, quais possibilidades elas oferecem para apoiar a formação continuada de professores da Educação Básica na apropriação e implementação de práticas investigativas e criativas de ensino? Busca-se compreender as características das propostas formativas disponibilizadas nessas plataformas e analisar como elas superam a exposição passiva de conteúdos.

Metodologia

Este estudo fundamenta-se como uma abordagem metodológica de caráter quali-quantitativo, uma vez que interpreta informações a partir da coleta de dados e estatísticas. De acordo com Schneider, Fujii e Corazza (2017), esse método propicia a análise da estrutura do fenômeno com métodos quantitativos, complementada por uma análise dos processos com métodos qualitativos. Para alcançar esse objetivo, portanto, foi realizado o mapeamento de plataformas educacionais digitais que oferecem cursos de formação, cujas informações foram organizadas e sistematizadas para posterior análise das características das propostas formativas.

Nesse sentido, a busca pelas plataformas foi realizada durante o mês de março de 2026, por meio do Google, utilizando os termos-chave “cursos educacionais”, “plataformas de cursos online” e “cursos de formação continuada”, bem como seus respectivos sinônimos. Assim, foram consideradas plataformas disponíveis em língua portuguesa que disponibilizam cursos relacionados à Educação.

Após a etapa inicial, procedeu-se à análise dos resultados encontrados, considerando como critérios de seleção: (i) plataforma ou ambiente digital que disponibilizasse curso de forma online; (ii) apresentar cursos relacionados à Educação; e (iii) oferta gratuita. A partir desses critérios, as 6 plataformas selecionadas foram organizadas em uma planilha para a sistematização das informações.

Com a compilação dos resultados das plataformas selecionadas, as informações foram examinadas tendo como foco questões orientadoras relacionadas às características e possibilidades formativas dos cursos identificados: “Quais estratégias metodológicas, recursos ou atividades são contemplados nos cursos oferecidos pelas plataformas?”; “De que forma as formações oferecidas possibilitam os professores planejar, adaptar e experimentar as práticas de ensino?”; “As propostas analisadas apresentam características investigativas e criativas ou permanecem centradas em uma abordagem tradicional de ensino?”. Sobre essas análises, apresentam-se os resultados a seguir.

Resultados e Discussões

Os resultados e discussões serão abordados nos próximos parágrafos de forma a responder as questões que foram levantadas anteriormente.

QUESTÃO 1 - Quais estratégias metodológicas, recursos ou atividades são contemplados nos cursos oferecidos pelas plataformas?

De modo geral, os cursos disponibilizam vídeo aulas, materiais de leituras, questionários e fóruns de discussão, organizados em módulos ou unidades temáticas. Os

recursos mais predominantes são vídeo aulas e materiais de leitura, que assumem o papel de apresentar e sistematizar os conteúdos abordados. Já os questionários e atividades avaliativas são utilizados principalmente para acompanhar a aprendizagem dos cursistas e também constituem pré-requisito para a validação da emissão dos certificados. Também foram identificados os fóruns de discussão em algumas propostas formativas, que possibilitam espaços para interação e para o esclarecimento de dúvidas com monitores do curso.

Ao comparar a organização dos cursos, também foram observadas diferenças na forma de apresentação e navegação das plataformas. No AVAMEC, por exemplo, os cursos analisados apresentam uma organização visual mais estruturada, com os conteúdos e atividades distribuídos de maneira clara ao longo dos módulos, o que favorece a localização dos materiais e a navegação na plataforma. Já a Escola Virtual Gov apresenta a estrutura visual mais simples e objetiva, com disposição mais tímida dos conteúdos e recursos. Embora ambas disponibilizam recursos semelhantes, observa-se que as plataformas adotam diferentes formas de apresentar o percurso formativo ao usuário.

Dessa forma, mesmo quando duas plataformas disponibilizam conteúdos semelhantes, o percurso realizado pelo cursista pode ser percebido e vivenciado de maneiras distintas. A organização do ambiente, constitui um elemento que interfere na experiência de formação, podendo favorecer ou dificultar o acesso e o acompanhamento das atividades. Assim, pensar em ambientes virtuais implica considerar não apenas o que é ofertado, mas também como é ofertado e quais condições são criadas para que o professor se aproprie desses recursos ao longo da formação.

QUESTÃO 2 - De que forma as formações oferecidas possibilitam os professores planejar, adaptar e experimentar as práticas de ensino?

As plataformas constituem importantes veículos para a formação continuada, ao ampliarem o acesso dos professores a conteúdos, metodologias e materiais formativos, especialmente por meio de cursos que podem ser realizados de maneira flexível. Entretanto, embora seja expressiva a oferta de materiais destinados à apresentação e apropriação dos conteúdos, não são incluídas muitas oportunidades para que os cursistas transformem seus conhecimentos em propostas concretas.

Essa questão pode ser observada, por exemplo, no curso sobre Aprendizagem Significativa disponibilizado na plataforma Aprenda Mais. A formação possibilita o contato com conceitos e fundamentos relacionados à abordagem, mas poderia propor situações em que o professor fosse convidado a elaborar uma atividade ou plano de aula fundamentada nos conhecimentos estudados. Essa proposta permitiria deslocar o cursista da posição de receptor dos conteúdos para a posição de agente ativo no processo formativo. Nesse sentido, responder questionários pode contribuir para a retomada e verificação dos conteúdos, mas não garante como o professor irá mobilizar esses conhecimentos no seu contexto escolar.

Assim, os resultados apontam para a ampliação das experiências formativas oferecidas pelas plataformas. De modo que os cursos ofertados podem potencializar sua

contribuição para a formação continuada ao criar condições para que o professor conheça novas possibilidades de ensino e também possa aplicar em sua própria prática.

QUESTÃO 3 - As propostas analisadas apresentam características investigativas e criativas ou permanecem centradas em uma abordagem tradicional de ensino?

Em relação às características investigativas e criativas, os resultados evidenciam uma oferta significativa de cursos cujas temáticas se aproximam de perspectivas que buscam superar modelos tradicionais de ensino. Foram identificadas formações voltadas às metodologias ativas, à abordagem STEAM, entre outras propostas que valorizam o protagonismo dos estudantes e a interdisciplinaridade. No AVAMEC, por exemplo, observa-se uma oferta expressiva de cursos relacionados à metodologias ativas, enquanto na plataforma Ápice disponibiliza formação voltada à implementação de espaços STEAM na escola.

A presença dessas temáticas representa uma possibilidade relevante para a formação continuada, pois amplia o contato dos professores com novas práticas de ensino. Dessa forma, as plataformas assumem um papel importante na disseminação de propostas que podem contribuir para que os docentes reflitam sobre os modelos de ensino predominantemente centrados na transmissão de conteúdo.

Contudo, embora cursos sobre STEAM e metodologias ativas surgiram um ensino pautado na investigação, autoria e criação, a estrutura interna das plataformas, mantém o cursista em uma posição predominantemente passiva. Assim, o caráter criativo e investigativo limita-se ao plano conceitual do discurso.

Conclusão

Nesse estudo, foi realizado um mapeamento de plataformas educacionais digitais que disponibilizam cursos gratuitos relacionados à educação. A partir da análise das seis plataformas selecionadas e das propostas formativas nelas disponibilizadas, tornou-se possível identificar os principais recursos, estratégias metodológicas e atividades presentes nos cursos, bem como compreender suas possibilidades para a apropriação e implementação de práticas investigativas e criativas de ensino.

De modo geral, verificou-se que ambientes virtuais de aprendizagem, ampliam o acesso a novas possibilidades de ensino. A presença de materiais de leitura, vídeo-aula, fóruns de discussão demonstra que as plataformas oferecem recursos que podem apoiar a aprendizagem de novos conceitos em diferentes contextos e com maior flexibilidade de acesso.

Entretanto, os resultados também apontam que a oferta dos recursos e cursos não garante necessariamente que a implementação das práticas aconteça. Diante disso, considera-se que o potencial das plataformas para a formação continuada pode ser ampliado. Além do contato com conceitos e metodologias, torna-se relevante propor

situações nas quais os cursistas possam analisar problemas de seu contexto, elaborar atividades, planejar intervenções pedagógicas e refletir sobre suas possibilidades de atuação. Dessa forma, a formação deixa de se limitar à apropriação de conceitos e passa a favorecer a mobilização dos conhecimentos em situações concretas, aproximando o percurso formativo dos desafios encontrados na prática docente.

Por fim, a partir desse trabalho foi possível evidenciar características e possibilidades relevantes a serem consideradas na oferta e no planejamento de cursos de formação continuada voltadas para o desenvolvimento de práticas que fomentem o protagonismo discente. Espera-se que esses achados contribuam para ampliar as discussões sobre a qualidade das experiências proporcionadas em ambientes virtuais de formação e para o desenvolvimento de propostas formativas que favoreçam uma participação mais ativa e autoral dos professores.

Referências

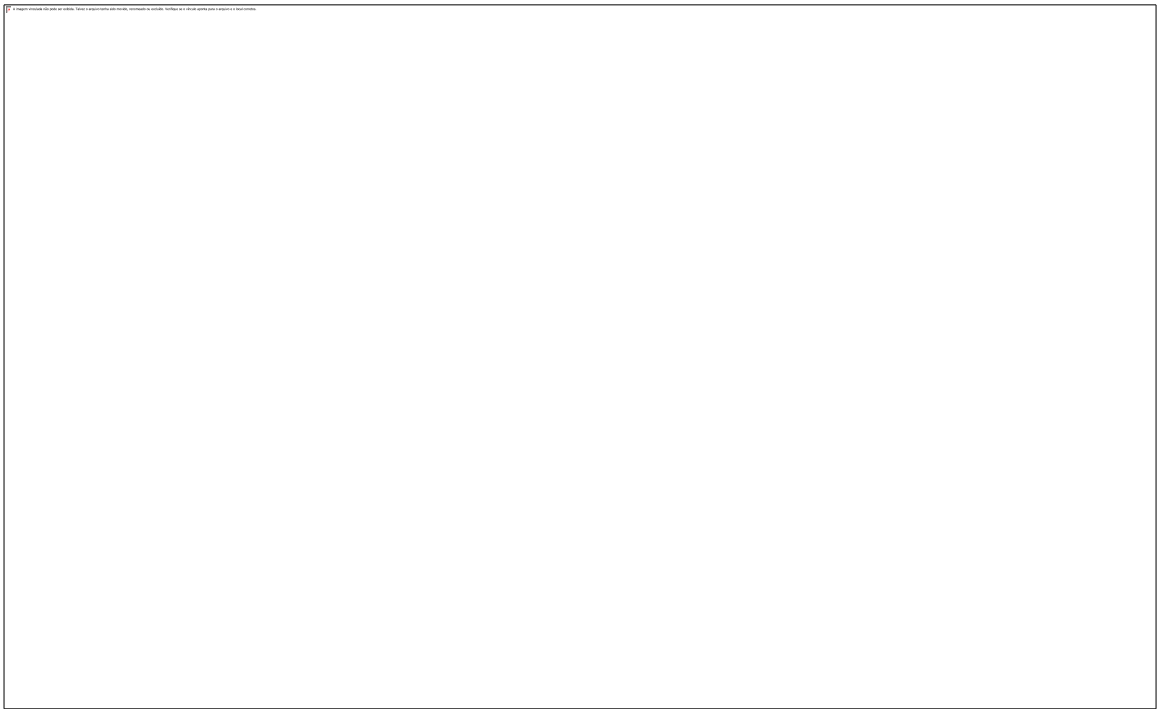
LIBÂNEO, José Carlos. Adeus professor, adeus professora?. Cortez editora, 2014. Disponível em: https://books.google.com/books?hl=pt-BR&lr=&id=BOK_AwAAQBAJ&oi=fnd&pg=PT6&dq=LIB%C3%82NEO,+Jos%C3%A9+Carlos.+Adeus+professor,+adeus+professora%3F.+Cortez+editora,+2014.&ots=Dun-tb1KEx&sig=aT5C61p4dl8CXAHlG7VvAZPhE1o. Acesso em 30 de jul. de 2026.

MAIA, Dennys et al. A abordagem STEAM como proposta pedagógica interdisciplinar para aprendizagem matemática. Revista Ensino Em Debate, v. 2, p. e2024016-e2024016, 2024. https://revistarede.ifce.edu.br/ojs/index.php/rede/article/download/48/37 . Acesso em 30 de jul. de 2026.

FREIRE, Paulo. Pedagogia da autonomia: saberes necessários à prática educativa. Editora Paz e terra, 2014. Disponível em: https://books.google.com/books?hl=pt-BR&lr=&id=Ae4nAwAAQBAJ&oi=fnd&pg=PT3&dq=FREIRE,+Paulo.+Pedagogia+da+autonomia:+saberes+necess%C3%A1rios+%C3%A0+pr%C3%A1tica+educativa.+Editora+Paz+e+terra,+2014.&ots=MYcy5BXtcn&sig=ry4Qlc-5Jb0rnEuzXb-0RnV6IZo. Acesso em 30 de jul. de 2026.

SCHNEIDER, Eduarda Maria; FUJII, Rosangela Araujo Xavier; CORAZZA, Maria Júlia. Pesquisas quali-quantitativas: contribuições para a pesquisa em ensino de ciências. Revista Pesquisa Qualitativa, v. 5, n. 9, p. 569-584, 2017. Acesso em 30 de jul. de 2026.

Anexos



Quadro 1 - Plataformas selecionadas

CÓDIGO: HS1501

AUTOR: EMANUEL KYWAL PINTO CABRAL FILHO

ORIENTADOR: DENNYS LEITE MAIA

TÍTULO: Evolução do Observatório Rede STEAM Potiguar: da prototipação à implementação de uma plataforma integrada para práticas investigativas e criativas

Resumo

Este trabalho apresenta a evolução do Observatório Rede STEAM Potiguar da prototipação para a implementação de uma plataforma web integrada destinada à circulação de práticas investigativas e criativas no Rio Grande do Norte. A pesquisa caracteriza-se como desenvolvimento tecnológico e foi conduzida de forma iterativa, abrangendo consolidação de requisitos, implementação da interface, integração entre módulos e verificação funcional. O sistema resultante reúne três eixos: uma rede para interação e compartilhamento de ações, um catálogo pesquisável de recursos educacionais e um observatório da rede STEAM Potiguar baseado em uma ferramenta de busca integrada ao Potimap, uma ferramenta de interação com o mapa do Rio Grande do Norte, suas mesorregiões, microrregiões e municípios. As telas implementadas evidenciam navegação unificada, busca e filtragem de conteúdos, seleção territorial e associação entre recortes do mapa e itens da plataforma. Como resultado, o projeto avança de uma representação conceitual para um artefato capaz de organizar e relacionar pessoas, iniciativas, recursos e territórios vinculados ao ecossistema STEAM potiguar. Como próximos passos, destacam-se validação com usuários, ampliação da base de dados e avaliações de acessibilidade e usabilidade.

Palavras-chave: STEAM. Observatório. Software educacional. Rede. Potiguar.

TITLE: Evolution of the Observatório Rede STEAM Potiguar: from prototyping to the implementation of an integrated platform for investigative and creative practices

Abstract

This paper presents the evolution of the Observatório Rede STEAM Potiguar from prototyping to the implementation of an integrated web platform designed to support the circulation of investigative and creative practices in Rio Grande do Norte. The research is characterized as technological development and was conducted iteratively, encompassing requirements consolidation, interface implementation, module integration, and functional verification. The resulting system is structured around three main components: a network for interaction and action sharing, a searchable catalog of educational resources, and an observatory of the Rede STEAM Potiguar based on a search tool integrated with Potimap, an interactive tool for exploring the map of Rio Grande do Norte, including its mesoregions, microregions, and municipalities. The implemented screens demonstrate unified navigation, content search and filtering, territorial selection, and the association between map areas and platform items. As a result, the project advances from a conceptual representation to a software artifact capable of organizing and relating people, initiatives, resources, and territories connected to the Potiguar STEAM ecosystem. Future work includes user validation, expansion of the database, and accessibility and usability assessments.

Keywords: STEAM. Observatory. Educational software. Network. Potiguar.

Introdução

A abordagem STEAM articula Ciência, Tecnologia, Engenharia, Artes e Matemática em práticas que favorecem investigação, criação, colaboração e resolução de problemas. Mais do que aproximar áreas do conhecimento, sua adoção pressupõe ambientes em que experiências possam ser registradas, compartilhadas e reinterpretadas por diferentes atores. Essa perspectiva é coerente com a valorização de competências relacionadas ao pensamento crítico, à criatividade, à comunicação e à colaboração, discutidas por Fullan e Langworthy (2014), e com a orientação da Base Nacional Comum Curricular para práticas formativas que mobilizem conhecimentos em situações contextualizadas (BRASIL, 2018).

No Rio Grande do Norte, estudos sobre experiências STEAM evidenciam uma produção relevante, mas distribuída de forma desigual pelo território e nem sempre conhecida pela própria comunidade educacional. Maia et al. (2024a) identificam participação expressiva de projetos potiguares em feiras científicas, enquanto Maia et al. (2024b) destacam o potencial da abordagem para aprendizagens interdisciplinares. Ao mesmo tempo, a literatura aponta obstáculos relacionados à formação docente, aos recursos disponíveis e à articulação entre sujeitos e instituições (MENDES et al., 2025). Nesse cenário, o problema não se limita à existência de boas práticas: envolve também a capacidade de fazê-las circular, conectar pessoas com diferentes expertises e tornar visível sua distribuição no território.

O Observatório Rede STEAM Potiguar foi concebido para responder a esse problema por meio de três frentes complementares: uma rede de interação entre participantes, um repositório de recursos educacionais e um espaço de observação territorial de dados relacionados à STEAM. Em uma etapa anterior, essas frentes foram apresentadas principalmente como proposta e protótipo de interface. Paralelamente, o Potimap foi desenvolvido como uma ferramenta cartográfica capaz de representar o Rio Grande do

Norte em diferentes recortes territoriais e associar dados aos elementos do mapa. Esses resultados estabeleceram as bases conceituais e técnicas do projeto, mas ainda permanecia o desafio de transformar componentes antes mostrados de forma separada em uma aplicação coesa e navegável.

Assim, o presente trabalho trata-se da implementação concreta do esboço funcional previamente definido, acompanhada da integração entre módulos que, até então, eram tratados como partes da proposta ou como ferramentas autônomas. A questão orientadora passa, portanto, a ser: como materializar o Observatório Rede STEAM Potiguar como uma plataforma integrada capaz de conectar colaboração, recursos educacionais e exploração territorial em um mesmo fluxo de uso?

Assim, o objetivo deste trabalho é descrever e analisar a evolução do Observatório Rede STEAM Potiguar da prototipação à implementação de uma plataforma web integrada. Especificamente, busca-se consolidar os requisitos dos três eixos do sistema, implementar uma interface comum e coerente entre eles, integrar o Potimap ao módulo territorial e analisar os resultados funcionais observáveis nas telas produzidas.

Metodologia

O trabalho caracteriza-se como pesquisa de desenvolvimento tecnológico, por ter como objeto principal a construção e a evolução de um artefato de software. A unidade de análise foi o próprio Observatório Rede STEAM Potiguar, considerando tanto os requisitos definidos em sua etapa de concepção quanto os componentes técnicos já desenvolvidos anteriormente. O processo foi conduzido de forma iterativa, seguindo a organização ágil adotada pelo projeto e priorizando entregas incrementais que pudessem ser verificadas durante a implementação.

A primeira etapa consistiu na consolidação dos requisitos funcionais. Em vez de tratar cada tela como uma peça isolada, os requisitos foram reorganizados em três eixos de uso: (i) Rede STEAM Potiguar, voltada à interação entre pessoas, divulgação de ações e estabelecimento de conexões; (ii) Recursos Educacionais, destinado à localização e consulta de materiais relacionados a práticas STEAM; e (iii) Observatório, responsável pela exploração territorial de pessoas, instituições, iniciativas, eventos e recursos. Essa organização permitiu preservar a proposta original do sistema e, ao mesmo tempo, estabelecer relações entre funcionalidades que anteriormente apareciam apenas no nível do protótipo.

Na segunda etapa, foi implementada a estrutura visual e de navegação da aplicação. Os protótipos produzidos no Figma foram utilizados como referência de composição, mas a atividade central passou a ser sua transformação em componentes funcionais de interface. A aplicação foi desenvolvida com React e TypeScript no frontend, Java com Spring Boot no backend e PostgreSQL para persistência dos dados. A navegação superior foi mantida como elemento comum às áreas da Rede, dos Recursos Educacionais, do Observatório e da página institucional, de forma a reduzir a percepção

de módulos desconectados. O desenvolvimento utilizou a organização de tarefas em ciclos curtos e controle de versões previsto pelo fluxo de desenvolvimento do projeto.

A terceira etapa concentrou-se na implementação dos comportamentos específicos de cada eixo. Na Rede, foram estruturados elementos de perfil, publicações, ações em projetos, convites de colaboração, conexões e sugestões de contato. No catálogo de Recursos Educacionais, foram implementados mecanismos de busca, filtros, classificação visual por tipo de material e áreas STEAM, além de paginação. No Observatório, a pesquisa textual e os filtros foram combinados à exploração do território, permitindo que o usuário altere o nível de recorte e relacione a seleção espacial aos resultados apresentados na mesma página.

A quarta etapa correspondeu à integração do Potimap ao Observatório. A ferramenta havia sido desenvolvida a partir da manipulação dos elementos de um mapa vetorial em SVG a partir do framework React com a linguagem TypeScript, possibilitando a gestão de municípios, microrregiões e mesorregiões, seleção de áreas e associação de informações por identificadores territoriais. Na versão integrada, o mapa deixa de funcionar apenas como demonstração cartográfica e passa a compor o fluxo de consulta do Observatório. A interface organiza opções para visualização do estado, mesorregiões, microrregiões e municípios e apresenta, ao lado do mapa, informações sobre o recorte selecionado. A seleção territorial pode ser relacionada aos itens retornados pela consulta, transformando o mapa em mecanismo de navegação pelos conteúdos da plataforma.

Por fim, realizou-se uma verificação técnico-funcional por cenários de navegação, observando a presença, a continuidade visual e a articulação das funcionalidades implementadas. Foram considerados como evidências os registros de tela da versão atual do sistema: página da Rede, catálogo de Recursos Educacionais, tela principal do Observatório, exploração do Potimap, seleção de município com resultados associados e página de apresentação do ecossistema. A análise foi qualitativa e centrada no artefato; portanto, não corresponde a um teste de usabilidade com participantes nem permite, nesta etapa, inferir impacto educacional, adesão dos usuários ou conformidade integral com critérios de acessibilidade.

Resultados e Discussões

Os resultados demonstram uma mudança de estágio do Observatório Rede STEAM Potiguar. Se a fase anterior permitia comunicar como a solução deveria funcionar, a versão atual oferece uma estrutura implementada na qual os principais eixos podem ser percorridos dentro da mesma aplicação. O resultado mais relevante, portanto, não é apenas a existência de novas telas, mas a passagem de uma representação de funcionalidades para uma composição integrada em que identidade visual, navegação, tipos de conteúdo e exploração territorial compartilham o mesmo contexto de uso.

A tela da Rede STEAM Potiguar evidencia essa materialização (Figura 1). O ambiente apresenta um perfil do participante, indicadores de conexões, ações, projetos e

instituições, campo para publicação e um fluxo de postagens com mecanismos de interação. Também são apresentados recursos para registrar uma ação em projeto, convidar colaboradores, adicionar mídia e encontrar novas conexões. Essa composição traduz para o software uma necessidade central da proposta: não apenas expor projetos, mas criar condições para que pessoas com conhecimentos distintos possam se localizar e interagir em torno de práticas investigativas e criativas. Em relação ao estágio conceitual anterior, o avanço está na articulação desses elementos em um fluxo semelhante ao de uma comunidade digital, no qual produção, interação e descoberta de pessoas coexistem.

A área de Recursos Educacionais constitui o segundo resultado da implementação (Figura 2). A tela apresenta busca textual, filtros, indicação da quantidade de itens localizados, paginação e cartões que organizam os materiais por título, descrição, origem, tipo e áreas STEAM. Essa estrutura é relevante porque transforma a ideia genérica de "repositório" em uma experiência de consulta. Em vez de concentrar os materiais em uma listagem indiferenciada, a interface oferece atributos que ajudam o usuário a compreender rapidamente a natureza de cada recurso e sua relação com áreas como Ciência, Tecnologia, Engenharia, Artes e Matemática. Com isso, a plataforma passa a exercer também uma função de organização e recuperação de informação, necessária para que a produção local possa ser utilizada como referência em novas práticas.

O terceiro resultado, e o que mais explicita a integração ao Potimap, encontra-se no Observatório territorial (Figuras 3 e 4). A interface combina apresentação do propósito da área, marcadores das cinco áreas STEAM, campo de busca, filtros e o Potimap incorporada diretamente ao fluxo da página. O usuário pode alternar entre uma visão geral do Rio Grande do Norte e recortes por mesorregiões, microrregiões e municípios. O mapa ocupa a região central da consulta, enquanto um painel lateral informa o território selecionado. Dessa forma, uma ferramenta que anteriormente havia sido validada como protótipo cartográfico passa a assumir uma função dentro da arquitetura informacional da plataforma.

Os registros de tela mostram ainda que a integração não termina na seleção visual do território. Em um dos cenários (Figura 5), o município de Assú aparece destacado no mapa e a área de resultados retorna dois cartões associados ao recorte selecionado: um evento e uma iniciativa, ambos acompanhados de descrições e classificações por áreas STEAM. O conteúdo utilizado na versão registrada tem caráter demonstrativo, mas o comportamento da interface evidencia uma mudança importante de função: o território deixa de ser apenas uma camada de visualização e passa a atuar como filtro e contexto para a descoberta de conteúdos. Essa relação entre mapa e itens da plataforma concretiza um dos objetivos centrais do Observatório, que é permitir observar onde pessoas, ações e recursos se distribuem e, a partir disso, explorar conexões possíveis.

A página institucional, apresentada em outro registro de tela (Figura 6), complementa a integração ao explicitar o ecossistema que sustenta a proposta. Nela, iniciativas como o GIIFE (Grupo Interdisciplinar de Estudos e Pesquisas em Informática na Educação), o IMD (Instituto Metrópole Digital), o projeto OBAMA (Objetos de Aprendizagem para Matemática) e o laboratório Esteamulab são apresentados como partes de um percurso que articula pesquisa, formação, desenvolvimento tecnológico e prática pedagógica.

Embora essa página não seja o núcleo funcional do sistema, ela desempenha papel relevante na comunicação do Observatório: situa a plataforma em uma rede institucional e ajuda a mostrar que o artefato não pretende funcionar como repositório isolado, mas como ponto de articulação entre ações já existentes.

A análise conjunta das telas também evidencia maior coerência de interação entre as diferentes áreas. Os quatro espaços principais compartilham cabeçalho, identidade visual e padrões de componentes, enquanto busca, filtros, cartões e marcadores temáticos reaparecem em diferentes contextos. Essa repetição controlada reduz a ruptura entre módulos e favorece a percepção de uma única plataforma.

Do ponto de vista pedagógico, a implementação cria condições técnicas para aproximar três ações que a literatura trata como relevantes para práticas STEAM: colaboração, acesso a referências e contextualização em problemas e territórios. Fullan e Langworthy (2014) enfatizam competências como colaboração, comunicação, criatividade e pensamento crítico; Maia et al. (2024b) discute a abordagem STEAM como proposta interdisciplinar; e Mendes et al. (2025) aponta desafios de articulação e disponibilidade de recursos para sua implementação. O Observatório não resolve, por si só, esses desafios, mas materializa uma infraestrutura digital na qual pessoas, materiais e território podem ser relacionados. Essa distinção é importante: o resultado alcançado é tecnológico e funcional, enquanto seus efeitos educacionais dependem de adoção, curadoria de conteúdo e usos concretos pela comunidade.

Também é necessário diferenciar implementação de validação de impacto. As telas demonstram que os fluxos e componentes previstos foram materializados, mas não constituem evidência suficiente para afirmar melhora de aprendizagem, ampliação de colaboração ou aumento da disseminação de práticas STEAM. Da mesma forma, a evolução visual não comprova, isoladamente, qualidade de usabilidade ou acessibilidade. Esses aspectos exigem métodos específicos de avaliação com usuários e critérios técnicos próprios. Assim, os resultados deste trabalho devem ser lidos como evidência de maturação do artefato: o projeto avança de uma proposta e de protótipos parciais para uma plataforma integrada que já permite observar, de forma concreta, como seus diferentes eixos se relacionam.

Por fim, a nova versão cria uma base mais adequada para as próximas etapas de pesquisa. Com a estrutura integrada, torna-se possível planejar avaliações que não seriam significativas em um protótipo fragmentado, como testes de tarefas envolvendo passagem entre Rede, Recursos e Observatório; análise da compreensão dos filtros territoriais; avaliação da descoberta de colaboradores; estudos de acessibilidade; e investigação da qualidade da curadoria dos recursos. Também se torna necessário substituir progressivamente os conteúdos demonstrativos por dados consolidados do ecossistema STEAM potiguar. Desse modo, a principal contribuição da evolução apresentada está em transformar uma ideia de observatório em um objeto de software suficientemente articulado para ser submetido a avaliações de uso e, posteriormente, analisado quanto a seus efeitos na circulação de práticas investigativas e criativas no Rio Grande do Norte.

Conclusão

O desenvolvimento apresentado marca a transição do Observatório Rede STEAM Potiguar de uma etapa predominantemente conceitual e prototipada para uma plataforma web integrada. A contribuição central não está em propor novamente as funcionalidades do projeto, mas em demonstrar sua materialização e a articulação entre componentes que antes eram apresentados como frentes independentes. A Rede STEAM, o catálogo de Recursos Educacionais e o Observatório passam a compartilhar uma mesma estrutura de navegação e a operar como partes de uma experiência contínua.

Entre os resultados, destacam-se a implementação de um ambiente de interação entre participantes, a organização pesquisável de materiais educacionais e, principalmente, a incorporação do Potimap ao fluxo de exploração do Observatório. A seleção territorial, associada à apresentação de resultados relacionados ao recorte escolhido, amplia o papel do mapa: de ferramenta cartográfica autônoma para mecanismo de navegação e contextualização dos conteúdos da plataforma. Os registros de tela da versão atual fornecem evidências dessa maturação funcional e mostram que o esboço apresentado em fases anteriores ganhou forma como sistema de software.

Ao mesmo tempo, os resultados delimitam com clareza o estágio da pesquisa. A existência de uma aplicação funcional não autoriza concluir que seus impactos pedagógicos ou colaborativos já foram alcançados. Avaliações com usuários, estudos de usabilidade e acessibilidade, testes sistemáticos de desempenho e segurança, além da ampliação e curadoria da base de dados, constituem etapas necessárias para verificar a qualidade e a efetividade do Observatório em contexto real.

Conclui-se, portanto, que a principal evolução do projeto está na integração: pessoas, recursos e território deixam de ser elementos apenas descritos na proposta e passam a coexistir em um artefato capaz de relacioná-los. Essa base possibilita que trabalhos futuros avancem da pergunta “como o Observatório deve funcionar?” para questões sobre “como ele é utilizado, compreendido e apropriado pela comunidade?”. Essa mudança de foco representa o amadurecimento tecnológico necessário para investigar, em etapas posteriores, o potencial do Observatório Rede STEAM Potiguar como infraestrutura de apoio à circulação e ao fortalecimento de práticas investigativas e criativas no estado.

Referências

BRASIL. Ministério da Educação. Base Nacional Comum Curricular. Brasília: MEC, 2018.

FULLAN, M.; LANGWORTHY, M. A rich seam: how new pedagogies find deep learning. London: Pearson, 2014.

MAIA, D. et al. Abordagem STEAM no Rio Grande do Norte: uma análise longitudinal a partir de experiências publicadas na FEBRACE. Revista Brasileira de Ensino de Ciência e Tecnologia, v. 17, p. 1–20, 2024a.

MAIA, D. et al. A abordagem STEAM como proposta pedagógica interdisciplinar para aprendizagem matemática. Revista Ensino em Debate, v. 2, e2024016, 2024b.

MENDES, A. A. et al. Mapeamento da abordagem STEAM na educação básica: analisando impactos e desafios atuais. Ensino de Ciências e Tecnologia em Revista – ENCITEC, v. 15, n. 2, p. 305–319, 2025.

Anexos

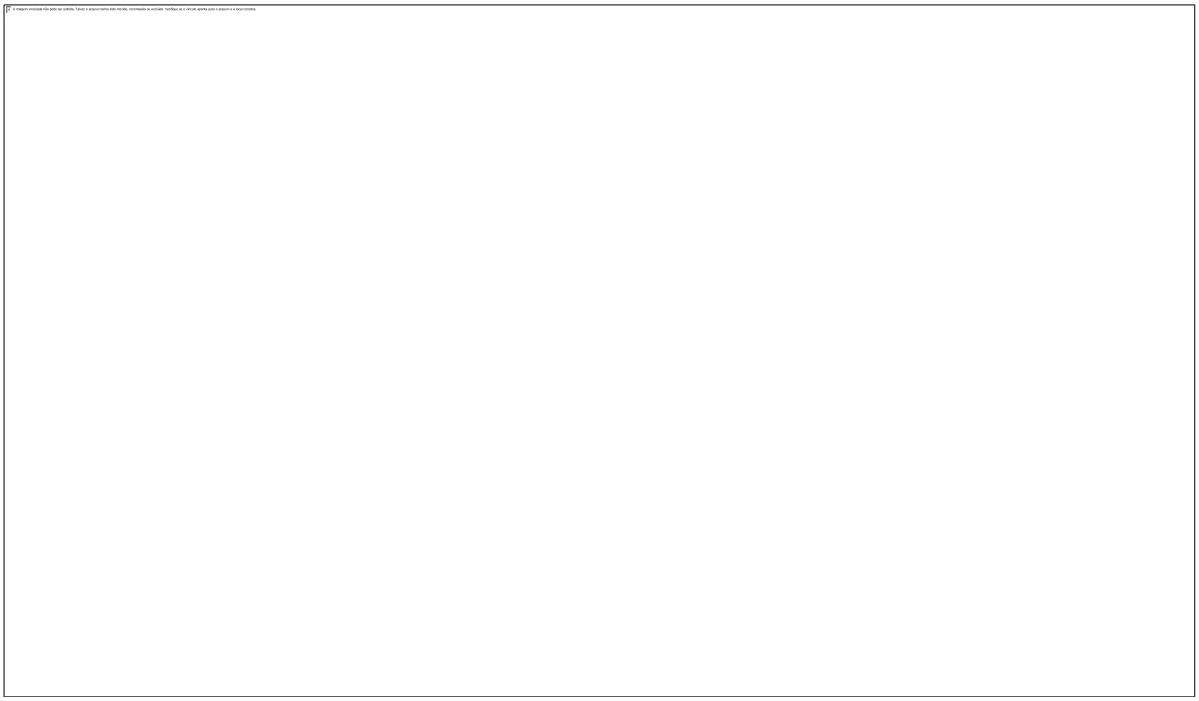


Figura 1

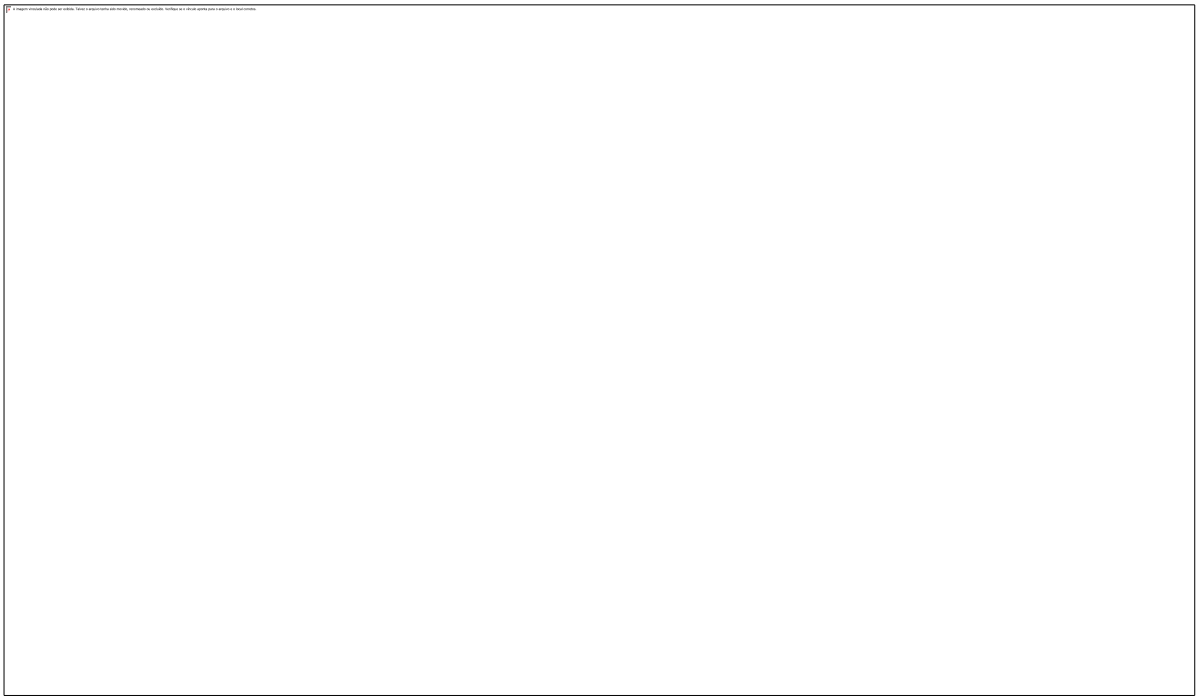


Figura 2

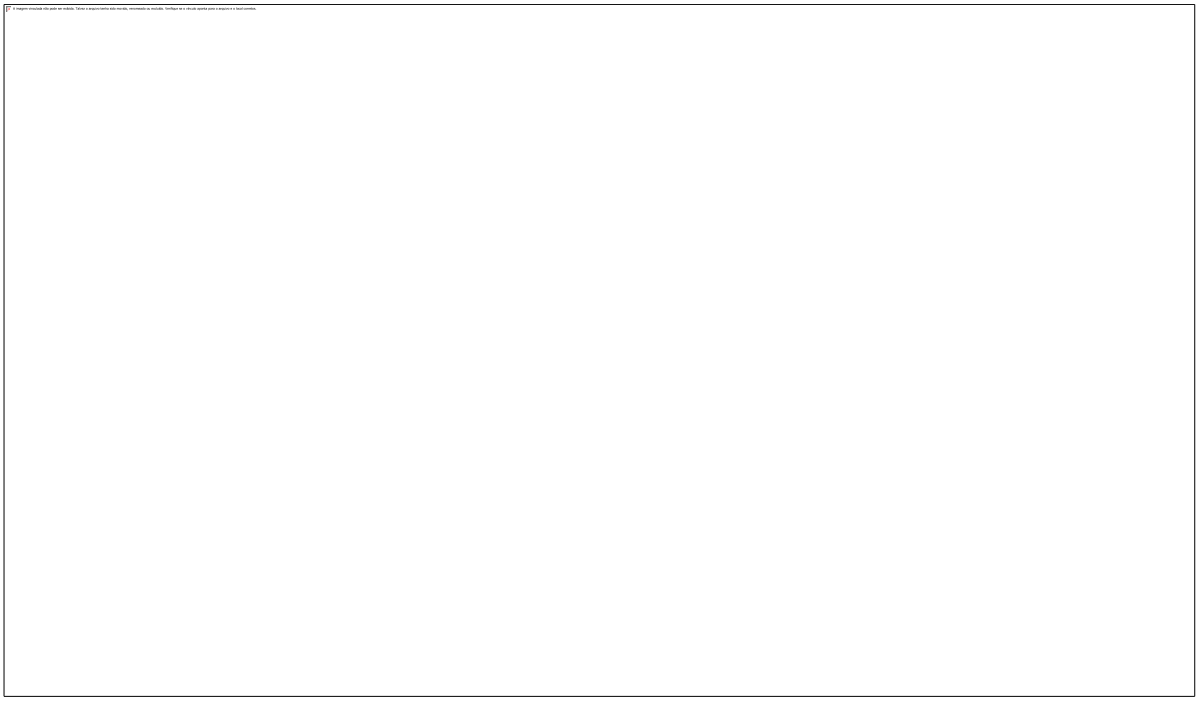


Figura 3



Figura 4

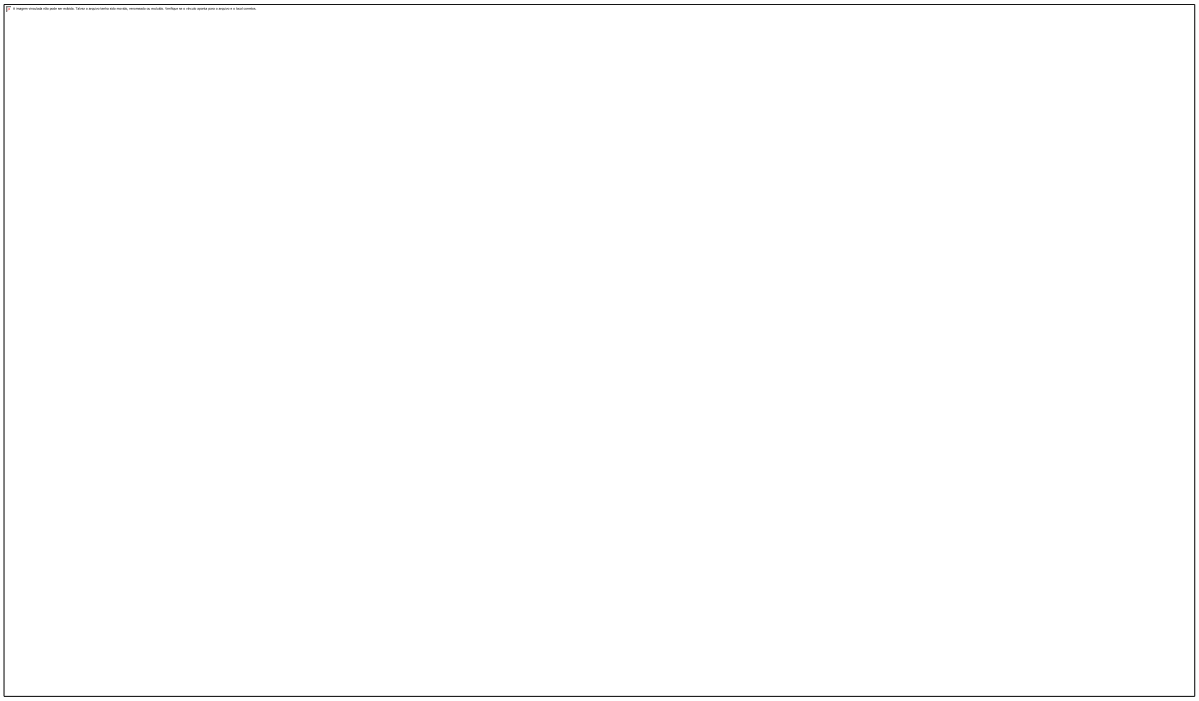


Figura 5



Figura 6

CÓDIGO: SB0137

AUTOR: LUIZ FERNANDO DIAS DE MELO

COAUTOR: GEAN CARLOS DA SILVA SANTOS

COAUTOR: ANTONINO ALVES FEITOSA NETO

ORIENTADOR: MARCIO DIAS PEREIRA

TÍTULO: Índices de desempenho de plântulas obtidos por mensuração manual discriminam o vigor de lotes de sementes de Mimosa caesalpiniifolia

Resumo

A avaliação do vigor de sementes por meio de índices de desempenho de plântulas tem sido tradicionalmente associada ao uso de sistemas automatizados de análise de imagens, limitando sua aplicação em laboratórios sem acesso a essas tecnologias. Neste estudo, avaliou-se a eficiência de índices obtidos a partir de mensurações manuais de plântulas na discriminação da qualidade fisiológica entre lotes de sementes de Mimosa caesalpiniifolia Benth. Sete lotes provenientes de diferentes matrizes localizadas na Fazenda Escola Jundiaí, UFRN, em Macaíba-RN, foram submetidos aos testes de germinação, primeira contagem, índice de velocidade de germinação, envelhecimento acelerado, emergência, índice de velocidade de emergência e massa seca de plântulas. Aos 5, 7 e 10 dias após a semeadura foram determinados os comprimentos da raiz primária, da parte aérea e total das plântulas, utilizados para o cálculo dos índices de crescimento, uniformidade, vigor e vigor corrigido. Os testes de vigor evidenciaram diferenças significativas entre os lotes, enquanto a germinação apresentou menor capacidade discriminatória. Os índices obtidos a partir das mensurações manuais apresentaram elevada associação com os testes convencionais de vigor, especialmente aos 10 dias após a semeadura. As análises de correlação e de componentes principais confirmaram a capacidade dos índices em discriminar os níveis de vigor entre os lotes, destacando os índices de crescimento, vigor e vigor corrigido como as variáveis m

Palavras-chave: análise de sementes; fenotipagem de plântulas; qualidade fisiológica;

TITLE: Seedling performance indices obtained by manual measurement discriminate the vigor of *Mimosa caesalpiniifolia* seed lots.

Abstract

associated with automated image analysis systems, limiting its application in laboratories lacking access to such technologies. This study evaluated the efficiency of indices Seed vigor assessment using seedling performance indices has traditionally been derived from manual seedling measurements in discriminating physiological quality among seed lots of Mimosa caesalpiniifolia Benth. Seven seed lots collected from different mother trees located at the Jundiaí Experimental Farm, Federal University of Rio Grande do Norte (UFRN), in Macaíba, Brazil, were subjected to germination, first count, germination speed index, accelerated aging, seedling emergence, emergence speed index, and seedling dry mass tests. At 5, 7, and 10 days after sowing, primary root length, shoot length, and total seedling length were measured and used to calculate growth, uniformity, vigor, and corrected vigor indices. Vigor tests revealed significant differences among lots, whereas germination showed lower discriminatory capacity. Indices derived from manual measurements exhibited strong associations with conventional vigor tests, particularly at 10 days after sowing. Correlation and principal component analyses confirmed the ability of these indices to discriminate vigor levels among lots, highlighting the growth index, vigor index, and corrected vigor index as the most sensitive variables. The results demonstrate that indices derived from manual seedling measurements constitute an accessible and efficien

Keywords: forest seeds; physiological quality; seed vigor; seedling phenotyping;

Introdução

A avaliação do vigor de sementes é fundamental para complementar as informações obtidas no teste de germinação e aumentar a capacidade de predição do desempenho das plântulas em viveiros e condições de campo (MARCOS-FILHO, 2015). Enquanto a germinação estima o potencial máximo de produção de plântulas normais sob condições favoráveis, o vigor representa o conjunto de atributos responsáveis pela rapidez, uniformidade da emergência e tolerância das sementes a condições ambientais menos favoráveis (AOSA, 2009; ISTA, 2024). Em espécies florestais, a correta estimativa do vigor é particularmente importante devido ao elevado custo de produção de mudas e à necessidade de obtenção de povoamentos uniformes e bem estabelecidos.

Entre as metodologias mais promissoras para avaliação do vigor destacam-se aquelas baseadas no crescimento inicial das plântulas, uma vez que o desenvolvimento inicial integra diversos processos fisiológicos relacionados à mobilização de reservas, reorganização de membranas e atividade metabólica das sementes (HOFFMASTER et al., 2003; MARCOS-FILHO, 2015). Nesse contexto, os índices de crescimento (GI), uniformidade (UI) e vigor (VI), propostos por Sako et al. (2001), bem como o índice de vigor corrigido (CVI), desenvolvido por Medeiros e Pereira (2018), têm demonstrado elevada sensibilidade para detectar diferenças na qualidade fisiológica entre lotes de sementes.

Esses índices foram originalmente desenvolvidos para utilização em sistemas automatizados de análise de imagens, como o SVIS®, SAPL® e Vigor-S, que permitem avaliações rápidas, objetivas e altamente reprodutíveis do desempenho das plântulas (SAKO et al., 2001; MEDEIROS; PEREIRA, 2018; CASTAN et al., 2018). Entretanto, a necessidade de equipamentos específicos, softwares especializados e infraestrutura

computacional ainda limita a utilização dessas ferramentas em muitos laboratórios de análise de sementes, especialmente em programas regionais de produção de sementes florestais e em instituições com menor disponibilidade de recursos tecnológicos (MEDEIROS et al., 2020).

Por outro lado, a mensuração manual do comprimento das plântulas utilizando régua milimetrada constitui um procedimento simples, de baixo custo e amplamente empregado em testes tradicionais de vigor (KRZYZANOWSKI et al., 2020). Apesar disso, ainda são escassos os estudos que validam a utilização dessas mensurações para o cálculo dos índices originalmente desenvolvidos para sistemas automatizados, particularmente em espécies florestais nativas brasileiras, configurando uma importante lacuna do conhecimento na área de tecnologia de sementes.

Entre essas espécies destaca-se Mimosa caesalpiniifolia Benth., conhecida popularmente como sabiá, amplamente utilizada em programas de recuperação de áreas degradadas, sistemas agroflorestais e produção de madeira, devido ao rápido crescimento, rusticidade e elevada adaptação às condições do semiárido brasileiro (CARVALHO, 2007; MAIA, 2012; LORENZI, 2020). Apesar de sua importância econômica e ecológica, ainda são limitadas as informações sobre a aplicação de índices de desempenho de plântulas na discriminação da qualidade fisiológica entre lotes dessa espécie.

Partiu-se da hipótese de que índices de desempenho de plântulas obtidos a partir de mensurações manuais apresentariam forte associação com os testes convencionais de vigor e seriam capazes de discriminar eficientemente diferenças de qualidade fisiológica entre lotes de sementes de Mimosa caesalpiniifolia. Assim, o objetivo deste estudo foi avaliar a eficiência de índices de desempenho de plântulas obtidos por mensurações manuais do comprimento de plântulas na discriminação de níveis de vigor entre diferentes lotes da espécie, bem como verificar seu potencial de utilização como ferramenta acessível para programas de controle de qualidade de sementes florestais.

Metodologia

Origem e caracterização dos lotes de sementes

Foram utilizados sete lotes de sementes de Mimosa caesalpiniifolia Benth. provenientes de matrizes distintas localizadas na Fazenda Escola Jundiaí da Escola Agrícola de Jundiaí da Universidade Federal do Rio Grande do Norte (EAJ/UFRN), no município de Macaíba, Rio Grande do Norte, Brasil (5°53’44’’ S; 35°21’52’’ W). As coletas foram realizadas durante o período de maturação fisiológica dos frutos, entre os meses de setembro e novembro de 2025.

Cada lote foi constituído por sementes coletadas individualmente de uma única árvore matriz, previamente selecionada com base em critérios fitossanitários, vigor vegetativo e produção de sementes, totalizando sete lotes independentes (Lote 1 a Lote 7). Após a coleta, os frutos foram acondicionados em sacos de papel e transportados ao Laboratório Inteligente de Sementes e Bioinovação (GETSem/EAJ-UFRN), onde

permaneceram em ambiente ventilado até a abertura natural das vagens e extração das sementes.

Após o beneficiamento manual, as sementes foram submetidas à limpeza e padronização, sendo eliminadas aquelas com danos mecânicos, deformações, sintomas de ataque de insetos ou fungos e diferenças acentuadas de tamanho e coloração. Os lotes foram armazenados separadamente em embalagens de papel Kraft sob condições refrigeradas (5 ± 2 °C e aproximadamente 60% de umidade relativa do ar) até o início das avaliações.

O teor de água inicial das sementes foi determinado pelo método da estufa a 105 ± 3 °C durante 24 horas, utilizando-se duas repetições de aproximadamente 4,5 g de sementes inteiras, conforme as Regras para Análise de Sementes (BRASIL, 2009).

Teste de germinação e índice de velocidade de germinação

O teste de germinação foi conduzido utilizando-se quatro repetições de 50 sementes por lote, distribuídas em rolos de papel Germitest® previamente umedecidos com água destilada na quantidade equivalente a 2,5 vezes a massa seca do substrato.

As sementes foram mantidas em germinador tipo BOD regulado a 25 °C e fotoperíodo de 12 horas, conforme recomendações para espécies florestais da família Fabaceae (BRASIL, 2013). As avaliações foram realizadas diariamente até o décimo dia após a semeadura, sendo consideradas germinadas as sementes que originaram plântulas normais.

Os resultados foram expressos em porcentagem de germinação (G)

Testes de vigor

O índice de velocidade de germinação (IVG) foi calculado de acordo com a metodologia proposta por Maguire (1962), a partir da contagem diária do teste de germinação:

$$IVG = (G_1/N_1) + (G_2/N_2) + \ldots + (G_n/N_n)$$

onde: G_1, G_2, …, G_n = número de sementes germinadas em cada contagem; N_1, N_2, …, N_n = número de dias decorridos da semeadura até cada contagem.

A emergência de plântulas foi conduzida em bandejas de poliestireno contendo areia lavada e esterilizada como substrato, utilizando-se quatro repetições de 50 sementes por lote.

As bandejas foram mantidas em casa de vegetação sob irrigação diária, suficiente para manutenção da umidade próxima à capacidade de campo do substrato. As avaliações foram realizadas diariamente durante dez dias, sendo consideradas emergidas as plântulas que apresentavam os cotilédones completamente expandidos acima da superfície do substrato.

A porcentagem de emergência (E) foi expressa em percentual e o índice de velocidade de emergência (IVE) foi calculado conforme Maguire (1962):

$$IVE = (E_1/N_1) + (E_2/N_2) + \ldots + (E_n/N_n)$$

onde: E_1, E_2, …, E_n = número de plântulas emergidas em cada contagem; N_1, N_2, …, N_n = número de dias decorridos da semeadura até cada avaliação.

O teste de envelhecimento acelerado foi conduzido utilizando-se 200 sementes por lote distribuídas sobre telas metálicas em caixas do tipo Gerbox®, contendo 40 mL de água destilada no fundo das caixas, permanecendo em câmara BOD a 41 °C e aproximadamente 100% de umidade relativa por 24 horas. Após esse período, as sementes foram submetidas ao teste de germinação, sendo os resultados expressos em porcentagem de plântulas normais.

A massa seca das plântulas foi determinada após a remoção dos cotilédones, sendo as amostras acondicionadas em sacos de papel e mantidas em estufa com circulação forçada de ar a 65 °C até massa constante, com resultados expressos em mg plântula^{-1}.

Avaliação do crescimento de plântulas e obtenção dos índices de desempenho

Para a avaliação do crescimento inicial, foi instalado um novo teste de germinação utilizando-se quatro repetições de 50 sementes para cada lote, conforme metodologia descrita por Krzyzanowski et al. (2020).

As avaliações foram realizadas aos 5, 7 e 10 dias após a semeadura, utilizando-se apenas plântulas normais. Em cada época de avaliação foram mensurados, com auxílio de régua milimetrada, o comprimento da raiz primária (RP), o comprimento da parte aérea (PA) e o comprimento total da plântula (PT), sendo os resultados expressos em centímetros por plântula.

Todas as mensurações foram realizadas pelo mesmo avaliador, buscando minimizar possíveis variações associadas ao operador e aumentar a repetibilidade das medidas.

Com base nos dados de crescimento foram calculados o índice de crescimento (IC), o índice de uniformidade (IU), o índice de vigor (IV) e o índice de vigor corrigido (IVC), conforme as metodologias propostas por Sako et al. (2001) e Medeiros e Pereira (2018).

Os cálculos foram realizados em planilha eletrônica desenvolvida no Microsoft Excel® (Microsoft Corporation, versão 365), utilizando as seguintes expressões:

Delineamento experimental

O experimento foi conduzido em delineamento inteiramente casualizado, constituído por sete tratamentos correspondentes aos sete lotes de sementes avaliados, com quatro repetições de 50 sementes por lote para cada variável analisada.

Cada repetição foi considerada uma unidade experimental independente.

Índice de

Crescimento (IC)

IC = mín. (wh × lh + wr × lr, 1000)

Índice de Uniformidade (IU)

IU = máx. [1000 − (wsh × sh + wsr × sr + wtotal × stotal + wsr/h × sr/h) − wm × n

Resultados e Discussões

. Resultados e Discussão

A análise de variância revelou diferenças significativas entre os sete lotes de sementes de Mimosa caesalpiniifolia para emergência (E), envelhecimento acelerado (EA) e índice de velocidade de emergência (IVE), evidenciando a existência de níveis contrastantes de qualidade fisiológica entre os materiais provenientes de diferentes matrizes da Fazenda Escola Jundiaí da UFRN, em Macaíba-RN (Figura 1). Por outro lado, germinação (G), primeira contagem (PC), índice de velocidade de germinação (IVG) e massa seca de plântulas (MS) não apresentaram diferenças significativas pelo teste F ($p \geq 0,05$), indicando menor sensibilidade dessas variáveis para discriminar o vigor dos lotes avaliados.

Embora os valores de germinação tenham variado entre aproximadamente 82 e 91%, todos os lotes apresentaram desempenho superior ao mínimo exigido para sementes florestais destinadas à produção de mudas, refletindo elevada viabilidade fisiológica do conjunto avaliado. Resultado semelhante foi observado para a primeira contagem de germinação e para a massa seca das plântulas, cujos coeficientes de variação relativamente baixos indicam boa precisão experimental e reduzida variabilidade entre os lotes para essas características (Figura 1). A ausência de diferenças significativas nessas variáveis reforça o entendimento de que o teste de germinação, quando utilizado isoladamente, possui capacidade limitada para identificar diferenças sutis de desempenho fisiológico entre lotes com elevada qualidade inicial (MARCOS-FILHO, 2015; KRZYZANOWSKI et al., 2020).

Em contraste, os testes tradicionalmente associados ao vigor apresentaram elevada sensibilidade para a estratificação fisiológica dos lotes. A emergência variou de aproximadamente 89% no lote 3 para valores próximos de 64% no lote 7, representando redução superior a 28% no potencial de estabelecimento das plântulas. O envelhecimento acelerado apresentou comportamento semelhante, com valores variando de 77% no lote 5 para aproximadamente 42% no lote 7, evidenciando diferenças marcantes na tolerância das sementes ao estresse e ao processo de deterioração. O índice de velocidade de emergência também apresentou redução expressiva entre os lotes de maior e menor vigor, indicando que as diferenças fisiológicas se manifestaram principalmente na velocidade e uniformidade do estabelecimento das plântulas.

Esses resultados corroboram a premissa de que o vigor representa um atributo fisiológico distinto da germinação, estando mais diretamente relacionado à rapidez de emergência, à capacidade de suportar condições adversas e ao desempenho em campo (AOSA, 2009; MARCOS-FILHO, 2015). Em espécies florestais, nas quais o sucesso da produção de mudas depende fortemente da uniformidade do estabelecimento e da capacidade competitiva inicial das plântulas, testes de vigor frequentemente apresentam maior relevância prática que o próprio percentual de germinação (KRZYZANOWSKI et al., 2020).

A caracterização fisiológica dos lotes permitiu identificar um gradiente consistente de qualidade entre os materiais avaliados (Figura 2). Os lotes 2, 3 e 5 apresentaram os maiores valores para emergência, envelhecimento acelerado e velocidade de emergência, caracterizando-se como lotes de alto vigor, enquanto o lote 7 apresentou desempenho inferior em praticamente todas as variáveis avaliadas, sendo classificado como lote de baixo vigor. Os lotes 1, 4 e 6 apresentaram comportamento intermediário, embora o lote 6 tenha apresentado redução mais acentuada em emergência e envelhecimento acelerado, aproximando-se do comportamento fisiológico observado no lote 7.

A ausência de diferenças estatísticas para germinação e primeira contagem, associada às diferenças expressivas observadas para emergência, envelhecimento acelerado e velocidade de emergência, demonstra a importância da utilização conjunta de testes de vigor na caracterização fisiológica de lotes florestais. Essa abordagem permite detectar diferenças que permaneceriam ocultas caso apenas o teste de germinação fosse utilizado, aumentando a precisão na classificação dos lotes e fornecendo estimativas mais realistas do potencial de desempenho em viveiro e em campo (MARCOS-FILHO, 2015; AOSA, 2009).

De maneira geral, os resultados das Figuras 1 e 2 evidenciam que os lotes provenientes de diferentes matrizes apresentaram diferenças reais e consistentes de qualidade fisiológica, justificando sua utilização como modelo experimental para a validação dos índices de desempenho de plântulas obtidos por mensurações manuais. A existência desse gradiente fisiológico entre os lotes constitui requisito fundamental para avaliar a capacidade discriminatória dos índices de crescimento, uniformidade e vigor, discutidos nas seções subsequentes do trabalho.

A análise de variância das variáveis relacionadas ao crescimento das plântulas revelou efeito significativo dos lotes para praticamente todas as características avaliadas aos diferentes períodos de mensuração, evidenciando elevada sensibilidade dessas variáveis para discriminar níveis de vigor em sementes de Mimosa caesalpiniifolia (Figura 3). Os comprimentos da raiz primária (RP), da parte aérea (PA) e o comprimento total das plântulas (PT), bem como os índices de crescimento (IC), uniformidade (IU), vigor (IV) e vigor corrigido (IVC), apresentaram diferenças significativas entre os lotes em pelo menos uma das épocas de avaliação, demonstrando que a deterioração fisiológica afeta diretamente o crescimento inicial das plântulas e a expressão do vigor.

Nos primeiros dias após a semeadura, as diferenças entre os lotes ainda eram relativamente discretas, especialmente para as variáveis relacionadas à parte aérea e aos índices compostos. Entretanto, à medida que o desenvolvimento das plântulas avançou,

tornou-se evidente uma ampliação do contraste fisiológico entre os lotes, refletindo a maior capacidade das sementes mais vigorosas em mobilizar reservas e sustentar taxas superiores de crescimento inicial. Esse comportamento é amplamente descrito na literatura e constitui um dos fundamentos fisiológicos para utilização do crescimento de plântulas como indicador de vigor (MARCOS-FILHO, 2015; HOFFMASTER et al., 2003).

A raiz primária apresentou elevada sensibilidade para discriminar a qualidade fisiológica dos lotes, especialmente aos sete e dez dias após a semeadura (Figura 4). Os lotes classificados previamente como de maior vigor apresentaram maiores valores de RP ao longo de todo o período experimental, enquanto os lotes de menor desempenho fisiológico apresentaram crescimento radicular significativamente reduzido. A elevada sensibilidade do sistema radicular é frequentemente atribuída ao fato de que a expansão celular da radícula constitui um dos primeiros processos metabólicos afetados pela deterioração das sementes, refletindo alterações na integridade das membranas, na atividade respiratória e na eficiência da mobilização de reservas (AOSA, 2009; MARCOS-FILHO, 2015).

Comportamento semelhante foi observado para o comprimento da parte aérea (PA), embora com menor magnitude das diferenças entre os lotes nos períodos iniciais de avaliação. Aos dez dias após a semeadura, entretanto, a variável apresentou maior capacidade discriminatória, evidenciando que os efeitos do vigor sobre o crescimento da parte aérea tendem a se manifestar de forma mais pronunciada em estágios mais avançados do desenvolvimento inicial das plântulas. Resultados semelhantes foram observados por Castan et al. (2018) e Medeiros et al. (2019), que identificaram aumento progressivo da sensibilidade das variáveis morfológicas ao longo do período de crescimento das plântulas.

O comprimento total das plântulas (PT) destacou-se como uma das variáveis mais eficientes na diferenciação dos lotes, integrando simultaneamente as respostas do sistema radicular e da parte aérea. A superioridade dos lotes de maior vigor tornou-se progressivamente mais evidente ao longo das avaliações, especialmente aos dez dias após a semeadura, quando a amplitude entre os lotes extremos atingiu seus maiores valores. O PT tem sido apontado como um dos melhores indicadores do desempenho fisiológico de sementes devido à sua capacidade de integrar múltiplos processos envolvidos no estabelecimento inicial das plântulas (SAKO et al., 2001; HOFFMASTER et al., 2003).

Os índices derivados das mensurações manuais apresentaram comportamento semelhante ao observado para as variáveis morfológicas individuais, confirmando a elevada eficiência da fenotipagem manual para avaliação do vigor em sementes florestais. O índice de crescimento (IC) destacou-se como o parâmetro mais sensível para discriminar os lotes, apresentando diferenças significativas já nas avaliações iniciais e ampliando sua capacidade de separação aos sete e dez dias após a semeadura. Esse resultado era esperado, uma vez que o IC é fortemente dependente do crescimento acumulado das plântulas e incorpora simultaneamente informações do desenvolvimento radicular e da parte aérea (SAKO et al., 2001).

O índice de uniformidade (IU), por sua vez, apresentou menor amplitude de variação entre os lotes, particularmente nas primeiras avaliações. Apesar disso, foi capaz de identificar diferenças fisiológicas em períodos mais avançados do crescimento, indicando que a sincronização do desenvolvimento das plântulas também é influenciada pela qualidade fisiológica das sementes. De maneira geral, a uniformidade tende a ser menos afetada pela deterioração que o crescimento absoluto das plântulas, o que explica sua menor capacidade discriminatória quando comparada ao IC e ao IV (SAKO et al., 2001).

O índice de vigor (IV), que combina crescimento e uniformidade em um único parâmetro, apresentou elevada eficiência na classificação dos lotes ao longo do período experimental. Os maiores valores foram observados nos lotes previamente classificados como de maior vigor pelos testes tradicionais, enquanto os menores valores ocorreram nos lotes de menor desempenho fisiológico. Esses resultados evidenciam forte associação entre os índices derivados das mensurações manuais e os testes convencionais de vigor, corroborando observações previamente relatadas para sistemas automatizados de análise de imagens (HOFFMASTER et al., 2003; CASTAN et al., 2018).

De forma semelhante, o índice de vigor corrigido (IVC) apresentou comportamento consistente com a classificação fisiológica previamente estabelecida pelos testes de emergência e envelhecimento acelerado. A incorporação do percentual de germinação ao cálculo do índice resultou em maior estabilidade dos valores, embora com ligeira redução da sensibilidade em algumas avaliações intermediárias, comportamento semelhante ao descrito por Medeiros e Pereira (2018) durante o desenvolvimento do sistema SAPL®. Ainda assim, o IVC apresentou elevada capacidade de discriminar os lotes extremos e mostrou potencial para utilização em programas de controle de qualidade de sementes florestais.

De maneira geral, os resultados das Figuras 3 e 4 demonstram que os comprimentos de plântulas obtidos por mensurações manuais e os índices derivados dessas variáveis foram capazes de discriminar de forma eficiente os diferentes níveis de vigor entre os lotes de Mimosa caesalpiniifolia. A ampliação progressiva das diferenças ao longo do período de avaliação indica que mensurações realizadas em períodos mais avançados do crescimento inicial, especialmente aos dez dias após a semeadura, tendem a apresentar maior poder discriminatório. Esses resultados reforçam a viabilidade da utilização de metodologias simples, acessíveis e de baixo custo como alternativa aos sistemas automatizados de análise de imagens, ampliando as possibilidades de aplicação dos índices de desempenho em laboratórios de sementes que não dispõem de infraestrutura especializada para fenotipagem digital.

A análise de correlação de Pearson (Figura 5) evidenciou associações consistentes entre os testes tradicionais de caracterização fisiológica e as variáveis de crescimento e índices de desempenho das plântulas avaliadas aos 5, 7 e 10 dias após a semeadura. De modo geral, observou-se aumento na magnitude das correlações ao longo do desenvolvimento das plântulas, indicando que a expressão do vigor torna-se mais evidente durante o estabelecimento inicial, conforme descrito por Marcos-Filho (2015).

Entre os testes fisiológicos, o índice de velocidade de germinação (IVG) apresentou as maiores correlações positivas com as variáveis de crescimento e com os índices obtidos a partir das mensurações manuais, especialmente aos 10 dias após a semeadura. Destacaram-se as associações com o comprimento total de plântulas (PT) (r = 0,81), índice de crescimento (IC) (r = 0,95), índice de vigor (IV) (r = 0,94) e índice de vigor corrigido (IVC) (r = 0,92), evidenciando a forte relação entre rapidez de germinação e desempenho subsequente das plântulas (MAGUIRE, 1962).

A emergência (E) e o índice de velocidade de emergência (IVE) também apresentaram elevadas correlações positivas com os comprimentos de plântulas e índices de desempenho, enquanto o envelhecimento acelerado (EA) apresentou correlações negativas com praticamente todas as variáveis avaliadas, confirmando sua elevada eficiência na estratificação do vigor entre lotes de sementes (KRZYZANOWSKI et al., 2020; MARCOS-FILHO, 2015).

Entre as variáveis morfológicas, o comprimento total das plântulas (PT) apresentou as associações mais elevadas com os índices derivados do crescimento, especialmente com IV (r = 0,95), IVC (r = 0,96) e IC (r = 0,94), demonstrando que a integração do crescimento da raiz primária e da parte aérea representa um indicador robusto do potencial fisiológico das sementes.

Os índices derivados das mensurações manuais destacaram-se entre as variáveis mais associadas aos testes convencionais de vigor. O IC apresentou correlações elevadas com germinação (r = 0,90), primeira contagem (r = 0,92) e IVG (r = 0,95), confirmando sua elevada sensibilidade para discriminar diferenças de qualidade fisiológica entre lotes. Resultados semelhantes foram observados para IV e IVC, enquanto o índice de uniformidade (IU) apresentou correlações moderadas, sugerindo menor capacidade discriminatória quando utilizado isoladamente.

De forma geral, os resultados demonstram que os índices obtidos a partir de mensurações manuais reproduzem adequadamente as relações fisiológicas tradicionalmente observadas em sistemas automatizados de análise de imagens, evidenciando o potencial da fenotipagem manual como ferramenta acessível e eficiente para programas de controle de qualidade de sementes florestais (SAKO et al., 2001; HOFFMASTER et al., 2003; MEDEIROS; PEREIRA, 2018).

A análise de componentes principais (PCA) permitiu integrar as informações provenientes dos testes tradicionais de caracterização fisiológica, das variáveis de crescimento e dos índices de desempenho obtidos a partir das mensurações manuais das plântulas, proporcionando uma visão global das relações existentes entre os lotes e as variáveis avaliadas (Figura 6). Os dois primeiros componentes principais explicaram conjuntamente 88,8% da variabilidade total dos dados, sendo 70,2% atribuídos ao primeiro componente (PC1) e 18,6% ao segundo componente (PC2), indicando elevada capacidade de representação da estrutura multivariada do experimento e reduzindo substancialmente a perda de informação associada à dimensionalidade dos dados.

O primeiro componente principal foi fortemente associado às variáveis tradicionalmente utilizadas na avaliação do vigor, incluindo germinação, primeira contagem, IVG, emergência e IVE, além dos comprimentos e índices de desempenho das plântulas,

evidenciando que essas variáveis compartilham uma mesma estrutura biológica subjacente relacionada ao potencial fisiológico das sementes. Resultados semelhantes têm sido relatados em diferentes espécies agrícolas e florestais, nos quais o primeiro eixo da PCA geralmente representa um gradiente de vigor, separando lotes de alta e baixa qualidade fisiológica (Marcos-Filho, 2015; Finch-Savage & Bassel, 2016).

Os lotes L1 e L2 foram posicionados próximos às variáveis associadas ao crescimento das plântulas e aos índices de desempenho, caracterizando-se como materiais de elevado vigor fisiológico. Por outro lado, L6 e L7 apresentaram associação com o envelhecimento acelerado e distanciamento das demais variáveis, comportamento típico de lotes com maior deterioração fisiológica e menor capacidade de estabelecimento inicial. Segundo Powell (1995) e Marcos-Filho (2015), sementes menos vigorosas apresentam menor eficiência nos processos de reorganização de membranas, mobilização de reservas e retomada metabólica durante a embebição, refletindo diretamente sobre a velocidade de crescimento das plântulas e sua uniformidade.

A proximidade espacial observada entre os testes tradicionais de vigor e os índices derivados das mensurações manuais reforça a elevada capacidade desses índices em representar o desempenho fisiológico das sementes. Em especial, os índices IC, IV e IVC determinados aos 10 dias apresentaram forte associação com germinação, primeira contagem, IVG e emergência, corroborando estudos realizados com sistemas automatizados de análise de imagens, nos quais índices derivados do crescimento das plântulas apresentaram elevada sensibilidade para discriminar diferenças de vigor entre lotes (Sako et al., 2001; Hoffmaster et al., 2003; Medeiros e Pereira, 2024). Esses resultados indicam que a integração entre crescimento e uniformidade das plântulas fornece uma medida mais abrangente do vigor do que avaliações isoladas de comprimento ou germinação.

Outro aspecto relevante observado na PCA foi o aumento progressivo da contribuição das variáveis obtidas aos 10 dias em relação às avaliações realizadas aos cinco dias após a semeadura. Esse comportamento sugere que o avanço do desenvolvimento das plântulas amplia a expressão fenotípica das diferenças fisiológicas entre os lotes, aumentando o poder discriminatório dos índices de desempenho. Resultados semelhantes foram relatados por Gomes-Junior et al. (2014) e Medeiros et al. (2020), que observaram maior eficiência dos sistemas de fenotipagem de plântulas em estágios mais avançados do crescimento inicial.

A orientação oposta do vetor correspondente ao envelhecimento acelerado em relação às demais variáveis confirma sua relação inversa com o vigor fisiológico, uma vez que sementes mais tolerantes ao estresse imposto pelo teste apresentam maior integridade de membranas, maior eficiência metabólica e melhor desempenho durante o estabelecimento inicial das plântulas (Delouche & Baskin, 1973; Marcos-Filho, 2015). Essa relação inversa também foi evidenciada na análise de correlação de Pearson, reforçando a consistência entre as diferentes abordagens estatísticas empregadas no estudo.

De maneira geral, a análise multivariada corroborou os resultados obtidos pelas análises univariadas e pela matriz de correlação, demonstrando que os índices obtidos a partir de mensurações manuais apresentam elevada eficiência na discriminação dos níveis de

vigor entre lotes de sementes de Mimosa caesalpiniifolia. Esses resultados ampliam as evidências sobre o potencial de utilização desses índices como ferramentas acessíveis e de baixo custo para laboratórios que não dispõem de sistemas automatizados de análise de imagens, contribuindo para a democratização das tecnologias de fenotipagem aplicadas à avaliação do vigor de sementes florestais.

Conclusão

Os índices obtidos a partir das mensurações manuais de plântulas foram eficientes na discriminação dos níveis de vigor entre lotes de sementes de Mimosa caesalpiniifolia. Os índices de crescimento (IC), vigor (IV) e vigor corrigido (IVC), especialmente aos 10 dias após a semeadura, apresentaram maior sensibilidade para detectar diferenças no potencial fisiológico dos lotes e elevada associação com os testes convencionais de vigor. No entanto, os índices derivados do crescimento de plântulas constituem ferramentas promissoras para a avaliação do vigor e a classificação da qualidade fisiológica de sementes da espécie, já a partir dos 5 dias após a semeadura.

Referências

ABUD, H. F.; GONÇALVES, N. R.; REIS, R. G. E.; GALLÃO, M. I.; INNECCO, R. Morfologia de sementes e plântulas de espécies florestais da Caatinga. Revista Ciência Agronômica, Fortaleza, v. 41, n. 3, p. 478–483, 2010.

ASSOCIATION OF OFFICIAL SEED ANALYSTS (AOSA). Seed vigor testing handbook. Contribution n. 32 to the Handbook on Seed Testing. Lincoln: AOSA, 2009. 341 p.

BEWLEY, J. D.; BRADFORD, K. J.; HILHORST, H. W. M.; NONOGAKI, H. Seeds: physiology of development, germination and dormancy. 3. ed. New York: Springer, 2013. 392 p. DOI: 10.1007/978-1-4614-4693-4.

CARVALHO, P. E. R. Sabiá (Mimosa caesalpiniifolia). Colombo: Embrapa Florestas, 2007. 10 p. (Circular Técnica, 135).

CASTAN, D. O. C.; GOMES-JUNIOR, F. G.; MARCOS-FILHO, J. Computerized seedling analysis for vigor evaluation of common bean seeds. Journal of Seed Science, Londrina, v. 40, n. 2, p. 167–176, 2018. DOI: 10.1590/2317-1545v40n2180414.

DELOUCHE, J. C.; BASKIN, C. C. Accelerated aging techniques for predicting the relative storability of seed lots. Seed Science and Technology, Zürich, v. 1, n. 2, p. 427–452, 1973.

FINCH-SAVAGE, W. E.; BASSEL, G. W. Seed vigour and crop establishment: extending performance beyond adaptation. Journal of Experimental Botany, Oxford, v. 67, n. 3, p. 567–591, 2016. DOI: 10.1093/jxb/erv490.

FINCH-SAVAGE, W. E.; LEUBNER-METZGER, G. Seed dormancy and the control of germination. New Phytologist, London, v. 171, n. 3, p. 501–523, 2006. DOI: 10.1111/j.1469-8137.2006.01787.x.

GOMES-JUNIOR, F. G.; CHAMMA, H. M. C. P.; CICERO, S. M. Automated image analysis of seedlings for vigor evaluation of common bean seeds. Acta Scientiarum. Agronomy, Maringá, v. 36, n. 2, p. 195–200, 2014. DOI: 10.4025/actasciagron.v36i2.18666.

HOFFMASTER, A. L.; FUJIMURA, K.; MCDONALD, M. B.; BENNETT, M. A. An automated system for vigor testing three-day-old soybean seedlings. Seed Science and Technology, Zürich, v. 31, n. 3, p. 701–713, 2003. DOI: 10.15258/sst.2003.31.3.15.

INTERNATIONAL SEED TESTING ASSOCIATION (ISTA). International rules for seed testing 2024. Bassersdorf: ISTA, 2024.

KRZYZANOWSKI, F. C.; VIEIRA, R. D.; FRANÇA-NETO, J. B.; MARCOS-FILHO, J. Vigor de sementes: conceitos e testes. Londrina: ABRATES, 2020. 601 p.

LORENZI, H. Árvores brasileiras: manual de identificação e cultivo de plantas arbóreas nativas do Brasil. 8. ed. Nova Odessa: Instituto Plantarum, 2020. v. 1.

MAGUIRE, J. D. Speed of germination-aid in selection and evaluation for seedling emergence and vigor. Crop Science, Madison, v. 2, n. 2, p. 176–177, 1962. DOI: 10.2135/cropsci1962.0011183X000200020033x.

MAIA, G. N. Caatinga: árvores e arbustos e suas utilidades. 2. ed. Fortaleza: Printcolor Gráfica e Editora, 2012. 413 p.

MARCOS-FILHO, J. Fisiologia de sementes de plantas cultivadas. 2. ed. Londrina: ABRATES, 2015. 660 p.

MARCOS-FILHO, J. Seed vigor testing: an overview of the past, present and future perspective. Scientia Agricola, Piracicaba, v. 72, n. 4, p. 363–374, 2015. DOI: 10.1590/0103-9016-2015-0007.

MARCOS-FILHO, J.; KIKUTI, A. L. P.; LIMA, L. B. C. Methods for soybean seed vigor testing in relation to seed storage and field emergence. Crop Breeding and Applied Biotechnology, Viçosa, v. 9, n. 2, p. 137–146, 2009. DOI: 10.12702/1984-7033.v09n02a05.

MEDEIROS, A. D.; PEREIRA, M. D. SAPL®: a free software for determining the physiological potential in soybean seeds. PLoS ONE, San Francisco, v. 13, n. 12, e0201929, 2018. DOI: 10.1371/journal.pone.0201929.

MEDEIROS, A. D.; SILVA, L. J.; CAPOBIANGO, N. P.; FIALHO, C. M. T.; DIAS, D. C. F. S. Assessing the physiological quality of soybean seeds using image analysis. Seed Science and Technology, Zürich, v. 48, n. 1, p. 11–24, 2020. DOI: 10.15258/sst.2020.48.1.02.

Anexos

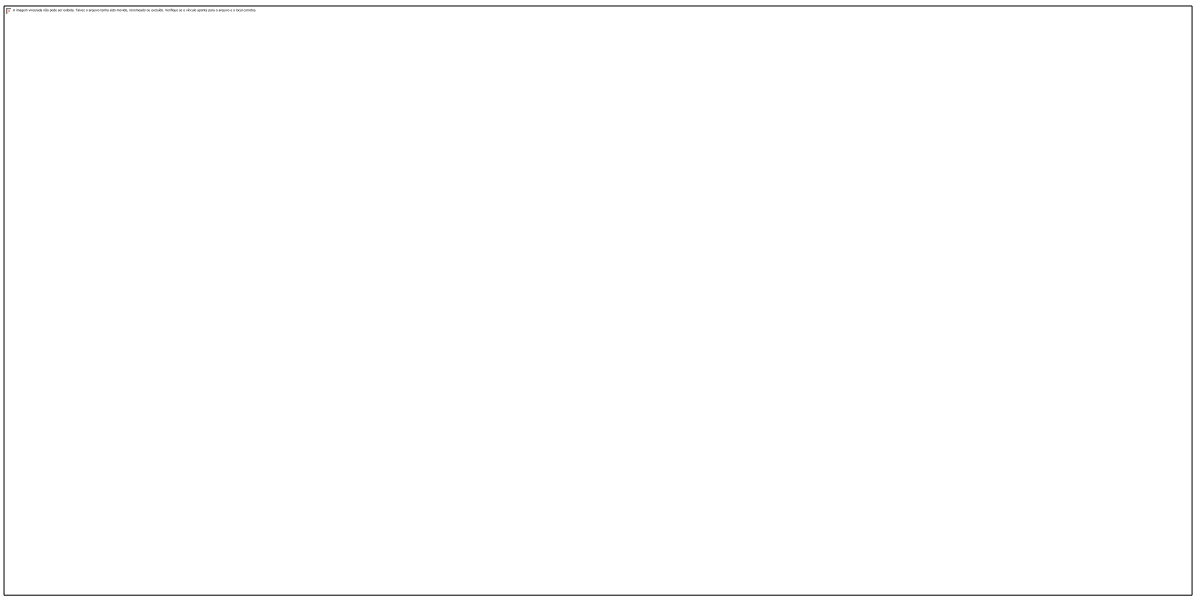


Figura 1. Análise de variância da germinação (G), primeira contagem (PC), índice de velocidade de germinação (IVG), massa seca de plântulas (MS), envelhecimento acelerado (EA), emergência (E) e índice de velocidade de emergência (IVE) de lotes de sementes

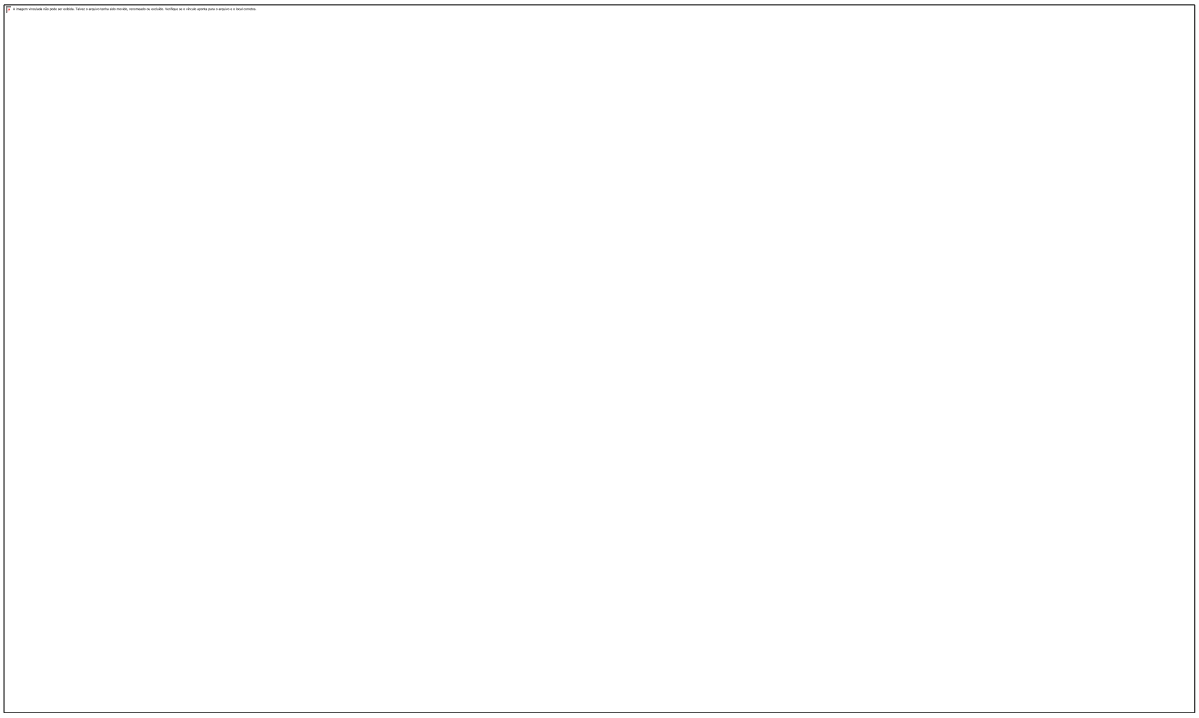


Figura 2. Médias da germinação (G), primeira contagem (PC), índice de velocidade de germinação (IVG), emergência (E), índice de velocidade de emergência (IVE), envelhecimento acelerado (EA) e massa seca de plântulas (MS) de lotes de sementes de Mimosa cae



Figura 3. Análise de variância dos comprimentos da raiz primária (RP), da parte aérea (PA) e total das plântulas (PT), bem como dos índices de crescimento (IC), uniformidade (IU), vigor (IV) e vigor corrigido (IVC) de plântulas de Mimosa caesalpiniifolia



Figura 4. Comprimento da raiz primária (RP), comprimento da parte aérea (PA), comprimento total das plântulas (PT), índice de crescimento (IC), índice de uniformidade (IU), índice de vigor (IV) e índice de vigor corrigido (IVC) de plântulas de Mimosa caes

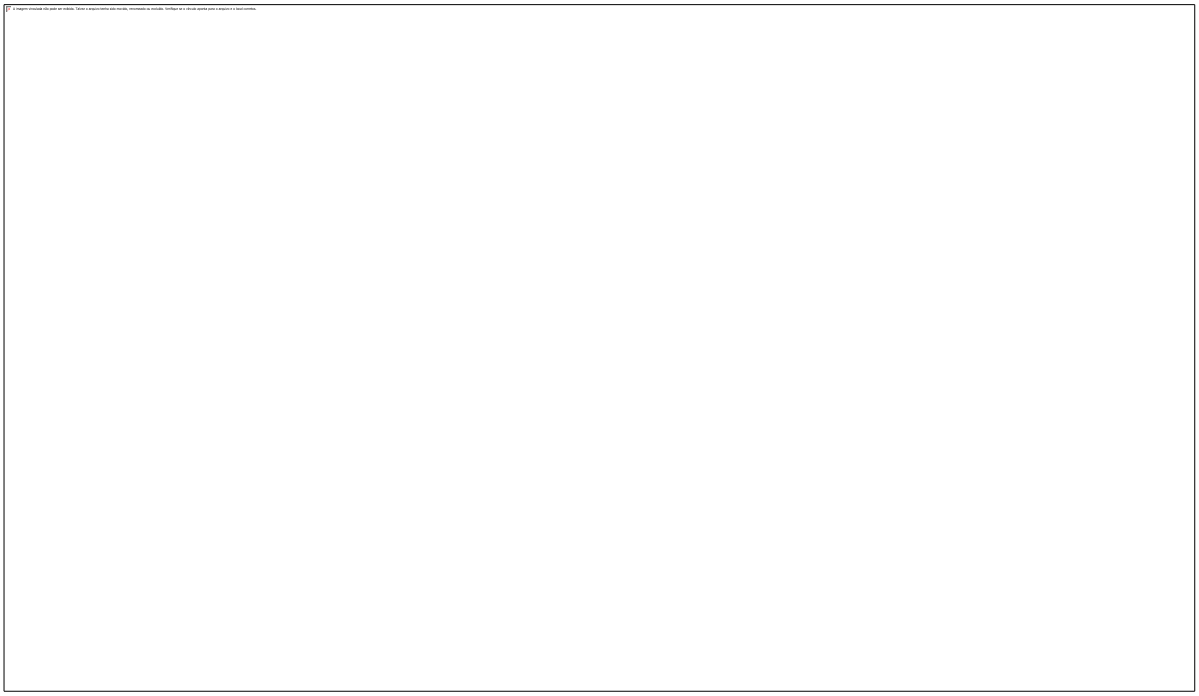


Figura 5. Mapa de calor da correlação de Pearson entre os testes de caracterização fisiológica das sementes e as variáveis de crescimento e índices de desempenho de plântulas de Mimosa caesalpiniifolia avaliadas aos 5, 7 e 10 dias após a semeadura.

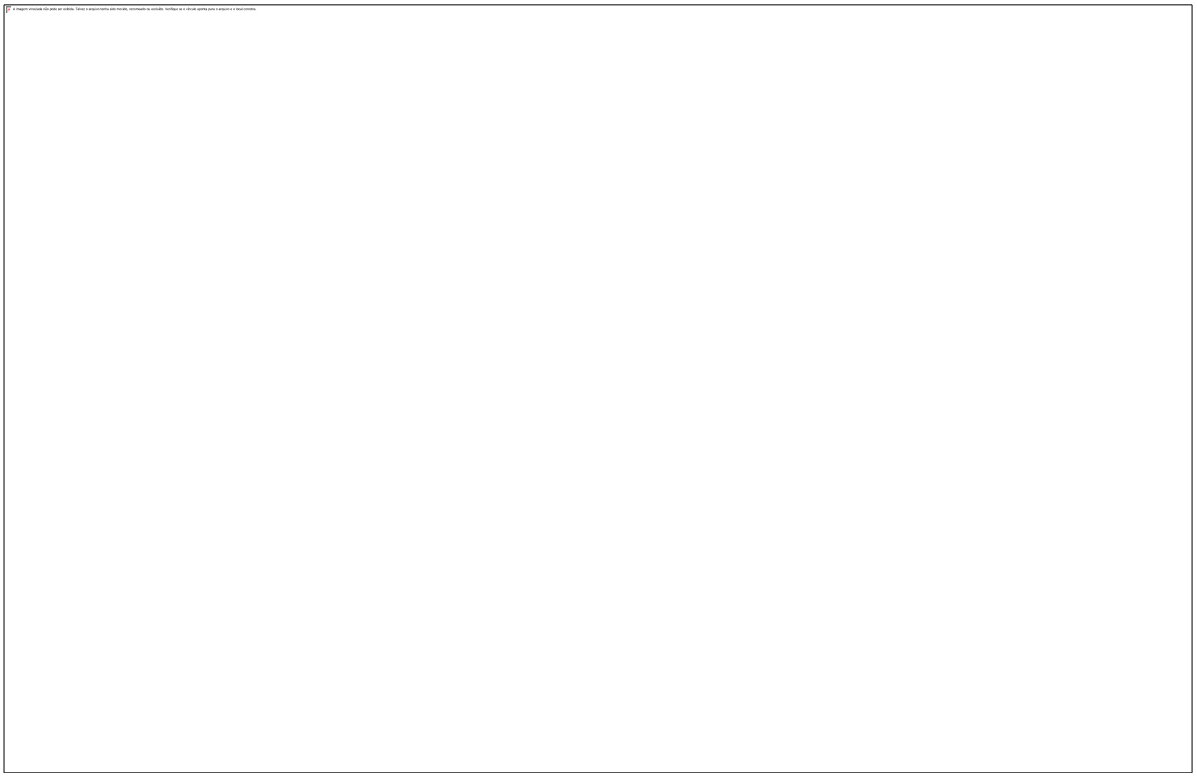


Figura 6. Biplot da análise de componentes principais (PCA) integrando testes tradicionais de vigor, variáveis de crescimento e índices de desempenho de plântulas para discriminação do potencial fisiológico de lotes de sementes de Mimosa caesalpiniifolia.

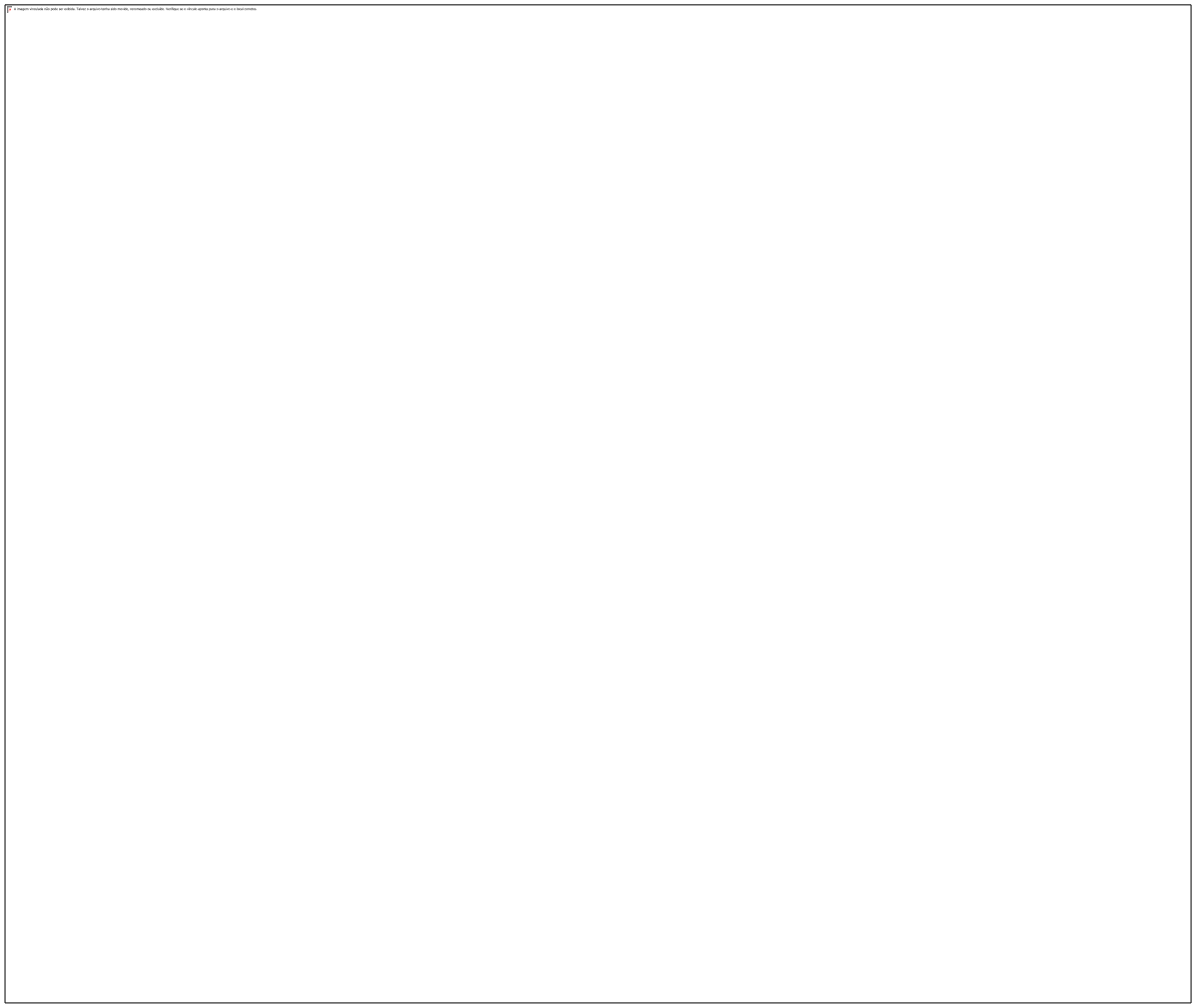


Tabela 1. Fórmulas utilizadas para cálculo dos índices de crescimento, uniformidade, vigor e vigor corrigido baseados na análise de plântulas.

CÓDIGO: SB0138

AUTOR: ALISON RIAN NUNES DA SILVA

COAUTOR: ANTONINO ALVES FEITOSA NETO

COAUTOR: GEAN CARLOS DA SILVA SANTOS

ORIENTADOR: MARCIO DIAS PEREIRA

TÍTULO: Automação da análise de sementes: redes neurais convolucionais e processamento digital de imagens.

Resumo

O estudo teve como objetivo avaliar a precisão da extração do comprimento da parte aérea e total de plântulas de Cenostigma pyramidale por meio de máscaras geradas pela segmentação realizada com o YOLOv8. As medições automatizadas foram comparadas às medições manuais por meio de testes de correlação, coeficiente de correlação intraclasse, análise de Bland-Altman e métricas de erro (MAE e RMSE). As análises indicaram uma diferença média entre os métodos de 0,16 cm para a parte aérea e 0,14 cm para o total, em plântulas com comprimento da parte aérea e total inferiores a 4,5 cm e 12 cm, respectivamente. No entanto, a diferença aumentou para plântulas maiores, indicando um viés proporcional no método e tendência a superestimar as medidas. Conclui-se que a utilização das redes neurais convolucionais na fenotipagem digital apresenta grande potencial, embora ajustes e aprimoramentos na metodologia sejam necessários para melhorar a acurácia para plântulas maiores.

Palavras-chave: Cenostigma pyramidale, Visão computacional, YOLOv8, Fenotipagem digita

TITLE: Automação da análise de sementes: redes neurais convolucionais e processamento digital de imagens.

Abstract

The study aimed to evaluate the accuracy of shoot and total length extraction in Cenostigma pyramidale seedlings using masks generated by YOLOv8 segmentation. Automated measurements were compared with manual measurements using correlation tests, intraclass correlation coefficient, Bland-Altman analysis, and error metrics (MAE and RMSE). The analyses indicated a mean difference between methods of 0.16 cm for the shoot and 0.14 cm for the total length, in seedlings with shoot and total lengths below 4.5 cm and 12 cm, respectively. However, the difference increased for larger seedlings, indicating a proportional bias in the method and a tendency to overestimate measurements. It is concluded that the use of convolutional neural networks in digital phenotyping has great potential, although adjustments and improvements in the methodology are necessary to improve accuracy for larger seedlings.

Keywords: Cenostigma pyramidale, Computer vision, YOLOv8, Digital phenotyping.

Introdução

A cenostigma pyramidale, popularmente conhecida como catingueira, é uma espécie de árvore nativa da caatinga brasileira. A espécie serve como local de construção para o ninho de espécies de abelhas nativas, exercendo assim, um papel fundamental no equilíbrio ecológico e sendo indicada para o plantio em regiões de conservação. Além disso, apresenta elevado potencial econômico devido à comercialização de sua madeira, a sua utilização no paisagismo urbano e graças ao seu rápido desenvolvimento, a espécie pode ser utilizada em projetos de reflorestamento.(OLIVEIRA, 1976; SILVA et., 2012). Nesse cenário, a identificação de sementes e mudas de catingueira de qualidade é fundamental para sua utilização nos contextos citados. Segundo Lopes e Nascimento (2009), a análise de sementes avalia parâmetros físicos, genéticos, fisiológicos e sanitários para determinar a qualidade das amostras. Essa etapa fundamental garante a produtividade e a sustentabilidade do plantio além de ser requisito obrigatório para a comercialização.

A análise do vigor das plântulas está relacionada à soma de um conjunto de características que determinam o potencial para a emergência rápida e uniforme de plântulas em diferentes condições (MARCOS FILHO, 2015). O vigor é uma das principais formas de avaliação e pode ser determinado a partir do comprimento das plântulas visto que sementes de maior vigor tendem a originar plântulas com maiores taxas de crescimento (DAN et al., 1987). No entanto, essas medições por serem realizadas manualmente, estão sujeitas à subjetividade e ao erro humano e, por vezes, podem ser destrutivas e laboriosas (BRUNES et al., 2016).

Nesse contexto, a utilização de ferramentas digitais para extração das métricas pode automatizar e padronizar o processo. O processamento digital de imagens (PDI) e a utilização de técnicas de visão computacional baseado em aprendizado profundo (deep learning), como as redes neurais convolucionais (CNN) são alternativas promissoras para a fenotipagem digital (MEDEIROS et al., 2018; BRUNES et al., 2016; ROCHA et al., 2018; BERNARDES et al., 2022; JIANG et al., 2019). As CNNs possuem a

capacidade de extrair características da imagem sem interferência manual o que possibilita o aprendizado de padrões complexos de forma automática e sua aplicação em diferentes áreas (LI et al., 2020).

No contexto das avaliações de vigor, uma arquitetura de rede neural convolucional pode ser utilizada para separar as plântulas do restante da imagem, possibilitando a medição individual do comprimento das estruturas. A tarefa de separar os objetos ou regiões de interesse dos outros elementos na imagem, é chamada de segmentação (GONZALEZ; WOODS, 2018). Uma arquitetura de rede neural convolucional para segmentação de instâncias, como o YOLOv8-seg apresenta a capacidade de isolar objetos de diferentes categorias e diferenciar indivíduos de uma mesma classe (MINAEE et al., 2022).

Diante desse cenário, o objetivo deste trabalho é avaliar a precisão da extração do comprimento total e da parte aérea de plântulas de Cenostigma pyramidale a partir da segmentação realizada pela rede neural convolucional YOLOv8. Para isso, foram definidos os objetivos específicos: definir uma configuração adequada para o problema, realizar o pré-processamento e a organização das imagens, treinar o modelo YOLOv8-seg, validar o modelo treinado e comparar os valores obtidos por meio do método proposto com as medições manuais realizadas por especialistas.

Metodologia

2.1 Aquisição e anotação dos dados

A base de dados utilizada para o treinamento do modelo e a validação dos resultados foi obtida pelo Grupo de Estudos em Sementes (GETSEM) da Escola Agrícola de Jundiaí (EAJ). O conjunto de dados é composto por 95 imagens de 8 lotes de plântulas com 4 repetições cada, as quais foram capturadas aos 5, 7 e 10 dias após a germinação. Para a realização dos registros, as amostras foram dispostas sobre uma superfície plana de cor preta e distribuídas de maneira a não ocorrer sobreposições. Em todas as imagens foi posicionada uma moeda de 50 centavos, objeto de tamanho conhecido para servir de referência para converter o tamanho das amostras de pixels para centímetros. A figura 1 exemplifica a estrutura utilizada para a captura das imagens. O YOLOv8-seg necessita da anotação prévia dos dados de treinamento, para realizar as anotações foi utilizada a plataforma CVAT com o auxílio da ferramenta de inteligência artificial SAM 3.0 para realizar as marcações de duas classes: parte aérea e plântula completa. Foram realizados ajustes manuais em situações onde o SAM não foi capaz de gerar adequadamente as máscaras, com o objetivo de manter as anotações próximas da estrutura real das plântulas.

2.2 Treinamento do modelo

Para a realização do treinamento dos modelos foi utilizada uma abordagem com dois níveis: um nível externo e um nível interno. No nível externo, foi utilizado a técnica de validação cruzada (cross-validation) com 8 folds, um para cada lote, dessa forma 7 lotes foram destinados ao nível interno e 1 lote foi reservado para a estimativa das medidas de comprimento. Essa abordagem foi adotada com o objetivo de garantir o desempenho

confiável dos modelos em dados nunca vistos, permitindo estimar os comprimentos de todos os lotes sem contaminação. No nível interno, visando determinar os melhores hiperparâmetros para o treinamento, foi utilizada a otimização bayesiana com 30 iterações utilizando a mAP50-95 como métrica alvo. Foram definidas 3 configurações possíveis para cada um dos seguintes hiperparâmetros: taxa de aprendizado (learning rate), a regularização (weight decay), o tamanho da imagem, quantidade de épocas e 3 configurações possíveis para cada uma das 3 técnicas de aumento de dados (data augmentation): ângulo de rotação (degrees), variação de brilho e desfoque gaussiano (gaussian blur), conforme apresentado na Tabela 1.
Além disso, o método holdout foi utilizado para dividir os dados na etapa de otimização, foi realizada a seleção aleatória de 20% dos dados, estratificados por lote, os quais foram destinados para validação e o restante dos dados destinados para o treinamento. Por fim, os melhores hiperparâmetros escolhidos no nível interno foram utilizados para treinar o modelo do nível externo.

2.3 Pipeline de processamento

Ao final da etapa de treinamento foram gerados 8 modelos, um para realizar as predições de cada um dos 8 lotes. As imagens dos lotes antes de serem submetidas a predição do seu respectivo modelo, foram redimensionadas mantendo a proporção da imagem original, para a mesma largura do tamanho de imagem com que o modelo foi treinado. Após essa etapa, foram aplicadas as predições, que retornam como resultado da operação, as máscaras binárias individuais de todos os objetos identificados. Para cada imagem, foi utilizada uma abordagem baseada na circularidade dos contornos para identificar a moeda e então, a partir do diâmetro em pixels do objeto, foi calculado um fator de escala para conversão das medidas de pixels para centímetros.
As máscaras foram ordenadas por meio da sua coordenada no eixo x e submetidas a uma operação de fechamento com um kernel 7x7 para preencher pequenas falhas nas máscaras e unir partes desconexas. Em seguida, as máscaras individuais foram filtradas, mantendo-se apenas o objeto conectado de maior área.
Para extrair o comprimento das plântulas, aplicou-se a operação de esqueletização em cada máscara e filtrou-se o esqueleto de maior comprimento do eixo principal, removendo ramificações secundárias. Para realizar essa filtragem, utilizou-se uma abordagem com duas buscas em largura, a primeira utiliza o primeiro pixel branco encontrado como base para encontrar o ponto mais distante dele, chamado de A, a segunda busca em largura parte de A para encontrar o ponto mais distante, chamado de B, o caminho entre A e B é o caminho mais longo possível do esqueleto. Em seguida, utilizou-se o cálculo da distância euclidiana entre os pontos que formam o esqueleto do eixo principal da plântula. O tamanho de cada parte foi convertido para centímetros, dividindo a quantidade de pixels pelo fator de escala calculado no início do processo. Por fim, as informações foram armazenadas em um arquivo .csv, contendo as colunas de fold, lote, repetição, dia do registro da imagem, nome do arquivo de imagem, número da plântula, comprimento da parte aérea, comprimento da raíz e comprimento total.

2.4 Análise estatística

Para as análises estatísticas das medidas extraídas pelo método computacional foram utilizadas para comparação as medidas manuais realizadas por 3 avaliadores especialistas. Os dados das duas abordagens foram concatenados em uma única tabela e

agrupados por repetição, em seguida realizou-se o pré-processamento dos dados e posteriormente o cálculo da média das medidas para cada repetição.
Em seguida, foi verificada a normalidade na distribuição dos dados por meio do teste de Shapiro-wilk, da avaliação de gráficos q-q e de histogramas. Para as variáveis que apresentaram uma distribuição normal ($p > 0,05$) o cálculo da correlação foi realizado por meio do método de Pearson e com o método de Spearman para as variáveis que não seguiam uma distribuição normal, com o objetivo de determinar se as medidas obtidas com os dois métodos aumentam ou diminuem juntas. Após a análise de correlação, foi utilizado o coeficiente de correlação intraclasse (ICC) para avaliar a confiabilidade entre os métodos. Em seguida, para detectar

Resultados e Discussões

3.1 Resultados da otimização dos hiperparâmetros e treinamento

A escolha dos melhores hiperparâmetros por meio da otimização bayesiana indicou que a utilização de imagens com tamanho de 1024 apresentou resultados melhores que os tamanhos inferiores, evidenciando que resoluções espaciais maiores contribuem para segmentações mais precisas das plântulas. A convergência consistente da taxa de aprendizado inicial e a quantidade de épocas entre a maioria dos folds indica que os valores são escolhas seguras dentro do espaço de busca determinado. Por outro lado, a divergência na escolha dos valores de regularização (weight decay) e as configurações de aumento de dados entre os diferentes folds, sugere que a escolha do melhor valor dentro do espaço de busca estabelecido pode ser influenciada por variações dos dados em cada fold, assim cada subconjunto demanda taxas de decaimento e configurações de aumento de dados específicas. A Tabela 2 apresenta os hiperparâmetros definidos para cada fold na otimização bayesiana.

Após a definição dos melhores hiperparâmetros para cada fold, foram realizados os treinamentos no nível externo e em seguida a validação. As métricas de desempenho apresentadas para a parte aérea foram elevadas, tanto para as caixas delimitadoras (bounding boxes)quanto para as máscaras. Por outro lado, o total apresentou métricas elevadas para as caixas e valores baixos para as máscaras, como exemplificado na Tabela 3.

A Figura 2 apresenta o gráfico de perda (loss) para o fold 1. As curvas de perda na validação e no treinamento apresentaram uma diminuição consistente durante as épocas e convergiram de forma estável, indicando que não ocorreu sobreajuste (overfitting) do modelo aos dados de treinamento. Padrões similares na curva de perda e consequentemente a ausência de sobreajuste foram identificados para todos os folds usados durante a validação cruzada.

3.2 Análise de normalidade e correlação

Os resultados dos teste de Shapiro-Wilk apresentaram, para todas as variáveis, um valor p menor que o nível de significância de 0,05, sugerindo que os dados não seguem uma distribuição normal. Os histogramas não apresentam o formato característico de sino e

os gráficos q-q apresentaram desvios dos pontos em relação a reta, dessa forma a análise complementar por meio dos histogramas e gráficos q-q reforçam a hipótese do teste de Shapiro-Wilk indicando a não normalidade dos dados.

A análise da normalidade indicou que nenhuma das colunas segue uma distribuição gaussiana, por isso para análise da correlação utilizou-se o método de Spearman. Os resultados indicam uma alta correlação positiva entre as medidas dos avaliadores apresentando valores acima de 0,9. Os coeficientes para o comprimento total das plântulas foram levemente superiores aos observados nas medidas da parte aérea. Entre o método computacional e os avaliadores, os resultados indicam uma forte associação entre as variáveis, embora os valores tenham sido inferiores para a parte aérea.A alta correlação positiva apresentada indica que à medida que os comprimentos registrados pelo avaliadores aumentam, o comprimento medido pelo método computacional também aumenta.

3.3 Avaliação do ICC e teste de Friedman

A análise da concordância entre as medições foi realizada por meio do coeficiente de correlação intraclasse com concordância absoluta, modelo aleatório de duas vias e único avaliador (ICC 2,1) (KOO; LI, 2016). Os resultados demonstraram uma excelente concordância entre os avaliadores humanos para a parte aérea (ICC = 0.97; IC 95% = (0.93 - 0.98, F (92, 184) = 153.07, $p < 0,001$) e para o comprimento total (ICC = 0.99; IC 95% = (0.99 - 1.0, F (92, 184) = 661.12, $p < 0,001$) indicando que os avaliadores são bem calibrados entre si.

Diante da elevada concordância observada entre os avaliadores, calculou-se a média dos medições humanas para cada repetição e esses valores foram utilizados como referência humana. Em seguida, realizou-se o cálculo do ICC entre o método computacional e a referência humana. Para a parte aérea observou-se uma concordância moderada, no limite inferior de uma classificação considerada boa (ICC = 0.75; IC 95% = (0.63 - 0.84), F (92, 92) = 8.00, $p < 0,001$) enquanto para o comprimento total, observou-se uma excelente concordância (ICC = 0.95; IC 95% = (0.93, 0.97), F (92, 92) = 43.98, $p < 0,001$).

O ICC entre os resultados individuais dos avaliadores juntamente com o método computacional apresentaram uma boa concordância para a parte aérea (ICC = 0,83; IC 95% = (0.77 - 0.88), F (92, 276) = 23.75, $p < 0,001$) e uma excelente concordância para o comprimento total (ICC = 0.97; IC 95% = (0.96 - 0.98), F (92, 276) = 147.25, $p < 0,001$).

Em seguida, aplicou-se o teste de Friedman para identificar diferenças entre as medições, a aplicação do teste resultou em um valor de p inferior a 0,05 indicando diferença estatisticamente significativa. Com o objetivo de identificar entre quais métodos ocorreram as diferenças, foi aplicado o teste de Nemenyi, os resultados estão apresentados na Figura 3. O teste evidenciou divergência entre as medições manuais dos avaliadores ($p < 0,05$), enquanto as medições modelo foram equivalentes ao avaliador 2. Para o comprimento total, o avaliador 3 se diferenciou dos demais avaliadores e do método automatizado, que por sua vez apresentou resultados semelhantes aos avaliadores 1 e 2. Esses resultados sugerem a ocorrência de diferenças sistemáticas

entre as medições dos avaliadores humanos. Essa divergência pode estar relacionada à delimitação do ponto de transição entre a raíz e a parte aérea, dessa forma cada avaliador pode determinar o comprimento da parte aérea levemente acima ou abaixo que os demais.

3.4 Análise das métricas de discordâncias, Bland-Altman e métricas de erro

O teste de Wilcoxon foi aplicado entre as métricas da discordância absoluta média entre os avaliadores e a discordância absoluta média entre o método computacional e os avaliadores. O teste indicou que as duas métricas de discordância são estatisticamente diferentes, tanto para a parte aérea quanto para o comprimento total ($p < 0,05$). Comparando a média de cada métrica identificou-se que a média da discrepância entre os avaliadores para a parte aérea foi de 0,189 enquanto entre o método computacional e os avaliadores foi de 0,676, para o comprimento total 0,210 e 0,670, respectivamente. Esses resultados indicam que o método automatizado discorda mais dos humanos do que os humanos discordam entre si.

O gráfico de Bland-Altman indica uma diferença média de -0,32 cm entre as medições manuais da parte aérea e as medições automatizadas, esse resultado negativo indica que o software tende a apresentar na média, valores maiores. Além disso, observa-se uma diferença média próxima a zero para os comprimentos médios abaixo de 4 cm e um aumento visível das diferenças conforme a média dos comprimentos cresce, esse comportamento evidencia uma tendência no software de superestimar as medições da parte aérea em amostras com comprimento médio superior a 4 cm. A média da diferença entre os dois métodos em plântulas com parte aérea com comprimento maior que 4 cm é de -0,88 cm, essa diferença aumenta para -1,43 cm em amostras com comprimento superior a 4,5 cm. A Figura 4 apresenta o gráfico de Bland-Altman entre a referência humana e o método para o comprimento da parte aérea.

Para o comprimento total, a Figura 5 evidencia uma diferença média de -0,14 cm entre os métodos, o que indica que o software mede em média 0,14 cm a mais. Observa-se que para plântulas com o comprimento médio menor que 10 cm, a diferença entre as medições é próxima a zero. Assim como para a parte aérea a discrepância entre os métodos tende a aumentar em plântulas maiores, com o método automatizado superestimando as medidas de plântulas com comprimento médio maior que 10 cm, resultando em uma diferença média de -0,82 cm. Essa diferença média aumenta para -1,13 cm em plântulas com comprimento médio maior que 12 cm.

Os resultados obtidos para as duas medidas evidenciam uma tendência no método a superestimar o comprimento em relação à referência humana à medida que o comprimento das plântulas aumenta. Em contrapartida, o software apresenta uma precisão elevada para plântulas menores, na faixa de valores abaixo de 4,5 cm para a parte aérea e abaixo de 12 cm para o comprimento total, a diferença média foi de 0,16 cm e 0,14 cm, respectivamente. Esse comportamento revela que o método tende a apresentar valores próximos aos obtidos com as medições manuais para plântulas em seus estágios iniciais.

Por fim, os resultados observados para as métricas de erro nas medições da parte aérea foram um MAE de 0,66 cm e um RMSE 0,92 cm, enquanto que para o comprimento

total, foram de 0,64 cm e 0,98 cm, respectivamente. O RMSE mais alto que o MAE indica que para algumas amostras ocorreram erros maiores, esse resultado também é coerente com a análise de Bland-Altman, que revelou maiores diferenças em relação a referência para plântulas de maiores dimensões. Além disso, embora o erro absoluto médio entre a parte aérea e o total apresentem valores aproximados, o erro nas estimativas para a parte aérea foi consideravelmente maior, com um erro percentual absoluto médio (MAPE) de 17,53% enquanto para o comprimento total foi de 6,52%.

3.5 Discrepância entre as métricas de segmentação e os resultados das medições

Apesar dos modelos apresentarem métricas reduzidas de segmentação para a plântula completa, as máscaras obtidas permitiram alcançar níveis de erro relativamente baixos, como o MAE de 0,64 cm. Além disso, a alta correlação observada entre os avaliadores e o método, o excelente coeficiente de correlação intraclasse apresentado, o viés médio de -0,14 cm, a ausência de evidências de diferença significativa em relação a 2 dos 3 avaliadores e uma melhor precisão do que a parte aérea, indicam que somente as métricas de segmentação não definem a preservação das informações necessárias na extração das medidas de comprimento.

Ainda que as métricas de segmentação para o total sejam baixas, as máscaras geradas aparentam ser adequadas em relação às plântulas reais. No entanto, eventualmente as máscaras podem apresentar regiões desconexas ou pequenas diferenças nas bordas do objeto. A Figura 6 apresenta a maioria das máscaras coerentes com a imagem real, apesar de uma das máscaras apresentar uma pequena falha na raíz.

Conclusão

Neste estudo, o método de mensuração das plântulas de cenostigma pyramidale por meio das máscaras geradas pelo modelo YOLOv8 apresentou resultados aproximados das medições humanas em plântulas com comprimento total médio inferior a 10 cm e parte aérea inferior a 4 cm, no entanto apresenta um viés proporcional para plântulas com comprimento superior a esses valores.

As diferenças de desempenho indicam uma maior eficácia do método proposto em medir corretamente plântulas em seus estágios iniciais. Em contrapartida, para as plântulas maiores, observou-se uma tendência a superestimar as medições. O comportamento observado pode não estar relacionado diretamente à qualidade das máscaras, as quais tendem a apresentar pequenas perdas de pixels nas extremidades do objeto, o que causaria subestimação das medidas, mas devido a metodologia de extração das métricas. Adicionalmente, a precisão reduzida para a parte aérea, pode indicar uma maior dificuldade em definir precisamente o ponto de transição (colo) entre a raíz e a parte aérea.

Assim, conclui-se que a utilização de redes neurais convolucionais, como YOLOv8, possui grande potencial para fenotipagem digital, apesar de serem necessários

aprimoramentos na metodologia utilizada para aumentar a precisão em estágios mais avançados de desenvolvimento. Em pesquisas futuras, recomenda-se a utilização de bases de imagens maiores, a comparação com outras arquiteturas de segmentação, aprimoramento do algoritmo de extração e o treinamento de modelos especializados para cada estágio de desenvolvimento das plântulas.

Referências

BERNARDES, Rodrigo Cupertino et al. Deep-learning approach for fusarium head blight detection in wheat seeds using low-cost imaging technology. Agriculture, v. 12, n. 11, p. 1801, 2022.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Regras para Análise de Sementes. Brasília, DF: MAPA/ACS, 2009. 399 p.

BRUNES, A. P. et al. Seedling length in wheat determined by image processing using mathematical tools. Revista Ciência Agronômica, v. 47, n. 2, p. 374-379, 2016.

BRUNES, A. P. et al. Rice seeds vigor through image processing of seedlings. Ciência Rural, v. 49, n. 8, e20180633, 2019.

DAN, E. L. et al. Transferência de matéria seca como método de avaliação do vigor de sementes de soja. Revista Brasileira de Sementes, v. 9, n. 3, p. 45-55, 1987.

DANTAS DE MEDEIROS, André; PEREIRA, Márcio; SILVA, Jackson. Processamento digital de imagens na determinação do vigor de sementes de milho. Revista Brasileira de Ciências Agrárias, v. 13, n. 3, p. 1-7, 2018.

DE OLIVEIRA, Odaci Fernandes. Algumas árvores do município de Mossoró. Revista Caatinga, v. 1, n. 1, p. 7–17, 2026. Disponível em: https://periodicos.ufersa.edu.br/index.php/caatinga/article/view/2397. Acesso em: 17 jul. 2026.

GONZALEZ, Rafael C.; WOODS, Richard E. Digital Image Processing. 4. ed. New York: Pearson, 2018.

JIANG, Yu et al. DeepSeedling: Deep convolutional network and Kalman filter for plant seedling detection and counting in the field. Plant Methods, v. 15, n. 1, p. 141, 2019.

KOO, Terry K.; LI, Mae Y. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. Journal of Chiropractic Medicine, v. 15, n. 2, p. 155–163, 2016.

LI, Zewen et al. A survey of convolutional neural networks: analysis, applications, and prospects. IEEE Transactions on Neural Networks and Learning Systems, v. 33, n. 12, p. 6999-7019, 2021.

LOPES, A. C. A.; NASCIMENTO, W. M. Análise de sementes de hortaliças. Brasília, DF: Embrapa Hortaliças, 2009.

MARCOS FILHO, Júlio. Fisiologia de sementes de plantas cultivadas. 2. ed. Londrina: ABRATES, 2015.

MINAEE, Shervin et al. Image segmentation using deep learning: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, v. 44, n. 7, p. 3523-3542, 2022.

ROCHA, C. R. M.; SILVA, V. N. da; CICERO, S. M. Avaliação do vigor de sementes de girassol por meio de análise de imagens de plântulas. Ciência Rural, v. 45, n. 6, p. 970-976, 2015. Disponível em: https://doi.org/10.1590/0103-8478cr20131455. Acesso em: 1 dez. 2025.

WANG, Peng et al. Leaf segmentation using modified YOLOv8-seg models. Life, v. 14, n. 6, p. 780, 2024.

Anexos

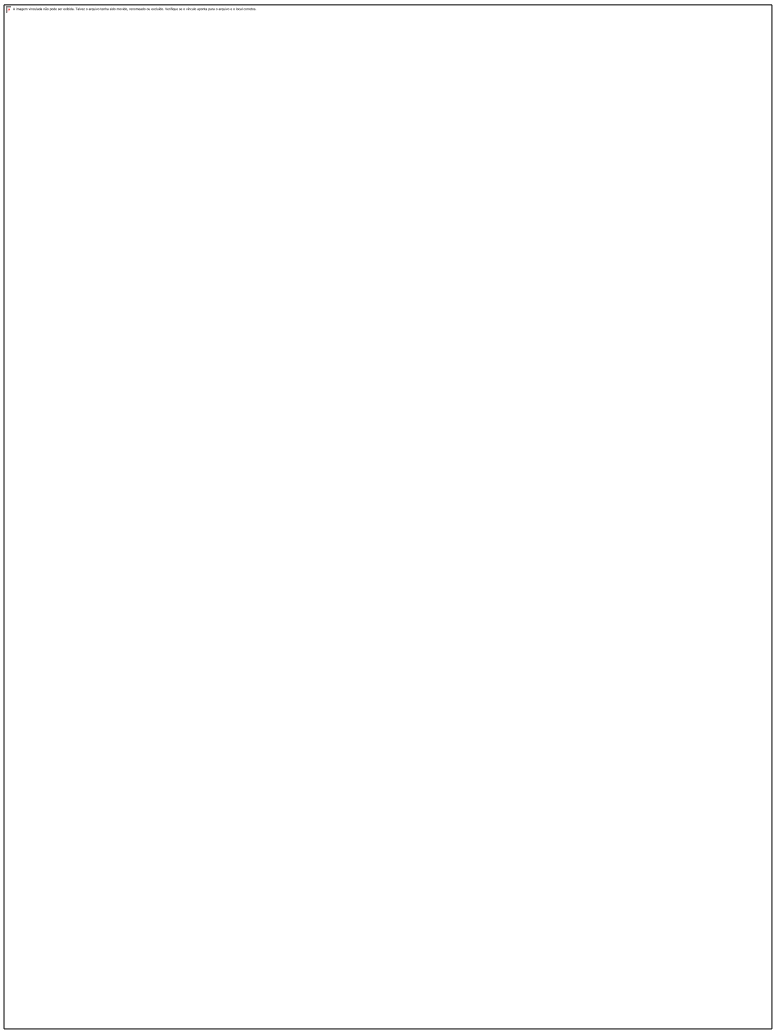


Figura 1 - Estrutura montada para a captura das imagens

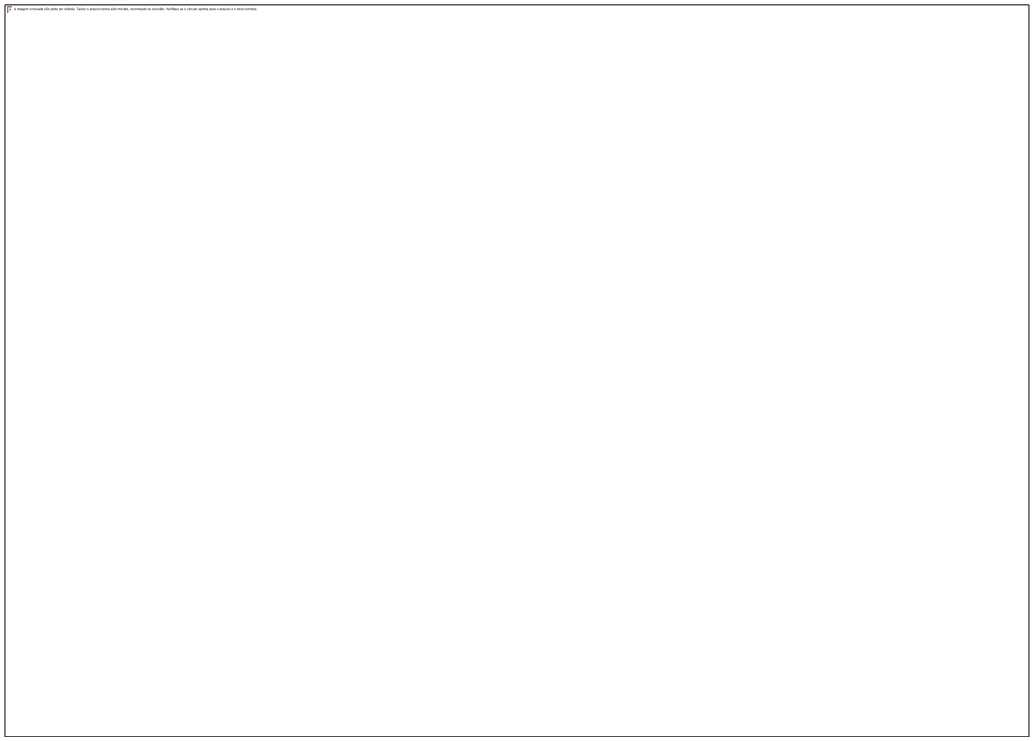


Figura 2 - Gráfico das curvas de perda para o fold 1.

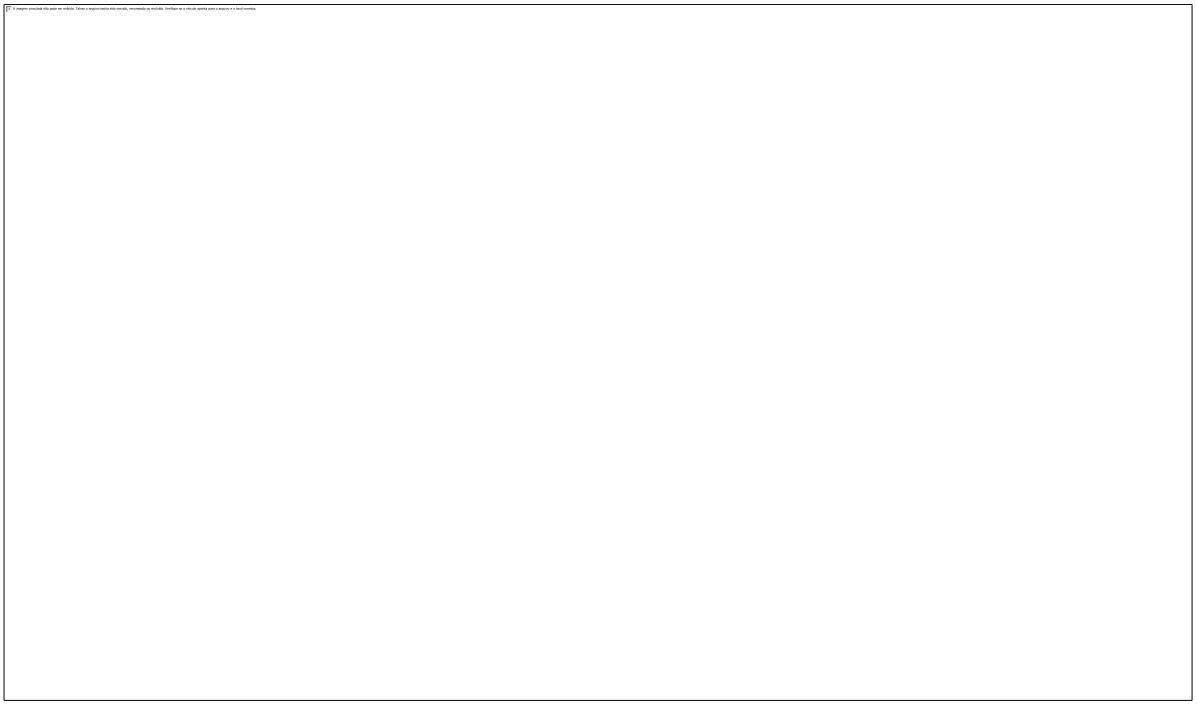


Tabela 1 - Valores utilizados no espaço de busca da otimização bayesiana

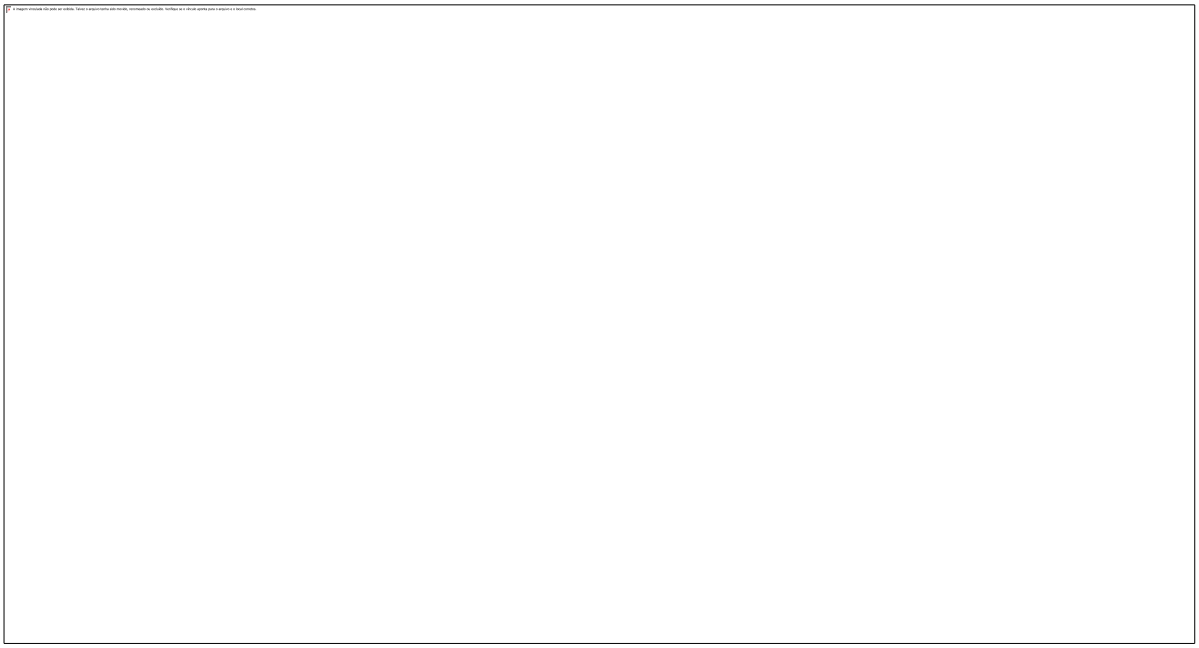


Tabela 2 - Melhores hiperparâmetros definidos para cada fold

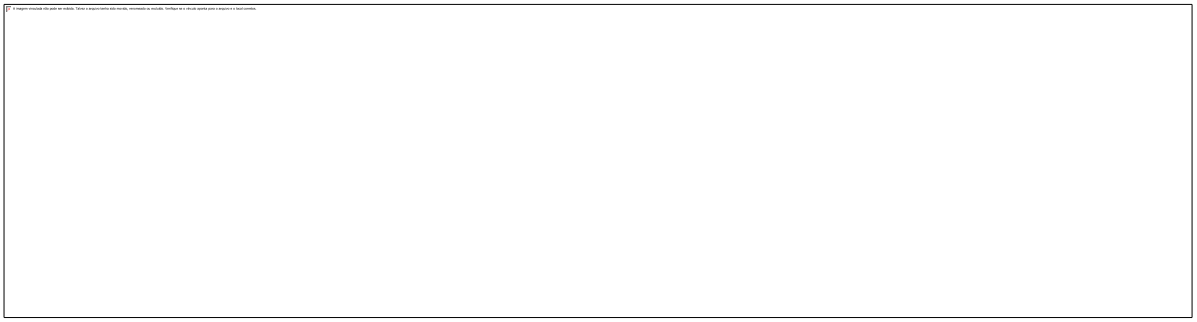


Tabela 3 - Média das métricas avaliativas apresentadas na validação do modelo

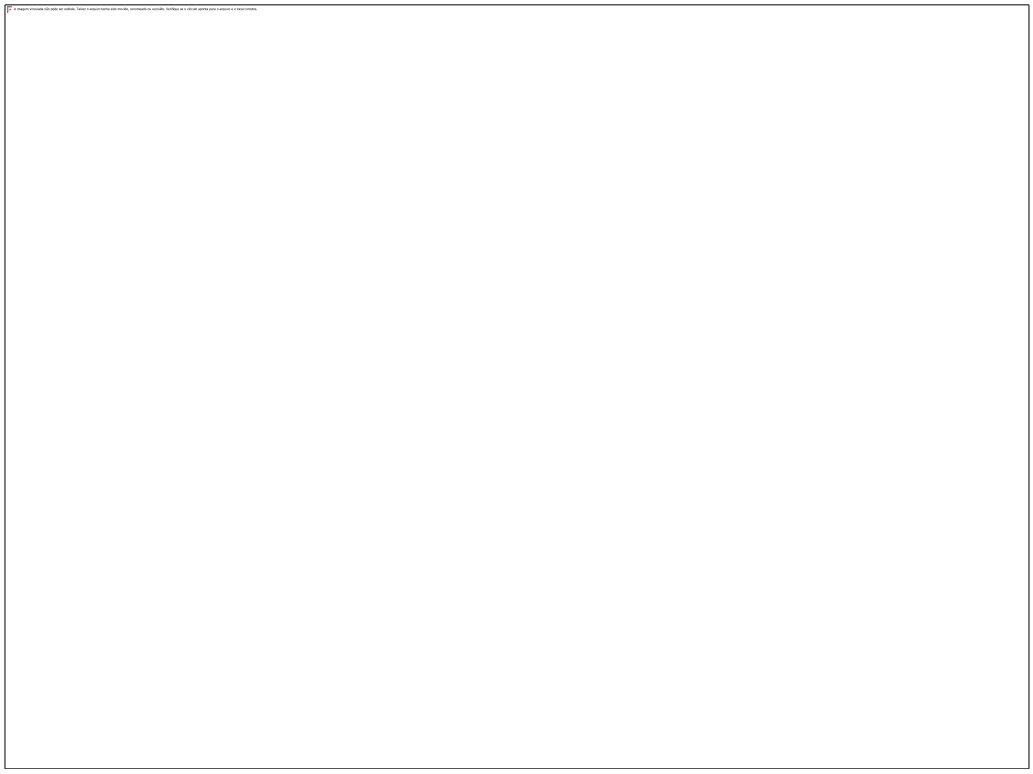


Figura 4 - Gráfico de Bland-Altman para a parte aérea

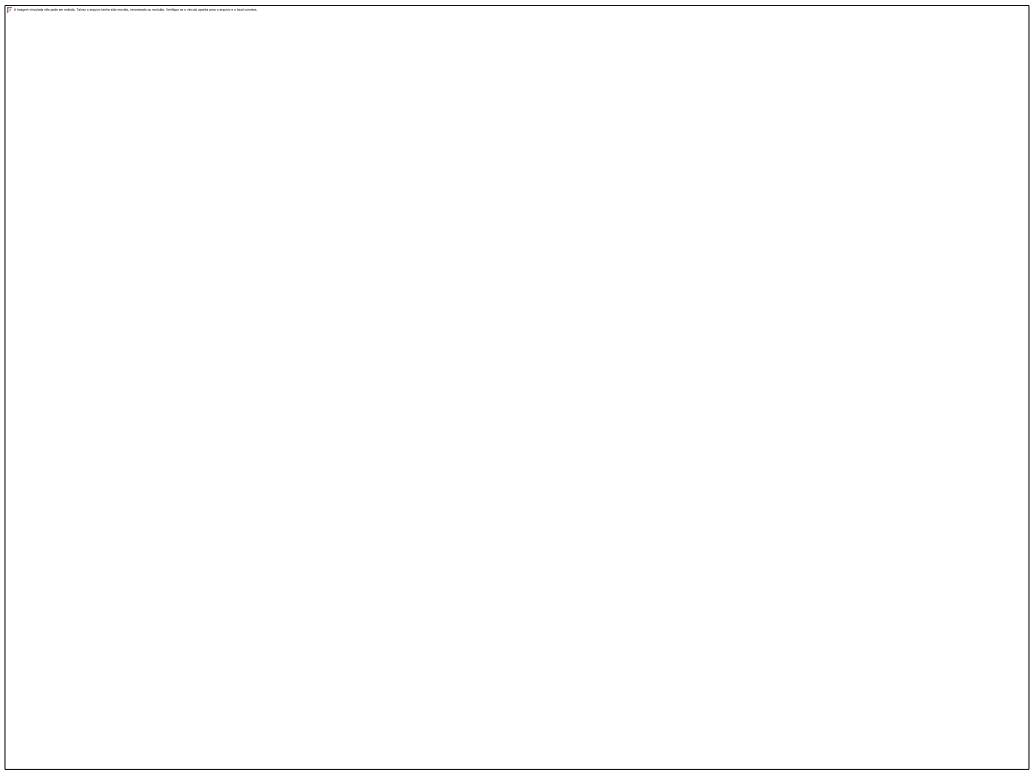


Figura 5 - Gráfico de Bland-Altman para o comprimento total

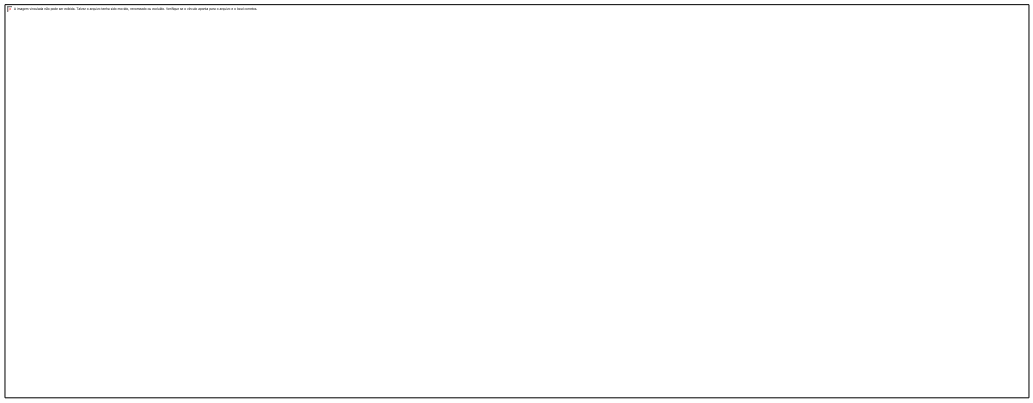


Figura 6 - Comparação entre imagem original e máscara gerada.



Figura 3 - Diagrama de diferença crítica entre as medições manuais e automatizadas

CÓDIGO: SB0159

AUTOR: MATHEUS DE MOURA FERREIRA

COAUTOR: ANTONINO ALVES FEITOSA NETO

COAUTOR: GEAN CARLOS DA SILVA SANTOS

ORIENTADOR: MARCIO DIAS PEREIRA

TÍTULO: Extrato aquoso de Gracilaria birdiae melhora o desempenho fisiológico de sementes de Piptadenia moniliformis sob salinidade

Resumo

A salinidade constitui um dos principais fatores limitantes à germinação e ao estabelecimento inicial de espécies florestais em ambientes semiáridos. Neste estudo, avaliou-se o efeito do condicionamento fisiológico com extrato aquoso da macroalga vermelha Gracilaria birdiae sobre a germinação, o vigor e o crescimento inicial de plântulas de Piptadenia moniliformis submetidas ao estresse salino. As sementes foram condicionadas com cinco concentrações do extrato (0, 5, 10, 15 e 20 mL L^{-1}) e posteriormente submetidas a seis níveis de salinidade induzidos por NaCl (0,0; 1,0; 2,2; 3,2; 4,4 e 5,4 g L^{-1}). Foram avaliadas a germinação, a primeira contagem, o índice de velocidade de germinação, o comprimento e a massa seca de plântulas. A análise de variância evidenciou efeito significativo do extrato sobre a germinação e o crescimento das plântulas, enquanto a salinidade afetou principalmente variáveis relacionadas ao vigor e ao desenvolvimento inicial, não sendo observada interação significativa entre os fatores. As análises multivariadas baseadas no índice de soma de ranks de Mulamba e Mock, na análise de componentes principais e no índice MGIDI identificaram os tratamentos contendo 15 e 20 mL L^{-1} de extrato associados a 1,0 g L^{-1} de NaCl como aqueles de melhor desempenho fisiológico global. Os resultados demonstram o potencial do extrato aquoso de G. birdiae como bioestimulante natural para a produção de mudas e o estabelecimento inicial de P. moniliformis em programas de restaura

Palavras-chave: bioestimulante; Caatinga; estresse salino; macroalga marinha; sementes

TITLE: Aqueous extract of Gracilaria birdiae improves the physiological performance of Piptadenia moniliformis seeds under salinity

Abstract

Salinity is one of the main factors limiting seed germination and seedling establishment in semiarid environments, directly affecting ecological restoration programs involving native forest species. This study evaluated the effects of seed priming with aqueous extract of the red seaweed Gracilaria birdiae on germination, vigor and early seedling growth of Piptadenia moniliformis under saline stress. Seeds were conditioned with five concentrations of the extract (0, 5, 10, 15 and 20 mL L^{-1}) and subsequently subjected to six salinity levels induced by NaCl (0.0, 1.0, 2.2, 3.2, 4.4 and 5.4 g L^{-1}). Germination, first germination count, germination speed index, seedling length and dry mass were evaluated. Analysis of variance indicated significant effects of the extract on germination and seedling growth, whereas salinity significantly affected only first count and seedling length, with no significant interaction between factors. Multivariate analyses based on the Mulamba and Mock rank sum index, principal component analysis and the MGIDI index consistently identified treatments involving 15 and 20 mL L^{-1} of G. birdiae extract combined with 1.0 g L^{-1} NaCl as those presenting the best overall physiological performance. The results demonstrate the potential of G. birdiae extract as a natural biostimulant for improving the production and establishment of P. moniliformis seedlings in restoration programs conducted in salt-affected environments of the Brazilian semiarid region.

Keywords: biostimulant; Caatinga; forest seeds; seed priming; saline stress;

Introdução

A Catanduva (Piptadenia moniliformis Benth.) é uma leguminosa arbórea nativa da Caatinga, com elevada importância ecológica para recomposição vegetal, sistemas agroflorestais, recuperação de áreas degradadas e manutenção da funcionalidade de ambientes semiáridos. Como espécie pioneira, sua capacidade de estabelecimento inicial é decisiva para o sucesso de programas de restauração em regiões sujeitas à irregularidade hídrica, altas temperaturas e crescente salinização dos solos (Albuquerque et al. 2012; Pereira et al. 2016; Alves-de-Oliveira et al. 2023). Nesse contexto, a germinação e o crescimento inicial das plântulas representam fases críticas do ciclo de vida, pois determinam a capacidade da espécie de ocupar ambientes sob estresse e formar populações viáveis em ecossistemas sazonalmente secos (Finch-Savage and Bassel 2016; Wolny et al. 2018).

A salinidade é um dos principais fatores limitantes à germinação e ao estabelecimento de plântulas, especialmente em regiões áridas e semiáridas. O excesso de sais no substrato reduz o potencial hídrico, dificulta a absorção de água pelas sementes e impõe estresse osmótico logo nas primeiras fases da germinação (Munns and Tester 2008; Machado and Serralheiro 2017; Acosta-Motos et al. 2017). Em fases posteriores, o acúmulo de íons Na^{+} eCl^{-} pode causar toxicidade iônica, desbalanço nutricional, alterações na integridade de membranas, aumento da produção de espécies reativas de oxigênio e redução do crescimento das plântulas (Parihar et al. 2015; Isayenkov and Maathuis 2019; van Zelm et al. 2020; Irik et al. 2024).

Durante a germinação, a absorção de água ocorre em três fases fisiologicamente distintas. A fase I é caracterizada pela rápida embebição; a fase II envolve reativação metabólica, reparo de membranas, síntese de proteínas e mobilização inicial de reservas; e a fase III corresponde à retomada do crescimento embrionário, protrusão da radícula e intensificação da divisão e expansão celular (Bewley 1997; Weitbrecht et al. 2011; Rajjou et al. 2012; Nonogaki 2014; Wolny et al. 2018). Sob estresse salino, a redução do potencial hídrico externo pode atrasar ou restringir a embebição, prolongar a fase II e comprometer a transição para a fase III, resultando em menor porcentagem de germinação, redução do índice de velocidade de germinação e menor crescimento inicial de raiz e parte aérea (Panuccio et al. 2014; Pereira et al. 2016; Manono et al. 2025). Portanto, monitorar a cinética de embebição é crucial para identificar a janela temporal exata em que mecanismos adaptativos podem ser estimulados antes da protrusão radicular.

Em sementes de P. moniliformis, estudos prévios demonstraram que a salinidade afeta diretamente a germinação, o vigor e o crescimento inicial das plântulas, com redução progressiva da qualidade fisiológica sob potenciais osmóticos mais negativos (Pereira et al. 2016). Esse comportamento é particularmente relevante para espécies da Caatinga, pois a limitação hídrica e o acúmulo de sais no solo podem comprometer a regeneração natural e a produção de mudas destinadas à restauração ecológica (Albuquerque et al. 2012; Rengasamy 2010; Daliakopoulos et al. 2016). Assim, estratégias capazes de melhorar a germinação sob salinidade podem ampliar o sucesso de estabelecimento de espécies nativas em ambientes degradados ou salinizados.

O condicionamento fisiológico de sementes, ou seed priming, é uma técnica pré-germinativa capaz de modular a hidratação controlada das sementes sem permitir a protrusão completa da radícula. Esse processo permite a ativação parcial do metabolismo germinativo, favorecendo reparo de membranas, reorganização metabólica, ativação antioxidante e maior velocidade de germinação após a semeadura (Paparella et al. 2015; Biswas et al. 2023; Hameed et al. 2025).

Metodologia

- Origem das sementes e procedimentos iniciais

As sementes de Catanduva (Piptadenia moniliformis Benth.) foram coletadas de matrizes localizadas em áreas de vegetação do semiárido do estado do Rio Grande do Norte, Brasil, durante o período de dispersão natural das sementes. Após a coleta, as sementes foram beneficiadas manualmente e selecionadas visualmente quanto à uniformidade de tamanho, coloração e ausência de danos mecânicos ou sinais de deterioração. As sementes foram armazenadas em sacos de papel kraft sob condições refrigeradas (5 ± 2 °C e aproximadamente 60% de umidade relativa) até o início dos experimentos.

O teor de água inicial das sementes foi determinado pelo método da estufa a 105 ± 3 °C por 24 h, seguindo os critérios e diretrizes estabelecidos pelas Regras para Análise de Sementes (BRASIL, 2025). Antes da instalação dos experimentos, as sementes foram

desinfestadas superficialmente por meio de imersão em solução de hipoclorito de sódio a 2,5% por 5 min e, posteriormente, submetidas a três lavagens sucessivas em água destilada.

Preparo do extrato aquoso de Gracilaria birdiae

O extrato aquoso da macroalga marinha Gracilaria birdiae foi preparado utilizando biomassa seca proveniente de material cultivado. O procedimento de extração seguiu a metodologia adaptada para a obtenção de bioestimulantes à base de algas descrita por Rady et al. (2018) .

A biomassa seca da alga foi homogeneizada em água destilada na proporção de 30 g L^{-1} (p v^{-1}), permanecendo sob agitação magnética contínua por 24 h em temperatura ambiente (25 ± 2 °C). Posteriormente, a suspensão foi filtrada em papel de filtro qualitativo para a obtenção do extrato aquoso concentrado, o qual foi mantido sob refrigeração (5 °C) até a sua utilização. As diluições foram preparadas imediatamente antes da aplicação dos tratamentos, originando as seguintes concentrações avaliadas: 0, 5, 10, 15 e 20 mL L^{-1}.

Condicionamento fisiológico das sementes

O condicionamento fisiológico (seed priming) foi realizado utilizando o método entre papel, adaptando o sistema tradicional de incubação em Gerbox. Foram utilizados rolos de papel Germitest® umedecidos com as soluções correspondentes às concentrações do extrato de G. birdiae (0, 5, 10, 15 e 20 mL L^{-1}).

O volume de solução utilizado para o umedecimento correspondeu a 2,5 vezes a massa do papel seco. As sementes permaneceram sob condicionamento pelo período pré-estabelecido de 36 h em câmara de germinação regulada a 25 °C e fotoperíodo de 12 h. Após o período de condicionamento, as sementes foram retiradas dos rolos de papel, submetidas à secagem superficial em temperatura ambiente até atingirem teor de água próximo à condição inicial e imediatamente destinadas aos testes de estresse salino.

Indução do estresse salino

Após o condicionamento fisiológico, as sementes foram submetidas ao estresse salino utilizando soluções de cloreto de sódio (NaCl) preparadas nas concentrações de 0; 1,0; 2,0; 3,2; 4,4; 5,4 e 6,6 g L^{-1}. As soluções salinas foram utilizadas de forma contínua no umedecimento dos substratos em todos os testes de germinação e vigor descritos a seguir.

O experimento foi conduzido em delineamento inteiramente casualizado (DIC), arranjado em esquema fatorial 5 × 7, correspondente a cinco concentrações do extrato de G. birdiae e sete níveis de salinidade, com quatro repetições de 50 sementes por tratamento.

- Germinação e cinética germinativa

O teste de germinação foi conduzido em rolos de papel Germitest® umedecidos com as

respectivas soluções salinas de cada tratamento, utilizando volume equivalente a 2,5 vezes a massa do papel seco. Os rolos foram mantidos em câmara de germinação a 25 °C com fotoperíodo de 12 h. As avaliações foram realizadas diariamente até a estabilização física da germinação. Foram analisadas as seguintes variáveis cinéticas:

- Germinação final (G): Foram contabilizadas as sementes que apresentaram protrusão radicular igual ou superior a 2 mm.

- Primeira contagem de germinação (PCG): Conduzida de acordo com os critérios estabelecidos para a dinâmica germinativa da espécie (BRASIL, 2025), registrando-se a porcentagem de plântulas normais.

- Índice de velocidade de germinação (IVG): Calculado com base nos dados diários de protrusão radicular, conforme a equação matemática proposta por Maguire (1962):

$$IVG = \Sigma (G_i/N_i)$$

Em que: G_i corresponde ao número de sementes germinadas computadas em cada contagem, e N_i representa o número de dias decorridos desde a semeadura. Crescimento de plântulas e acúmulo de biomassa

Ao final do teste de germinação, as plântulas normais de cada repetição foram avaliadas seguindo as diretrizes de avaliação morfológica de Krzyzanowski et al. (2020) . Foram mensurados o comprimento da parte aérea, o comprimento da raiz primária e o comprimento total da plântula com o auxílio de uma régua milimetrada, sendo os resultados expressos em cm plântula^{-1}.

Para a determinação da massa seca das plântulas, os eixos embrionários foram extraídos (com a remoção dos cotilédones), acondicionados em sacos de papel e transferidos para estufa de circulação forçada de ar mantida a 65 °C até atingirem peso constante. Os resultados de biomassa seca foram expressos em mg plântula^{-1}. Análise estatística e seleção multicaracterística dos tratamentos

Os dados foram inicialmente submetidos aos testes de normalidade dos resíduos e homogeneidade das variâncias. Atendidas as pressuposições, realizou-se a análise de variância em esquema fatorial, considerando as concentrações do extrato aquoso de Gracilaria birdiae, os níveis de salinidade e a interação entre esses fatores. A significância dos efeitos foi avaliada pelo teste F, adotando-se o nível de 5% de probabilidade.

Considerando a baixa contribuição da interação entre os fatores e o interesse em avaliar simultaneamente o desempenho fisiológico global de cada combinação de tratamento, foram empregadas análises multivariadas e índices de seleção.

Resultados e Discussões

Análise de variância e resposta geral das sementes de Piptadenia moniliformisao extrato de Gracilaria birdiae e à salinidade

A análise de variância evidenciou que as concentrações do extrato de Gracilaria birdiae exerceram efeito significativo sobre a germinação final (G) e o comprimento de plântulas, ambos ao nível de 1% de probabilidade pelo teste F ($p < 0,01$), enquanto não foram observados efeitos significativos para a primeira contagem de germinação (PCG), índice de velocidade de germinação (IVG) e massa seca das plântulas (Tabela 1). Por sua vez, os níveis de salinidade influenciaram significativamente apenas a primeira contagem de germinação ($p < 0,05$) e o comprimento das plântulas ($p < 0,01$), ao passo que a interação entre as concentrações do extrato e os níveis salinos não apresentou significância estatística para nenhuma das variáveis avaliadas. Os coeficientes de variação variaram entre 12,45 e 18,27%, indicando adequada precisão experimental e boa confiabilidade das estimativas obtidas, valores considerados satisfatórios para experimentos envolvendo fisiologia e vigor de sementes.

O efeito isolado do extrato de algas sobre a germinação e o crescimento das plântulas sugere que o condicionamento fisiológico com G. birdiae promoveu benefícios fisiológicos independentemente do ambiente salino imposto após o tratamento das sementes. Extratos de macroalgas marinhas constituem fontes importantes de compostos bioativos, incluindo aminoácidos, polissacarídeos sulfatados, compostos fenólicos, minerais, osmólitos e moléculas sinalizadoras capazes de modular a ativação metabólica inicial, favorecer o reparo de membranas e aumentar a eficiência do estabelecimento inicial das plântulas sob condições adversas (Khan et al., 2009; Rouphael e Colla, 2020; Raja et al., 2023; Kumar et al., 2024). Dessa forma, os resultados indicam que o período de hidrocondicionamento adotado foi suficiente para promover uma pré-ativação metabólica das sementes, refletindo-se em maior capacidade germinativa e em maior crescimento inicial das plântulas.

Em contraste, a salinidade apresentou efeito relativamente limitado sobre a maior parte das variáveis avaliadas. Embora o estresse salino seja reconhecido como um dos principais fatores limitantes à germinação e ao estabelecimento inicial de plântulas, a magnitude dessa resposta depende fortemente da espécie e do estádio de desenvolvimento considerado (Munns e Tester, 2008; Acosta-Motos et al., 2017; Isayenkov e Maathuis, 2019). No presente estudo, a reduzida sensibilidade da germinação final aos níveis crescentes de NaCl provavelmente está associada às características ecológicas de P. moniliformis, espécie nativa da Caatinga e naturalmente adaptada a ambientes submetidos à irregularidade hídrica, elevada demanda evaporativa e frequente acúmulo de sais nos horizontes superficiais do solo.

Espécies arbóreas do semiárido frequentemente apresentam mecanismos fisiológicos que permitem a manutenção da germinação mesmo sob condições moderadas de estresse osmótico, incluindo maior eficiência no ajuste osmótico, preservação da integridade das membranas celulares e tolerância ao desbalanço iônico durante as fases iniciais do desenvolvimento (Albuquerque et al., 2012; Pereira et al., 2016; Alves-de-Oliveira et al., 2023). Estudos anteriores com a própria espécie já demonstraram que P. moniliformis apresenta capacidade de manter elevadas taxas germinativas mesmo sob potenciais osmóticos negativos, embora ocorram reduções progressivas no vigor e no

crescimento inicial das plântulas à medida que a intensidade do estresse aumenta (Pereira et al., 2016).

Esse comportamento ajuda a explicar a significância observada para a primeira contagem de germinação e para o comprimento das plântulas, mas não para a germinação final. Enquanto a germinação representa uma medida da capacidade máxima de emissão da radícula ao final do teste, a primeira contagem e o crescimento inicial constituem indicadores mais sensíveis do vigor, sendo capazes de detectar alterações sutis na velocidade de mobilização de reservas, na retomada do metabolismo e nos processos de expansão celular (Finch-Savage e Bassel, 2016; Wolny et al., 2018). Dessa forma, os resultados sugerem que o principal efeito da salinidade ocorreu sobre a dinâmica temporal da germinação e sobre o crescimento subsequente das plântulas, e não necessariamente sobre a capacidade final de germinar.

Outro aspecto relevante foi a ausência de interação significativa entre os fatores extrato de algas e salinidade. Estatisticamente, este resultado indica que o comportamento relativo dos tratamentos permaneceu relativamente constante ao longo do gradiente salino avaliado, sem ocorrência de mudanças expressivas na classificação dos tratamentos em função do aumento da concentração de NaCl. Em outras palavras, os benefícios promovidos pelo condicionamento com G. birdiae tenderam a ocorrer de maneira relativamente uniforme nos diferentes ambientes salinos, caracterizando um efeito predominantemente aditivo e não dependente da intensidade do estresse.

Sob o ponto de vista fisiológico, esse comportamento sugere que os compostos bioativos presentes no extrato atuaram principalmente promovendo melhorias gerais na qualidade fisiológica das sementes, e não exclusivamente ativando mecanismos específicos de tolerância ao sal. Resultados semelhantes têm sido reportados para bioestimulantes à base de algas, nos quais os incrementos em germinação e crescimento inicial são observados tanto em condições favoráveis quanto sob estresses abióticos, refletindo um aumento global da competência metabólica das sementes e da eficiência dos processos de estabelecimento inicial (Bhattacharyya et al., 2015; Bulgari et al., 2019; Biswas et al., 2023; Hameed et al., 2025).

Considerando a reduzida contribuição da interação entre os fatores e a predominância de respostas aditivas, optou-se pela utilização de abordagens multivariadas complementares para integrar simultaneamente o comportamento de todas as variáveis avaliadas e identificar os tratamentos mais promissores. Assim, análises baseadas no índice de soma de ranks de Mulamba e Mock, na análise de componentes principais (PCA) e no índice MGIDI foram empregadas para selecionar as combinações de tratamentos com maior desempenho global e maior proximidade do ideótipo desejado para o estabelecimento inicial de P. moniliformis sob condições de salinidade.

Considerando a baixa significância da interação entre os fatores observada na análise de variância e a necessidade de integrar simultaneamente as diferentes variáveis fisiológicas avaliadas, foi empregada a metodologia de soma de ranks proposta por Mulamba e Mock (1978), utilizando pesos diferenciados para germinação (3), primeira contagem de germinação (3), índice de velocidade de germinação (2), comprimento de plântulas (1) e massa seca de plântulas (1) (Figura 1). Essa abordagem permite a seleção simultânea de múltiplos caracteres de interesse, evitando decisões baseadas em uma

única variável e proporcionando uma avaliação mais robusta do desempenho fisiológico global dos tratamentos.

Os resultados evidenciaram ampla variabilidade entre as combinações de extrato e salinidade, com valores do índice variando de 70,0 a 291,0. Os tratamentos A4_S2, A5_S2 e A1_S6 apresentaram os menores valores do índice de Mulamba e Mock, sendo classificados, respectivamente, como primeiro, segundo e terceiro melhores desempenhos globais. O tratamento A4_S2 destacou-se principalmente pela elevada estabilidade entre as variáveis avaliadas, apresentando simultaneamente elevados ranks para germinação, primeira contagem, velocidade de germinação e crescimento inicial. De forma semelhante, A5_S2 apresentou excelente desempenho germinativo e elevada produção de biomassa, enquanto A1_S6 demonstrou comportamento particularmente interessante por manter elevado desempenho mesmo sob uma das maiores intensidades de estresse salino avaliadas.

Por outro lado, os tratamentos associados às maiores concentrações salinas, especialmente aqueles pertencentes ao nível S4, concentraram-se entre os menores desempenhos globais, destacando-se A3_S4, A5_S4 e A1_S4 como os tratamentos mais distantes do ideótipo desejado. Esse resultado indica que, embora a germinação final tenha apresentado relativa estabilidade frente ao estresse salino, os componentes de vigor e crescimento inicial mostraram-se mais sensíveis ao aumento da concentração de NaCl, corroborando observações anteriores para espécies florestais da Caatinga e para a própria Piptadenia moniliformis (Pereira et al., 2016; Alves-de-Oliveira et al., 2023).

A superioridade dos tratamentos contendo maiores concentrações do extrato de Gracilaria birdiae sugere a atuação dos compostos bioativos presentes na alga na manutenção da qualidade fisiológica das sementes e no estabelecimento inicial das plântulas. Polissacarídeos sulfatados, aminoácidos, compostos fenólicos e moléculas sinalizadoras presentes em extratos de macroalgas podem atuar promovendo ajuste osmótico, maior eficiência antioxidante e melhor reorganização metabólica durante a germinação, favorecendo a tolerância a condições adversas (Rouphael e Colla, 2020; Raja et al., 2023; Kumar et al., 2024).

A utilização do índice de Mulamba e Mock mostrou-se particularmente adequada para este estudo por permitir a identificação de tratamentos com desempenho equilibrado entre os diferentes componentes do vigor, característica essencial em programas de produção de mudas e restauração ecológica, nos quais a germinação isoladamente nem sempre representa o melhor indicador do potencial de estabelecimento em campo. Resultados semelhantes têm sido observados em estudos recentes envolvendo índices multicaracterísticos aplicados à seleção de tratamentos e bioestimulantes em sementes submetidas a estresses abióticos, demonstrando maior eficiência em comparação às análises univariadas tradicionais (Olivoto e Nardino, 2021; Zuffo et al., 2022).

Análise de componentes principais e identificação dos tratamentos superiores

A análise de componentes principais (PCA) permitiu integrar simultaneamente as variáveis de germinação, vigor e crescimento inicial, reduzindo a dimensionalidade dos dados e evidenciando os padrões de associação entre os tratamentos avaliados (Figura 2). Os dois primeiros componentes principais explicaram conjuntamente 60,0% da

variabilidade total dos dados, sendo 38,1% atribuídos ao primeiro componente (PC1) e 21,9% ao segundo componente (PC2), valores considerados adequados para a interpretação biológica de experimentos envolvendo múltiplos caracteres fisiológicos em sementes e plântulas.

O primeiro componente principal apresentou associação positiva principalmente com comprimento e massa seca de plântulas, além de contribuir, em menor intensidade, para a germinação e a primeira contagem de germinação. Dessa forma, tratamentos localizados no quadrante positivo de PC1 apresentaram desempenho superior para crescimento e estabelecimento inicial das plântulas. Em contraste, tratamentos posicionados no lado negativo desse eixo estiveram associados à redução do crescimento e menor acúmulo de biomassa, refletindo maior impacto do estresse sobre o desenvolvimento inicial.

O segundo componente principal mostrou maior associação com o índice de velocidade de germinação (IVG), evidenciando um gradiente relacionado principalmente ao vigor inicial das sementes. A disposição dos vetores indicou ainda forte correlação positiva entre germinação, primeira contagem, comprimento e massa seca, sugerindo que tratamentos que favoreceram a germinação também promoveram maior crescimento e acúmulo de biomassa. Esse comportamento é frequentemente observado em estudos de vigor de sementes, uma vez que a rápida retomada do metabolismo durante a germinação tende a resultar em maior eficiência na mobilização de reservas e no crescimento subsequente das plântulas (Finch-Savage e Bassel, 2016; Rajjou et al., 2012).

Os tratamentos A4_S2, A5_S2 e A1_S6 destacaram-se claramente no quadrante positivo de PC1, posicionando-se próximos aos vetores associados à germinação, vigor e crescimento, indicando elevada proximidade com o ideótipo desejado para estabelecimento inicial sob condições salinas. Esses resultados corroboram diretamente a classificação obtida pelo índice de Mulamba e Mock, reforçando a robustez da seleção multicaracterística e indicando que esses tratamentos apresentaram desempenho superior de forma consistente entre as diferentes metodologias empregadas.

Particularmente interessante foi o comportamento do tratamento A1_S6, que apresentou elevado desempenho mesmo sob um dos maiores níveis de salinidade avaliados. Esse resultado sugere que Piptadenia moniliformis possui mecanismos intrínsecos de tolerância ao estresse osmótico durante a fase germinativa, característica já descrita anteriormente para espécies da Caatinga e para a própria espécie (Albuquerque et al., 2012; Pereira et al., 2016). Por outro lado, os tratamentos A4_S2 e A5_S2 indicam que concentrações intermediárias e elevadas do extrato de Gracilaria birdiae potencializaram a expressão desse potencial fisiológico, favorecendo simultaneamente germinação, vigor e crescimento inicial.

Em oposição, os tratamentos associados ao grupo localizado no quadrante inferior esquerdo, especialmente aqueles envolvendo o nível salino S4, apresentaram os menores desempenhos globais, caracterizados por baixos valores de germinação, crescimento e acúmulo de biomassa. A separação espacial desses tratamentos evidencia que o aumento da intensidade do estresse salino passa a comprometer principalmente processos relacionados à expansão celular e ao crescimento das plântulas, variáveis

reconhecidamente mais sensíveis ao estresse do que a germinação final propriamente dita (Munns e Tester, 2008; Acosta-Motos et al., 2017; van Zelm et al., 2020).

A proximidade dos tratamentos superiores aos vetores de germinação, crescimento e massa seca sugere que os compostos bioativos presentes no extrato de G. birdiae atuaram favorecendo mecanismos fisiológicos relacionados ao ajuste osmótico, manutenção da integridade das membranas e maior eficiência antioxidante durante o estabelecimento inicial das plântulas. Polissacarídeos sulfatados, aminoácidos, compostos fenólicos e moléculas sinalizadoras presentes em extratos de macroalgas têm sido frequentemente associados ao aumento da tolerância a estresses abióticos e ao fortalecimento do metabolismo germinativo (Rouphael e Colla, 2020; Raja et al., 2023; Kumar et al., 2024).

De forma geral, a PCA confirmou os resultados obtidos pela análise univariada e pelo índice de Mulamba e Mock, evidenciando que os tratamentos A4_S2, A5_S2 e A1_S6 apresentaram a melhor combinação entre germinação, vigor e crescimento inicial, constituindo as alternativas mais promissoras para a produção de mudas de P. moniliformis destinadas a programas de restauração e revegetação em ambientes sujeitos à salinização crescente dos solos semiáridos.

Seleção simultânea dos tratamentos pelo índice MGIDI

A análise pelo índice MGIDI (Multi-trait Genotype-Ideotype Distance Index) permitiu a seleção simultânea dos tratamentos considerando todas as variáveis fisiológicas avaliadas e a proximidade em relação ao ideótipo previamente definido, caracterizado por elevados valores de germinação, vigor e crescimento inicial das plântulas (Figura 3). Adotando-se uma intensidade de seleção de 15%, correspondente à escolha dos cinco melhores tratamentos dentre as trinta combinações avaliadas, foram selecionados os tratamentos A4_S1, A5_S3, A4_S4, A1_S2 e A4_S2, os quais apresentaram os menores valores de MGIDI e, consequentemente, maior proximidade do perfil fisiológico ideal.

Os valores do índice variaram de 0,908 a 6,458, evidenciando ampla variabilidade no desempenho fisiológico entre os tratamentos avaliados. O tratamento A4_S1 apresentou a menor distância em relação ao ideótipo (MGIDI = 0,908), indicando elevada estabilidade e desempenho equilibrado entre todas as variáveis consideradas. De forma semelhante, os tratamentos A5_S3, A4_S4 e A4_S2 apresentaram baixos valores do índice, sugerindo que concentrações intermediárias e elevadas do extrato de Gracilaria birdiae favoreceram simultaneamente múltiplos componentes do vigor e do crescimento inicial das plântulas.

A predominância dos tratamentos contendo as maiores concentrações do extrato entre os selecionados pelo MGIDI reforça a hipótese de que os compostos bioativos presentes em G. birdiae atuaram promovendo benefícios fisiológicos amplos, incluindo maior eficiência metabólica durante a germinação, melhor aproveitamento das reservas e maior capacidade de crescimento sob condições de estresse. Extratos de macroalgas são reconhecidamente ricos em polissacarídeos sulfatados, aminoácidos, compostos fenólicos e moléculas sinalizadoras capazes de estimular mecanismos relacionados ao ajuste osmótico, atividade antioxidante e manutenção da integridade celular em ambientes salinos (Rouphael e Colla, 2020; Raja et al., 2023; Kumar et al., 2024).

Por outro lado, os maiores valores do índice foram observados nos tratamentos A2_S5, A3_S6 e A1_S5, indicando maior distância em relação ao ideótipo e menor desempenho fisiológico global. Esses resultados sugerem que determinadas combinações entre níveis salinos mais elevados e concentrações intermediárias do extrato podem não ter sido suficientes para compensar completamente os efeitos negativos do estresse osmótico sobre o desenvolvimento inicial das plântulas.

Embora exista concordância parcial entre os tratamentos selecionados pelo MGIDI e aqueles identificados pelo índice de Mulamba e Mock e pela PCA, observa-se que cada metodologia enfatiza diferentes aspectos do desempenho fisiológico. Enquanto o índice de Mulamba e Mock privilegia a ordenação relativa das variáveis e a PCA destaca os padrões de associação multivariada, o MGIDI considera simultaneamente a distância euclidiana ao ideótipo desejado após a redução da dimensionalidade dos dados, sendo atualmente considerado um dos métodos mais robustos para seleção simultânea de múltiplos caracteres (Olivoto e Nardino, 2021).

A divergência parcial entre os métodos não representa inconsistência estatística, mas sim complementaridade analítica, uma vez que diferentes abordagens multivariadas capturam distintos componentes da variabilidade biológica dos dados. Nesse contexto, a convergência observada para tratamentos como A4_S2 e A4_S1 entre as diferentes metodologias empregadas reforça a elevada estabilidade e superioridade fisiológica dessas combinações, indicando elevado potencial para utilização em programas de produção de mudas e restauração ecológica em ambientes sujeitos à salinização dos solos.

Conclusão

O condicionamento fisiológico com extrato aquoso de Gracilaria birdiae melhorou o desempenho fisiológico de sementes de Piptadenia moniliformis, promovendo maior germinação e crescimento inicial das plântulas.

A salinidade apresentou efeito limitado sobre a germinação final da espécie, evidenciando elevada tolerância durante a fase germinativa, embora variáveis relacionadas ao vigor e ao crescimento tenham sido mais sensíveis ao aumento da concentração salina.

As análises multivariadas identificaram o condicionamento com 15 mL L^{-1} de extrato de G. birdiae associado a 1,0 g L^{-1} de NaCl e 20 mL L^{-1} de extrato associado a 1,0 g L^{-1} de NaCl como as combinações de melhor desempenho global, reunindo elevados valores de germinação, vigor e crescimento inicial das plântulas.

Os resultados demonstram o potencial do extrato de G. birdiae como bioestimulante natural para a produção de mudas e o estabelecimento inicial de P. moniliformis em ambientes sujeitos à salinização dos solos do semiárido brasileiro.

Referências

ACOSTA-MOTOS, J. R.; ORTUÑO, M. F.; BERNAL-VICENTE, A.; DIAZ-VIVANCOS, P.; SANCHEZ-BLANCO, M. J.; HERNANDEZ, J. A. Plant responses to salt stress: adaptive mechanisms. Agronomy, v. 7, n. 1, p. 18, 2017. DOI: https://doi.org/10.3390/agronomy7010018.

ALBUQUERQUE, M. B.; SANTOS, R. C.; LIMA, L. M.; MELO FILHO, P. A.; SANTOS, C.; GONÇALVES, E. P.; FERNANDES, P. D. Behavior of Piptadenia moniliformis Benth. seeds under water stress. Industrial Crops and Products, v. 36, n. 1, p. 316-320, 2012. DOI: https://doi.org/10.1016/j.indcrop.2011.10.023.

ALVES-DE-OLIVEIRA, J.; SILVA, A. C.; MEDEIROS, R.; SANTOS, F. L. Germination and initial growth of Caatinga forest species under abiotic stress. Journal of Arid Environments, v. 208, p. 104870, 2023. DOI: https://doi.org/10.1016/j.jaridenv.2022.104870.

BEWLEY, J. D. Seed germination and dormancy. The Plant Cell, v. 9, n. 7, p. 1055-1066, 1997. DOI: https://doi.org/10.1105/tpc.9.7.1055.

BHATTACHARYYA, D.; BABGOHARI, M. Z.; RATHOR, P.; PRITHIVIRAJ, B. Seaweed extracts as biostimulants in agriculture: past, present and future. Journal of Applied Phycology, v. 27, p. 1931-1956, 2015. DOI: https://doi.org/10.1007/s10811-014-0466-2.

BISWAS, S.; SARAF, S.; RAHMAN, M. Seed priming: a technological intervention to modulate seed germination and seedling vigor under environmental stress. Plant Growth Regulation, v. 99, p. 15-32, 2023. DOI: https://doi.org/10.1007/s10725-022-00892-1.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Regras para análise de sementes. Brasília: MAPA, 2009.

BULGARI, R.; FRANZONI, G.; FERRANTE, A. Biostimulants application in horticultural crops under abiotic stress conditions. Agronomy, v. 9, n. 9, p. 545, 2019. DOI: https://doi.org/10.3390/agronomy9090545.

CALVO, P.; NELSON, L.; KLOEPPER, J. W. Agricultural uses of plant biostimulants. Plant and Soil, v. 383, p. 3-41, 2014. DOI: https://doi.org/10.1007/s11104-014-2131-8.

CRAIGIE, J. S. Seaweed extracts as biostimulants in horticulture and agriculture. Journal of Applied Phycology, v. 23, p. 371-393, 2011. DOI: https://doi.org/10.1007/s10811-010-9560-4.

DALIAKOPOULOS, I. N.; TSANIS, I. K.; KOUTROULIS, A.; KASTE, K. The threat of soil salinity: a European scale review. Science of the Total Environment, v. 573, p. 727-739, 2016. DOI: https://doi.org/10.1016/j.scitotenv.2016.08.177.

DU JARDIN, P. Plant biostimulants: definition, concept, main categories and regulation. Scientia Horticulturae, v. 196, p. 3-14, 2015. DOI: https://doi.org/10.1016/j.scienta.2015.09.021.

FINCH-SAVAGE, W. E.; BASSEL, G. W. Seed vigour and crop establishment: extending performance beyond adaptation. Journal of Experimental Botany, v. 67, n. 3, p. 567-591, 2016. DOI: https://doi.org/10.1093/jxb/erv490.

HAMEED, A.; HUSSAIN, S.; IMRAN, M. Advanced seed priming techniques for sustainable crop production under changing climate scenarios. Journal of Plant Growth Regulation, v. 44, p. 112-129, 2025. DOI: https://doi.org/10.1007/s00344-024-11324-x.

IRIK, S.; KARATAS, A.; UNAL, S. Ion toxicity and membrane lipid peroxidation in woody plants under saline conditions. Environmental and Experimental Botany, v. 218, p. 105580, 2024. DOI: https://doi.org/10.1016/j.envexpbot.2023.105580.

ISAYENKOV, S. V.; MAATHUIS, F. J. M. Plant salinity stress: many unanswered questions remain. Frontiers in Plant Science, v. 10, p. 80, 2019. DOI: https://doi.org/10.3389/fpls.2019.00080.

KHAN, W.; RAYIRATH, U. P.; SUBRAMANIAN, S.; JITHESH, M. N.; RAYORATH, P.; HODGES, D. M.; CRITCHLEY, A. T.; CRAIGIE, J. S.; NORRIE, J.; PRITHIVIRAJ, B. Seaweed extracts as biostimulants of plant growth and development. Journal of Plant Growth Regulation, v. 28, p. 386-399, 2009. DOI: https://doi.org/10.1007/s00344-009-9103-x.

KRZYZANOWSKI, F. C.; VIEIRA, R. D.; FRANÇA-NETO, J. B. Vigor de sementes: conceitos e testes. 2. ed. Londrina: ABRATES, 2020.

KUMAR, M.; ROY, S.; JHA, B. Algal biostimulants: signaling molecules and regulatory pathways in plant abiotic stress tolerance. Algal Research, v. 78, p. 103395, 2024. DOI: https://doi.org/10.1016/j.algal.2024.103395.

MACHADO, R. M. A.; SERRALHEIRO, R. P. Soil salinity: effect on vegetable crop growth and management practices to prevent and mitigate soil salinization. Horticulturae, v. 3, n. 2, p. 30, 2017. DOI: https://doi.org/10.3390/horticulturae3020030.

MAGUIRE, J. D. Speed of germination-aid in selection and evaluation for seedling emergence and vigor. Crop Science, v. 2, p. 176-177, 1962.

MAMEDE, M. E.; SILVA, J. L.; CASTRO, A. M. Sulfated polysaccharides from red seaweed Gracilaria genus: extraction, bioactivity, and agricultural potential. Carbohydrate Polymers, v. 312, p. 120810, 2023. DOI: https://doi.org/10.1016/j.carbpol.2023.120810.

MANONO, J.; ZHOU, Y.; ZHANG, X. Dynamics of seed imbibition phases under saline stress: a physiological and molecular review. Seed Science Research, v. 35, p. 15-28, 2025. DOI: https://doi.org/10.1017/S096025852400012X.

MEDEIROS, A. D.; PEREIRA, M. D. SAPL®: a free software for determining the physiological potential in soybean seeds. Pesquisa Agropecuária Tropical, v. 48, p. 222-228, 2018. DOI: https://doi.org/10.1590/1983-40632018v4852340.

MUNNS, R.; TESTER, M. Mechanisms of salinity tolerance. Annual Review of Plant Biology, v. 59, p. 651-681, 2008. DOI: https://doi.org/10.1146/annurev.arplant.59.032607.092911.

NONOGAKI, H. Seed dormancy and germination: emerging mechanisms and new perspectives. Frontiers in Plant Science, v. 5, p. 564, 2014. DOI: https://doi.org/10.3389/fpls.2014.00564.

OLIVOTO, T.; NARDINO, M. MGIDI: toward an effective multivariate selection in biological experiments. Bioinformatics, v. 37, n. 10, p. 1383-1389, 2021. DOI: https://doi.org/10.1093/bioinformatics/btaa981.

PANUCCIO, M. R.; JACOBSEN, S. E.; AKHTAR, S. S.; MUSCOLO, A. Effect of salt stress on germination and early seedling growth of different quinoa cultivars. Plants, v. 3, n. 1, p. 19-31, 2014. DOI: https://doi.org/10.3390/plants3010019.

PAPARELLA, S.; ARAUJO, S. S.; ROSSI, G.; WIJAYASINGHE, M. M.; CARBONERA, D.; BALESTRAZZI, A. Seed priming: state of the art and new perspectives. Plant Cell Reports, v. 34, p. 1281-1293, 2015. DOI: https://doi.org/10.1007/s00299-015-1784-y.

PARIHAR, P.; SINGH, S.; SINGH, R.; SINGH, V. P.; PRASAD, S. M. Effect of salinity stress on plants and its tolerance strategies: a review. Environmental Science and Pollution Research, v. 22, p. 4056-4075, 2015. DOI: https://doi.org/10.1007/s11356-014-3729-1.

PEREIRA, M. D.; ALVES, F. L.; SILVA, T. R. Germinação e vigor de sementes de Piptadenia moniliformis Benth. sob estresse salino. Ciência Florestal, v. 26, n. 3, p. 1025-1034, 2016. DOI: https://doi.org/10.5902/1980509825044.

PEREIRA, M. D.; DIAS, M.; SANTOS, A. C. Aqueous extract of the red seaweed Gracilaria birdiae as a natural biostimulant for agricultural crops. Journal of Applied Phycology, v. 37, p. 415-427, 2025. DOI: https://doi.org/10.1007/s10811-024-03350-x.

RADY, M. M.; EL-YAZAL, M. A. S.; TAHA, R. S. Extracts of the cyanobacterium Arthrospira platensis as natural biostimulants enhance growth, yield, and physiological performance of plants under stress. Journal of Applied Phycology, v. 30, p. 2341-2354, 2018. DOI: https://doi.org/10.1007/s10811-018-1475-9.

RAJA, R.; MISHRA, A.; JHA, B. Seaweed liquid extracts as biostimulants: modern tools for sustainable agriculture. Trends in Plant Science, v. 28, n. 8, p. 892-905, 2023. DOI: https://doi.org/10.1016/j.tplants.2023.03.012.

RAJJOU, L.; DUVAL, M.; GALLARDO, K.; CATUSSE, J.; BALLY, J.; JOB, C.;

JOB, D. Seed germination and vigor. Annual Review of Plant Biology, v. 63, p. 507-533, 2012. DOI: https://doi.org/10.1146/annurev-arplant-042811-105550.

RENGASAMY, P. Soil processes affecting crop production in salt-affected soils. Functional Plant Biology, v. 37, p. 613-620, 2010. DOI: https://doi.org/10.1071/FP09249.

ROUPHAEL, Y.; COLLA, G. Toward a sustainable agriculture through plant biostimulants: from herbage extract to bio

Anexos

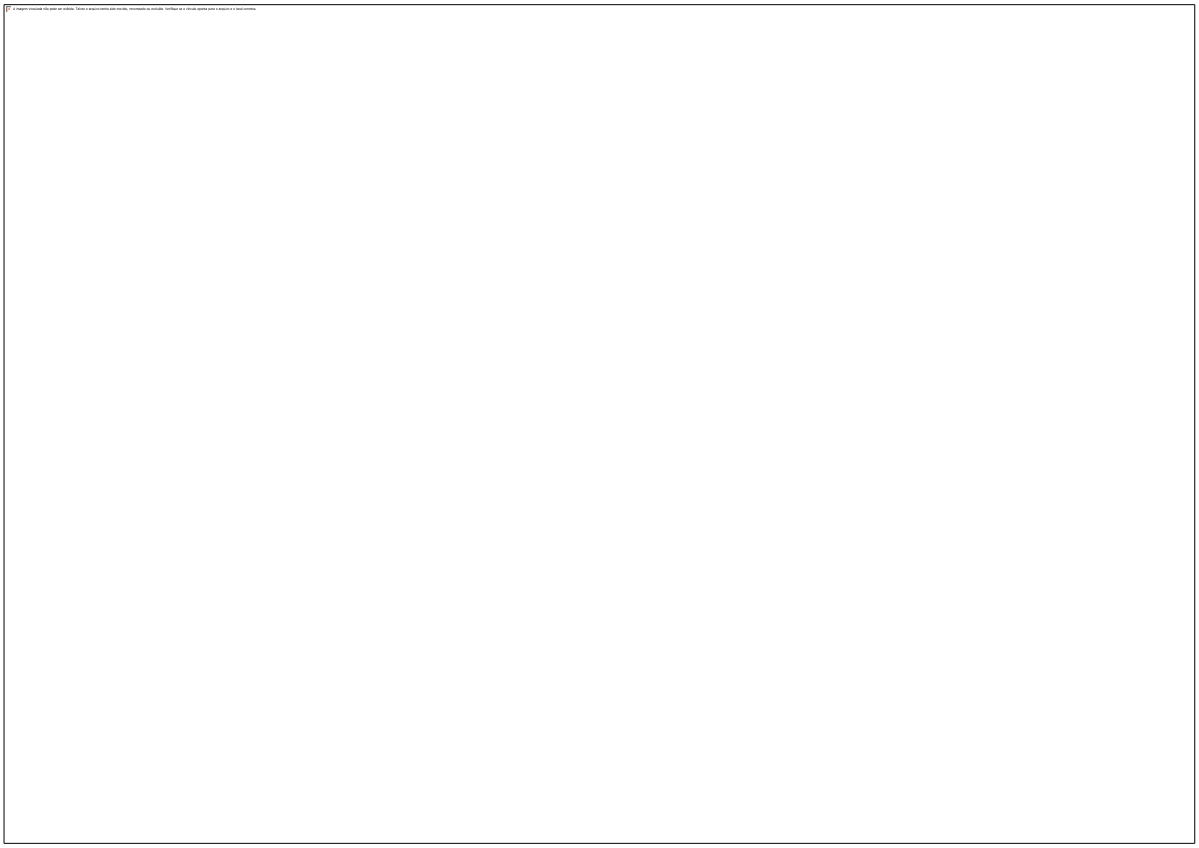


tabela 1. Quadrados médios da análise de variância para germinação, vigor e crescimento inicial de plântulas de Piptadenia moniliformis submetidas ao condicionamento com extrato de Gracilaria birdiae sob diferentes níveis de salinidade.

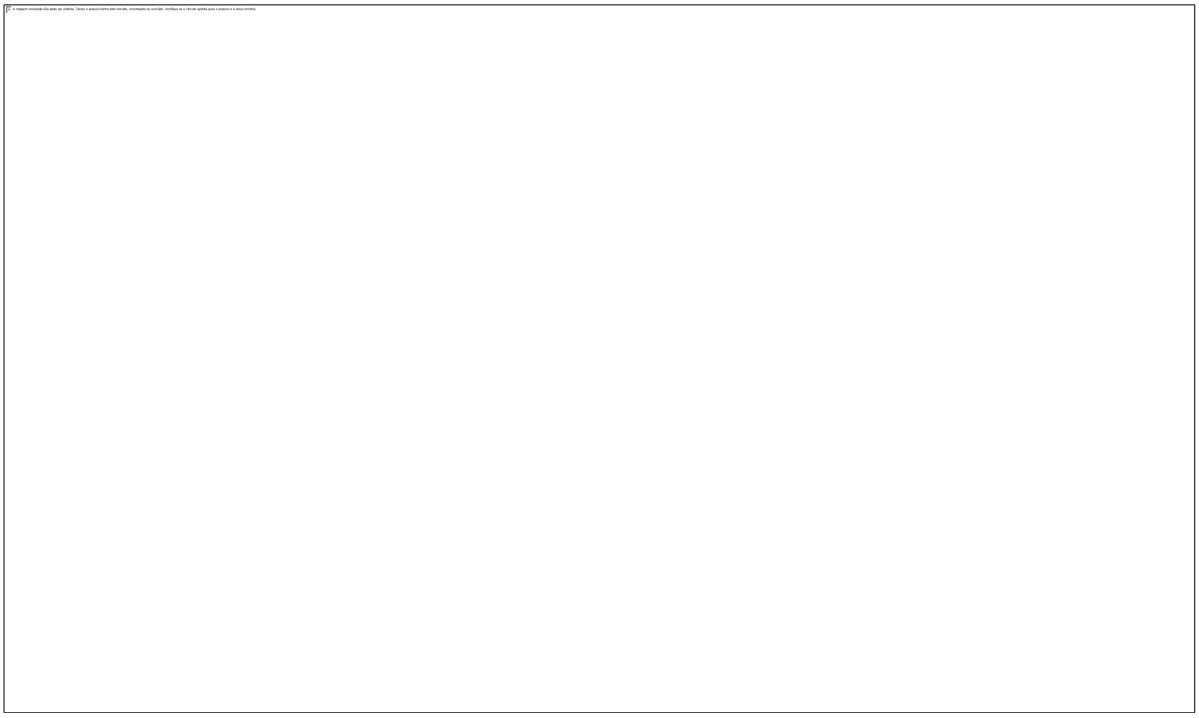


figura 1. Heatmap dos ranks e classificação dos tratamentos pelo índice de soma de ranks de Mulamba e Mock (1978), considerando variáveis de germinação, vigor e crescimento inicial de plântulas de Piptadenia moniliformis submetidas ao condicionamento com



figura 2. Análise de componentes principais (PCA) das variáveis de germinação, vigor e crescimento inicial de plântulas de Piptadenia moniliformis submetidas ao condicionamento com extrato de Gracilaria birdiae sob diferentes níveis de salinidade.

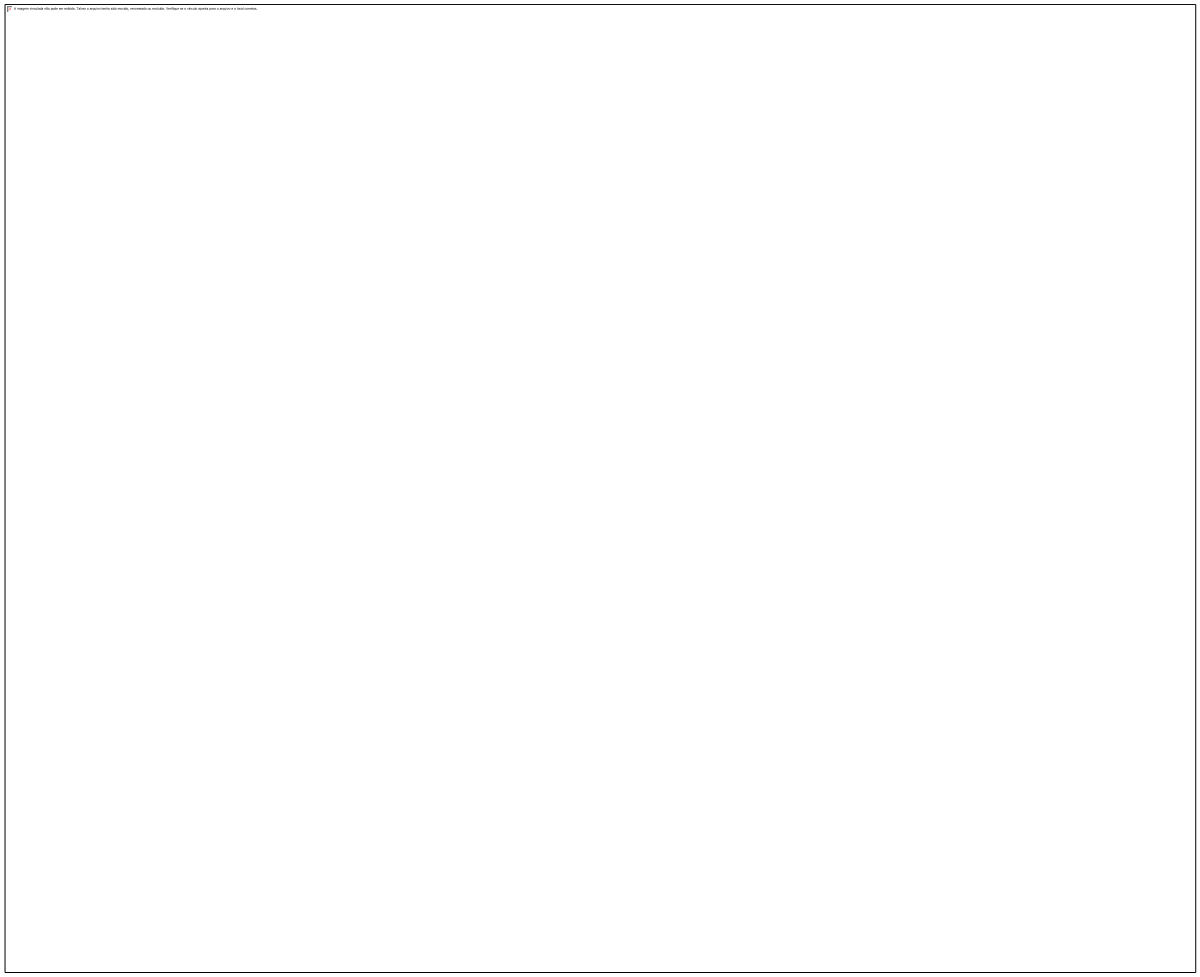


figura 3. Seleção simultânea dos tratamentos pelo índice MGIDI (Multi-trait Genotype-Ideotype Distance Index) considerando variáveis de germinação, vigor e crescimento inicial de plântulas de Piptadenia moniliformis condicionadas com extrato de Gracilaria

CÓDIGO: SB0463

AUTOR: EDSON DE CARVALHO FERREIRA FILHO

COAUTOR: PEDRO MELO ARRUDA

ORIENTADOR: TETSU SAKAMOTO

TÍTULO: Aprendizagem de máquina para identificação sorológica do gênero Leptospira (Leptospiraceae, Bacteria)

Resumo

Leptospira é um gênero altamente diversificado, tradicionalmente classificado por métodos sorológicos em mais de 30 sorogrupos e 300 sorovares. Entretanto, essa classificação pode apresentar resultados ambíguos devido a reações cruzadas entre antígenos, além de exigir métodos laboriosos, culturas vivas e difícil padronização. Para superar essas limitações, compilamos dados genômicos de 721 amostras patogênicas de Leptospira, provenientes dos bancos NCBI RefSeq e BIGSdb (Institut Pasteur), para desenvolver modelos de aprendizado de máquina capazes de prever a classificação sorológica a partir de dados genômicos. A abordagem utiliza o locus rfb, associado à biossíntese do antígeno-O do lipopolissacarídeo, em duas etapas: inicialmente, as amostras são agrupadas em quatro classes sorológicas e, posteriormente, classificadas em seus respectivos sorogrupos. Os modelos apresentaram alto desempenho, com desempenho máximo na primeira etapa e F1 score médio de 0,948 na segunda. A análise de importância revelou que a discriminação dos sorogrupos é determinada por padrões combinatórios de presença e ausência de genes no locus rfb. Com base na coerência genética e nas características antigênicas compartilhadas, propomos o termo “soroclasse” para esses agrupamentos. A abordagem oferece uma alternativa escalável e reprodutível aos métodos sorológicos tradicionais, com potencial aplicação em vigilância epidemiológica, investigação de surtos e desenvolvimento de vacinas.

Palavras-chave: Leptospira; Microbiologia; Leptospirose; Aprendizado de máquina;

TITLE: Machine learning for the serological identification of the genus Leptospira (Leptospiraceae, Bacteria)

Abstract

Leptospira is a highly diverse genus, traditionally classified by serological methods into more than 30 serogroups and 300 serovars. However, this classification can present ambiguous results due to cross-reactions between antigens, in addition to requiring laborious methods, live cultures and difficult standardization. To overcome these limitations, we compiled genomic data from 721 Leptospira pathogenic samples from the NCBI RefSeq and BIGSdb (Institut Pasteur) databases to develop machine learning models capable of predicting serological classification from genomic data. The approach uses the rfb locus, associated with the biosynthesis of the lipopolysaccharide O-antigen, in two stages: initially, the samples are grouped into four serological classes and, subsequently, classified into their respective serogroups. The models showed high performance, with maximum performance in the first stage and an average F1 score of 0.948 in the second. Importance analysis revealed that serogroup discrimination is determined by combinatorial patterns of presence and absence of genes at the rfb locus. Based on genetic coherence and shared antigenic characteristics, we propose the term “seroclass” for these groupings. The approach offers a scalable and reproducible alternative to traditional serological methods, with potential application in epidemiological surveillance, outbreak investigation and vaccine development.

Keywords: Leptospira; Microbiology; Leptospirosis; Machine learning

Introdução

O gênero Leptospira é composto por bactérias espiroquetas caracterizadas por morfologia helicoidal e alta motilidade. Suas espécies patogênicas são responsáveis pela leptospirose, uma das zoonoses mais difundidas no mundo, transmitida pelo contato com água ou solo contaminado com urina de animais infectados, com desfechos clínicos que variam de febre leve à doença de Weil grave e hemorragia pulmonar.

Existem duas abordagens principais para classificar isolados de Leptospira. A classificação taxonômica, baseada em relações evolutivas, reconhece atualmente 77 espécies distribuídas em quatro clados filogenéticos: P1 (patogênico sensu stricto), P2 (intermediário), S1 (saprófico sensu stricto) e S2 (ambiental). Espécies como L. interrogans, L. kirschneri e L. borgpetersenii estão associadas a doenças humanas e animais, enquanto espécies intermediárias e saprófitas ocupam nichos ambientais. A classificação sorológica, por sua vez, baseia-se em diferenças antigênicas detectadas pelo teste de aglutinação microscópica (MAT) e pelo teste de absorção cruzada por aglutinação (CAAT), que levaram à descrição de mais de 300 sorovares agrupados em 30 sorogrupos. Embora considerados padrão ouro, esses métodos são laboriosos, exigem manutenção de grandes coleções vivas e são propensos à interpretação subjetiva.

Em nível molecular, a diversidade do locus rfb — que codifica enzimas envolvidas na biossíntese do antígeno-O dos lipopolissacarídeos (LPS) — é a principal responsável pela diversidade antigênica de Leptospira, constituindo a base molecular da classificação de sorogrupos e sorovares. Variações na composição e organização gênica desse locus determinam diferenças estruturais no LPS, resultando em perfis sorológicos distintos entre cepas, e representam uma ligação direta entre dados genômicos e classificação sorológica tradicional.

Apesar dos avanços em sequenciamento e genômica comparativa, conectar a diversidade genômica de Leptospira às suas propriedades antigênicas permanece desafiador. Métodos de aprendizado de máquina (ML) surgem como ferramentas poderosas para identificar padrões complexos em dados genômicos de alta dimensão, possibilitando a predição de fenótipos como resistência antimicrobiana, especificidade de hospedeiro ou sorotipo. No entanto, a predição de sorogrupos baseada em ML em Leptospira permanece amplamente inexplorada.

Metodologia

O desenvolvimento do framework de aprendizado de máquina para classificação de sorogrupos de Leptospira foi estruturado em etapas sequenciais e criteriosas. A primeira fase consistiu na coleta e curadoria de dados, em que sequências genômicas e anotações sorológicas foram adquiridas dos bancos de dados NCBI RefSeq e BIGSdb. Foram aplicados critérios de filtragem rigorosos, retendo apenas isolados não mutantes do grupo P1 com montagens de menos de 300 contigs, além da remoção de redundâncias, resolução de inconsistências e exclusão de amostras anômalas. Esse processo resultou em um conjunto principal de 721 genomas, enquanto outros 30 genomas recentes foram separados para constituir um conjunto de validação independente.

Na sequência, procedeu-se à construção da matriz de características genômicas. Cepas representativas de alta qualidade para cada sorogrupo foram selecionadas a fim de extrair as proteínas codificadas no locus rfb. Essas proteínas passaram por uma análise de agrupamento utilizando a ferramenta CD-HIT a 80% de similaridade, visando diminuir a redundância e gerando um total de 549 proteínas representativas. Posteriormente, empregou-se o algoritmo TBLASTN para comparar essas proteínas com as montagens genômicas dos isolados do conjunto de dados. Isso permitiu a criação de uma matriz de entrada onde os valores correspondiam à porcentagem de identidade de aminoácidos corrigida pelo comprimento da query.

Para refinar ainda mais os dados, realizou-se uma etapa final de redução de redundância e exploração visual. As amostras que apresentavam uma correlação extremamente alta (coeficiente de Pearson superior a 0,999) foram agrupadas, mantendo-se apenas um genoma representativo de cada cluster para evitar possíveis vieses de classificação, o que reduziu o conjunto de treinamento para 384 genomas. A estrutura, o agrupamento e a variabilidade desses dados foram então explorados visualmente em um espaço bidimensional por meio de uma Análise de Componentes Principais (PCA).

A etapa de modelagem de aprendizado de máquina foi elaborada como um pipeline de classificação hierárquica dividido em duas etapas, baseando-se em modelos binários de Floresta Aleatória Balanceada (BRF) para discriminar as instâncias entre membros e não-membros de cada grupo-alvo. Na primeira etapa, os modelos processaram o conjunto completo filtrado de 384 amostras para classificar cada genoma em uma das quatro grandes "soroclasses" antigênicas. Já na segunda etapa, os isolados foram classificados em seus respectivos sorogrupos por modelos BRF separados e específicos para cada soroclasse, os quais utilizaram exclusivamente os dados e amostras pertinentes àquele subgrupo.

Por fim, a eficácia preditiva de todos os classificadores foi avaliada detalhadamente através de métricas de exatidão, precisão, recordação e F1-score. As estratégias de avaliação variaram conforme a disponibilidade de dados e o nível de classificação: os modelos da primeira etapa foram testados usando validação cruzada do tipo k=5. Em contrapartida, devido ao baixo número de amostras positivas disponíveis em certos sorogrupos, os modelos da segunda etapa foram validados utilizando a abordagem leave-one-out (LOO).

Resultados e Discussões

Caracterização dos Dados e Análise Exploratória

A matriz de dados consolidada para o treinamento contou com 384 amostras genômicas e 549 proteínas. Estas amostras abrangeram 22 sorogrupos distintos distribuídos em quatro grandes agrupamentos de ordem superior, denominados "soroclasses". A soroclasse I foi a mais volumosa, contendo 138 genomas e nove sorogrupos, com destaque para a prevalência de Icterohaemorrhagiae e Pyrogenes. A soroclasse II agrupou 129 genomas, destacando-se os sorogrupos Pomona e Grippotyphosa; a soroclasse III reuniu 86 genomas, predominantemente do sorogrupo Sejroe; e a soroclasse IV contou com 31 genomas, liderados por Tarassovi. A espécie Leptospira interrogans foi a mais frequente no conjunto, e a análise revelou que a maioria dos sorogrupos abriga mais de uma espécie, evidenciando que a classificação sorológica transcende a estrutura filogenética. Durante a exploração visual via Análise de Componentes Principais (PCA), constatou-se que a variação genômica reflete fielmente a estrutura das soroclasses e gera agrupamentos bem definidos, embora a soroclasse II tenha apresentado leve sobreposição entre sorogrupos como Autumnalis, Pomona e Grippotyphosa.

Desempenho do Modelo de Classificação

O pipeline hierárquico operou de forma precisa em duas etapas. Na primeira, modelos binários de Floresta Aleatória Balanceada (BRF) atribuíram as amostras às suas respectivas soroclasses com desempenho perfeito (100% de acerto) na validação cruzada, refletindo o alto poder de separação genômica deste nível estrutural. Na segunda etapa, responsáveis por alocar a amostra no sorogrupo específico, os modelos atingiram um elevado desempenho preditivo na validação leave-one-out, registrando

uma exatidão média de 0,967, precisão de 0,961 e um F1-score de 0,931. Sorogrupos robustos como Ballum, Icterohaemorrhagiae e Sejroe ultrapassaram a marca de 0,95 no F1-score. Contudo, sorogrupos com representação insuficiente no treinamento, a exemplo de Celledoni, Cynopteri e Panama, sofreram classificações incorretas sistemáticas, demonstrando que a escassez de dados limita a aprendizagem de padrões discriminatórios.

Importância de Variáveis e Validação

A análise de importância das variáveis nos modelos evidenciou que os genes cruciais para a discriminação localizam-se majoritariamente na primeira metade do locus rfb. Estas proteínas — incluindo glicosiltransferases e epimerases — gerenciam a biossíntese de carboidratos, que ditam a variabilidade dos antígenos de superfície. O modelo provou que a classificação não depende apenas da presença de genes exclusivos, mas de um padrão combinatório de presença e ausência gênica. Ao isolar essas variáveis mais informativas, modelos reduzidos mantiveram a alta eficácia preditiva, elevando o F1-score médio da segunda etapa para 0,948. Quando validado em um conjunto independente de 30 novos genomas, o sistema exibiu precisão consistente entre diferentes linhagens. A única falha ocorreu com uma amostra Djasiman (classificada como Grippotyphosa devido à falta de um modelo específico no treinamento), mas o sistema sinalizou incerteza ao atribuir a esta predição uma probabilidade de apenas 68,5%.

Discussão e Implicações Translacionais

Esta pesquisa é pioneira ao empregar aprendizado de máquina para prever sorogrupos de Leptospira utilizando apenas informações genômicas. O sucesso do framework comprova a relação biológica direta entre o fenótipo sorológico e a estrutura do locus rfb, preenchendo a lacuna histórica entre a taxonomia clássica e as análises moleculares. Os autores observam que as técnicas sorológicas tradicionais (MAT e CAAT) sofrem com reações cruzadas e subjetividade, sugerindo que ruídos nas anotações de bancos de dados públicos podem ter mascarado um desempenho algorítmico potencialmente ainda maior.

Apesar de limitações referentes à falta de genomas de variantes raras e à necessidade futura de modelos não supervisionados para descobrir novos sorogrupos, o estudo oferece ferramentas valiosas para a prática diagnóstica. Os genes do locus rfb identificados como os de maior importância discriminatória formam uma base empírica ideal para o desenvolvimento de exames de PCR. Isso permitiria uma tipagem epidemiológica rápida e barata, dispensando o sequenciamento de genomas completos no fluxo clínico diário. Finalmente, a forte coesão genômica evidenciada nos resultados fundamenta a adoção formal do termo "soroclasse", uma nomenclatura necessária para organizar as similaridades antigênicas de ordem superior que a estrutura sorológica atual falha em capturar.

Conclusão

Em conclusão, nossos resultados demonstram que modelos de aprendizado de máquina treinados com variáveis genômicas derivadas do rfb são capazes de reproduzir com precisão a classificação sorológica tradicional da Leptospira. Essa abordagem oferece uma alternativa escalável, padronizada e reprodutível aos ensaios sorológicos e contribui para abrir caminho para um sistema baseado na genômica para a classificação e vigilância deste gênero bacteriano complexo e diversificado.

Referências

ADLER, B.; DE LA PEÑA MOCTEZUMA, A. Leptospira and leptospirosis. Veterinary Microbiology, v. 140, n. 3-4, p. 287-296, 2010. DOI: 10.1016/j.vetmic.2009.03.012.

ALTSCHUL, S. F. et al. Basic local alignment search tool. Journal of Molecular Biology, v. 215, n. 3, p. 403-410, 1990. DOI: 10.1016/S0022-2836(05)80360-2.

BULACH, D. M. et al. Lipopolysaccharide biosynthesis in Leptospira. Journal of Molecular Microbiology and Biotechnology, v. 2, n. 4, p. 375-380, 2000.

CAI, C.-S. et al. Development of O-antigen gene cluster-specific PCRs for rapid typing six epidemic serogroups of Leptospira in China. BMC Microbiology, v. 10, p. 67, 2010. DOI: 10.1186/1471-2180-10-67.

CAIMI, K.; RUYBAL, P. Leptospira spp., a genus in the stage of diversity and genomic data expansion. Infection, Genetics and Evolution, v. 81, p. 104241, 2020. DOI: 10.1016/j.meegid.2020.104241.

CHINCHILLA, D. et al. Phylogenomics of Leptospira santarosai, a prevalent pathogenic species in the Americas. PLoS Neglected Tropical Diseases, v. 17, n. 11, e0011733, 2023. DOI: 10.1371/journal.pntd.0011733.

CIURARIU, E. et al. Diagnostic Advances in Leptospirosis: A Comparative Analysis of Paraclinical Tests with a Focus on PCR. Microorganisms, v. 13, n. 3, p. 667, 2025. DOI: 10.3390/microorganisms13030667.

COSTA, F. et al. Global Morbidity and Mortality of Leptospirosis: A Systematic Review. PLoS Neglected Tropical Diseases, v. 9, n. 9, e0003898, 2015. DOI: 10.1371/journal.pntd.0003898.

DE LA PEÑA-MOCTEZUMA, A.; BULACH, D. M.; ADLER, B. Genetic differences among the LPS biosynthetic loci of serovars of Leptospira interrogans and Leptospira borgpetersenii. FEMS Immunology and Medical Microbiology, v. 31, n. 1, p. 73-81, 2001. DOI: 10.1111/j.1574-695X.2001.tb01589.x.

DE LA PEÑA-MOCTEZUMA, A. et al. Comparative analysis of the LPS biosynthetic loci of the genetic subtypes of serovar Hardjo: Leptospira interrogans subtype Hardjoprajitno and Leptospira borgpetersenii subtype Hardjobovis. FEMS Microbiology Letters, v. 177, n. 2, p. 319-326, 1999.

DOS SANTOS RIBEIRO, P. et al. Environmental Biofilms from an Urban Community in Salvador, Brazil, Shelter Previously Uncharacterized Saprophytic Leptospira. Microbial Ecology, v. 86, n. 4, p. 2488-2501, 2023. DOI: 10.1007/s00248-023-02253-3.

DROUIN, A. et al. Interpretable genotype-to-phenotype classifiers with performance guarantees. Scientific Reports, v. 9, p. 4071, 2019. DOI: 10.1038/s41598-019-40561-2.

FAINE, S. et al.Leptospira and leptospirosis. Melbourne: MediSci, 1999.

FERNANDES, L. G. V. et al. Leptospira sanjuanensis sp. nov., a pathogenic species of the genus Leptospira isolated from soil in Puerto Rico. International Journal of Systematic and Evolutionary Microbiology, v. 72, n. 10, p. 005560, 2022. DOI: 10.1099/ijsem.0.005560.

FERREIRA, L. C. et al. Genetic structure and diversity of the rfb locus of pathogenic species of the genus Leptospira. Life Science Alliance, v. 7, n. 1, e202302478, 2024. DOI: 10.26508/lsa.202302478.

FU, L. et al. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics, v. 28, n. 23, p. 3150-3152, 2012. DOI: 10.1093/bioinformatics/bts565.

GIRAUD-GATINEAU, A. et al. Shaping the future of Leptospira serotyping. Journal of Medical Microbiology, v. 74, p. 002059, 2025. DOI: 10.1099/jmm.0.002059.

GOLDFARB, T. et al. NCBI RefSeq: reference sequence standards through 25 years of curation and annotation. Nucleic Acids Research, v. 53, n. D1, p. D243-D257, 2025. DOI: 10.1093/nar/gkae1038.

HAAKE, D. A.; LEVETT, P. N. Leptospirosis in humans. Current Topics in Microbiology and Immunology, v. 387, p. 65-97, 2015. DOI: 10.1007/978-3-662-45059-8_5.

HAMOND, C. et al. Leptospira gorisiae sp. nov, L. cinconiae sp. nov, L. mgodei sp. nov, L. milleri sp. nov and L. iowaensis sp. nov: five new species isolated from water sources in the Midwestern United States. International Journal of Systematic and Evolutionary Microbiology, v. 75, p. 006595, 2025. DOI: 10.1099/ijsem.0.006595.

HUNTER, J. D. Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering, v. 9, n. 3, p. 90-95, 2007. DOI: 10.1109/MCSE.2007.55.

JOLLEY, K. A.; MAIDEN, M. C. BIGSdb: Scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics, v. 11, p. 595, 2010. DOI: 10.1186/1471-2105-11-595.

KALAMBAHETI, T. et al. Genetic organization of the lipopolysaccharide O-antigen biosynthetic locus of Leptospira borgpetersenii serovar Hardjobovis. Microbial Pathogenesis, v. 27, n. 2, p. 105-117, 1999. DOI: 10.1006/mpat.1999.0285.

KO, A. I.; GOARANT, C.; PICARDEAU, M. Leptospira: the dawn of the molecular genetics era for an emerging zoonotic pathogen. Nature Reviews Microbiology, v. 7, n. 10, p. 736-747, 2009. DOI: 10.1038/nrmicro2208.

KORBA, A. A. et al. Leptospira ainlahdjerensis sp. nov., Leptospira ainazelensis sp. nov., Leptospira abararensis sp. nov. and Leptospira chreensis sp. nov., four new species isolated from water sources in Algeria. International Journal of Systematic and Evolutionary Microbiology, v. 71, n. 11, p. 005148, 2021. DOI: 10.1099/ijsem.0.005148.

LEES, J. A. et al. Improved Prediction of Bacterial Genotype-Phenotype Associations Using Interpretable Pangenome-Spanning Regressions. mBio, v. 11, n. 4, e01344-20, 2020. DOI: 10.1128/mBio.01344-20.

LEHMANN, J. S. et al. Leptospiral pathogenomics. Pathogens, v. 3, n. 2, p. 280-308, 2014. DOI: 10.3390/pathogens3020280.

LEVETT, P. N. Leptospirosis. Clinical Microbiology Reviews, v. 14, n. 2, p. 296-326, 2001. DOI: 10.1128/CMR.14.2.296-326.2001.

MEDEIROS, E. J. S. et al. Genetic basis underlying the serological affinity of leptospiral serovars from serogroups Sejroe, Mini and Hebdomadis. Infection, Genetics and Evolution, v. 103, p. 105345, 2022. DOI: 10.1016/j.meegid.2022.105345.

MITCHISON, M. et al. Identification and characterization of the dTDP-rhamnose biosynthesis and transfer genes of the lipopolysaccharide-related rfb locus in Leptospira interrogans serovar Copenhageni. Journal of Bacteriology, v. 179, n. 4, p. 1262-1267, 1997.

NIEVES, C. et al. Horizontal transfer of the rfb cluster in Leptospira is a genetic determinant of serovar identity. Life Science Alliance, v. 6, n. 2, e202201480, 2022. DOI: 10.26508/lsa.202201480.

PEDREGOSA, F. et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, v. 12, p. 2825-2830, 2011.

PICARDEAU, M. Virulence of the zoonotic agent of leptospirosis: still terra incognita? Nature Reviews Microbiology, v. 15, n. 5, p. 297-307, 2017. DOI: 10.1038/nrmicro.2017.5.

WORLD HEALTH ORGANIZATION; INTERNATIONAL LEPTOSPIROSIS SOCIETY. Human leptospirosis: guidance for diagnosis, surveillance and control. Geneva: World Health Organization, 2003.

VINCENT, A. T. et al. Revisiting the taxonomy and evolution of pathogenicity of the genus Leptospira through the prism of genomics. PLoS Neglected Tropical Diseases, v. 13, n. 5, e0007270, 2019. DOI: 10.1371/journal.pntd.0007270.

WASKOM, M. seaborn: statistical data visualization. Journal of Open Source Software, v. 6, n. 60, p. 3021, 2021. DOI: 10.21105/joss.03021.

WENDERLEIN, J. et al. Detection and Identification of Pathogenic Leptospira spp. Serogroups in Europe between 2017 and 2020 Applying a Novel Gene-Based Molecular Approach. Transboundary and Emerging Diseases, v. 2024, p. 1101841, 2024. DOI: 10.1155/2024/1101841.

Anexos

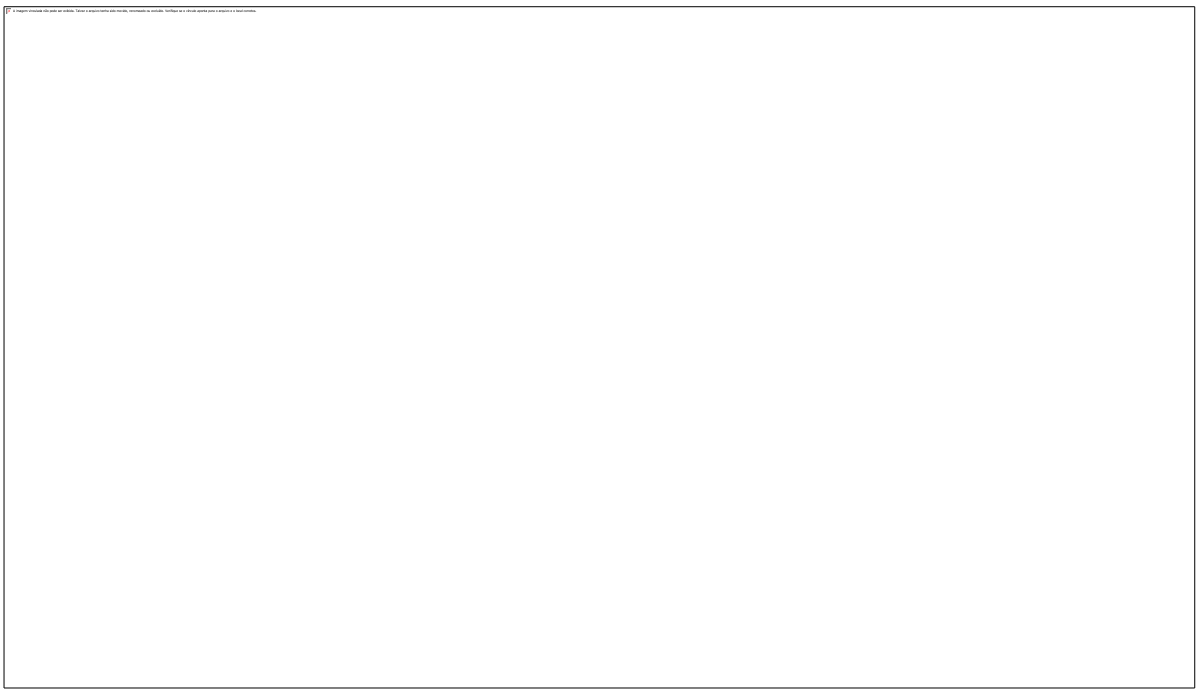


Tabela 1. Amostras representativas de Leptospira para cada sorogrupo.

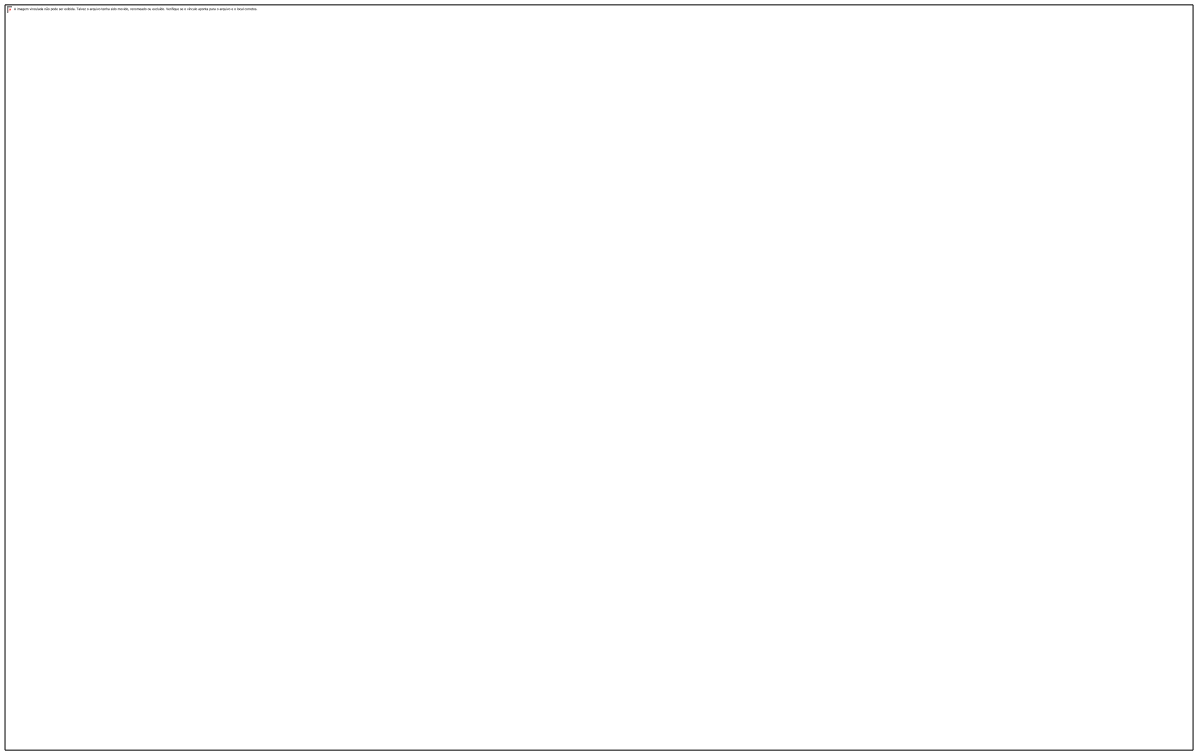


Tabela 2: Distribuição dos isolados de Leptospira entre as classes, sorogrupos e espécies incluídas nas análises de aprendizado de máquina.



Tabela 3: Desempenho dos modelos da segunda etapa (classificação por sorogrupo)

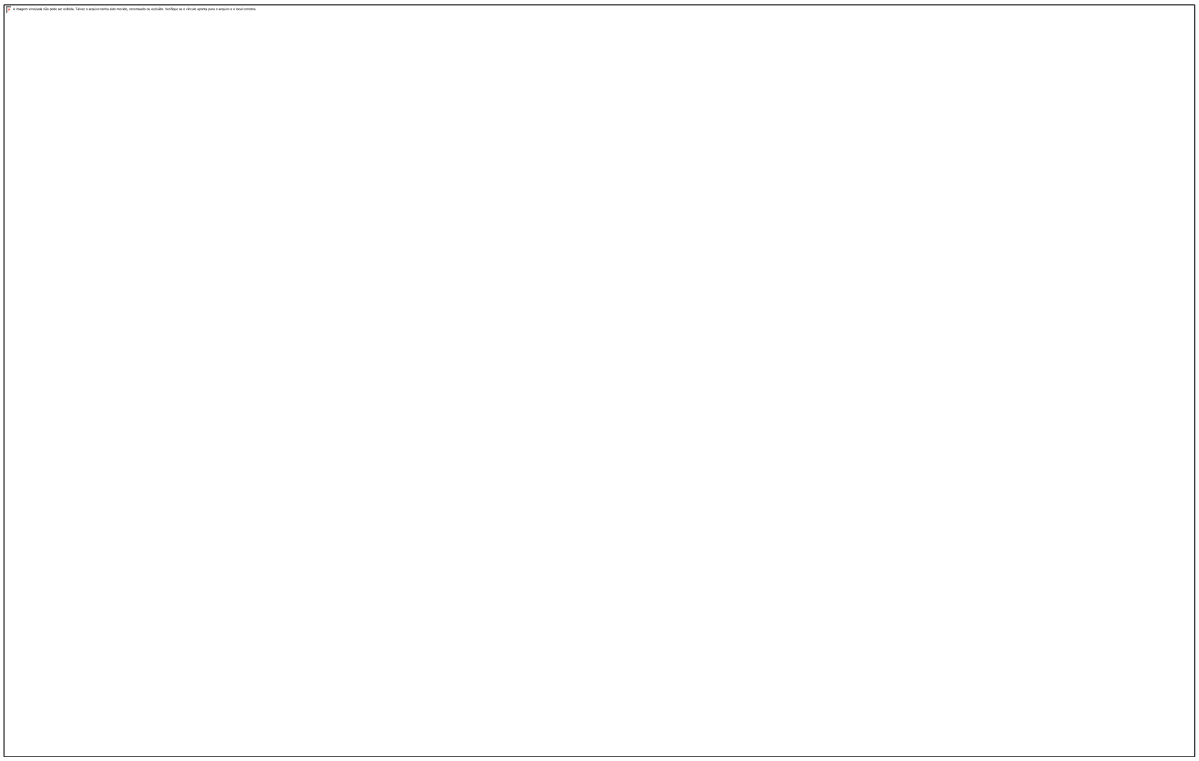


Tabela 4: Matriz de classificação dos modelos de sorogrupos de Leptospira e número de amostras positivas.

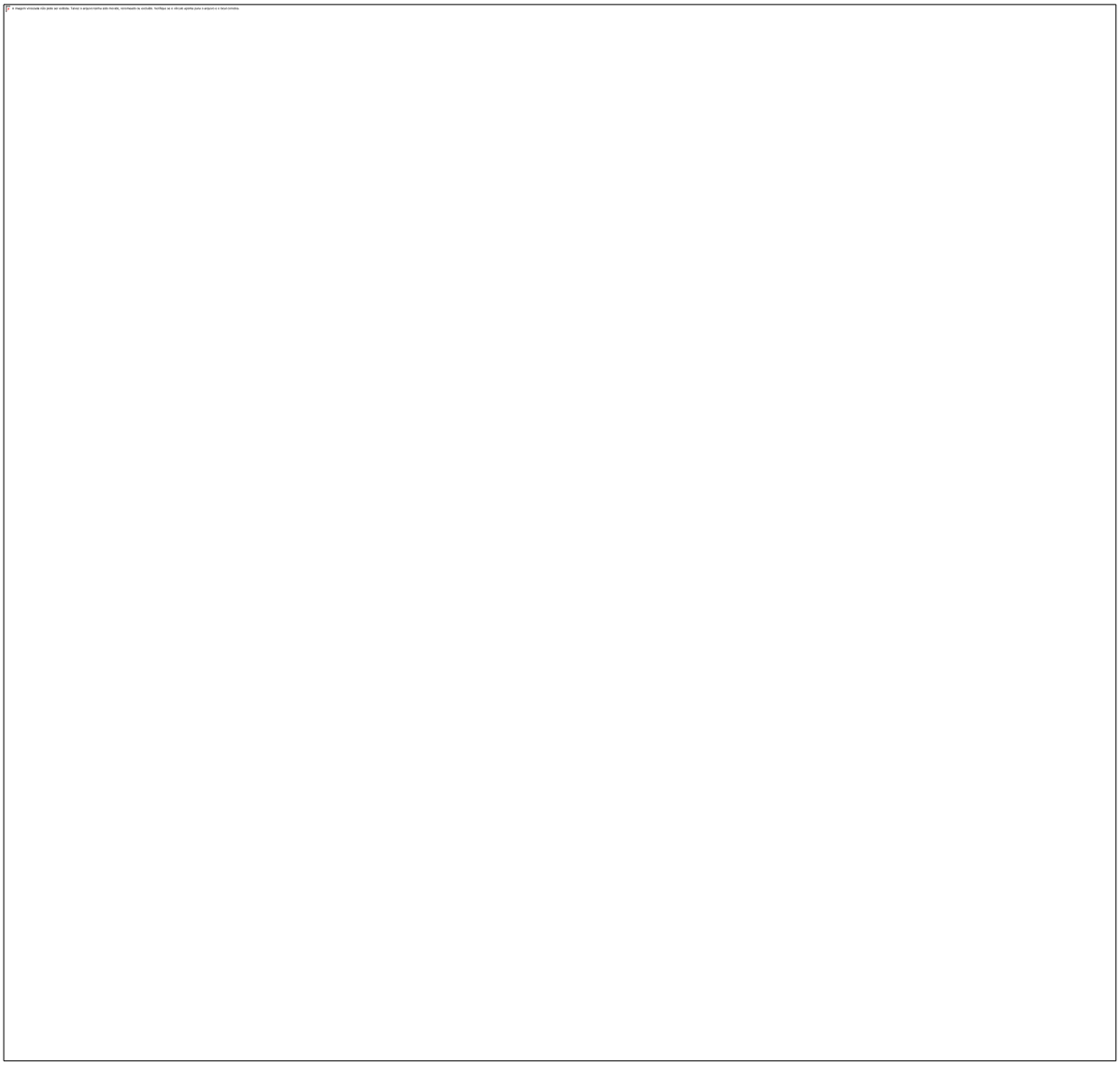


Figura 2: Perfis de importância das variáveis para cada modelo específico de sorogrupo.

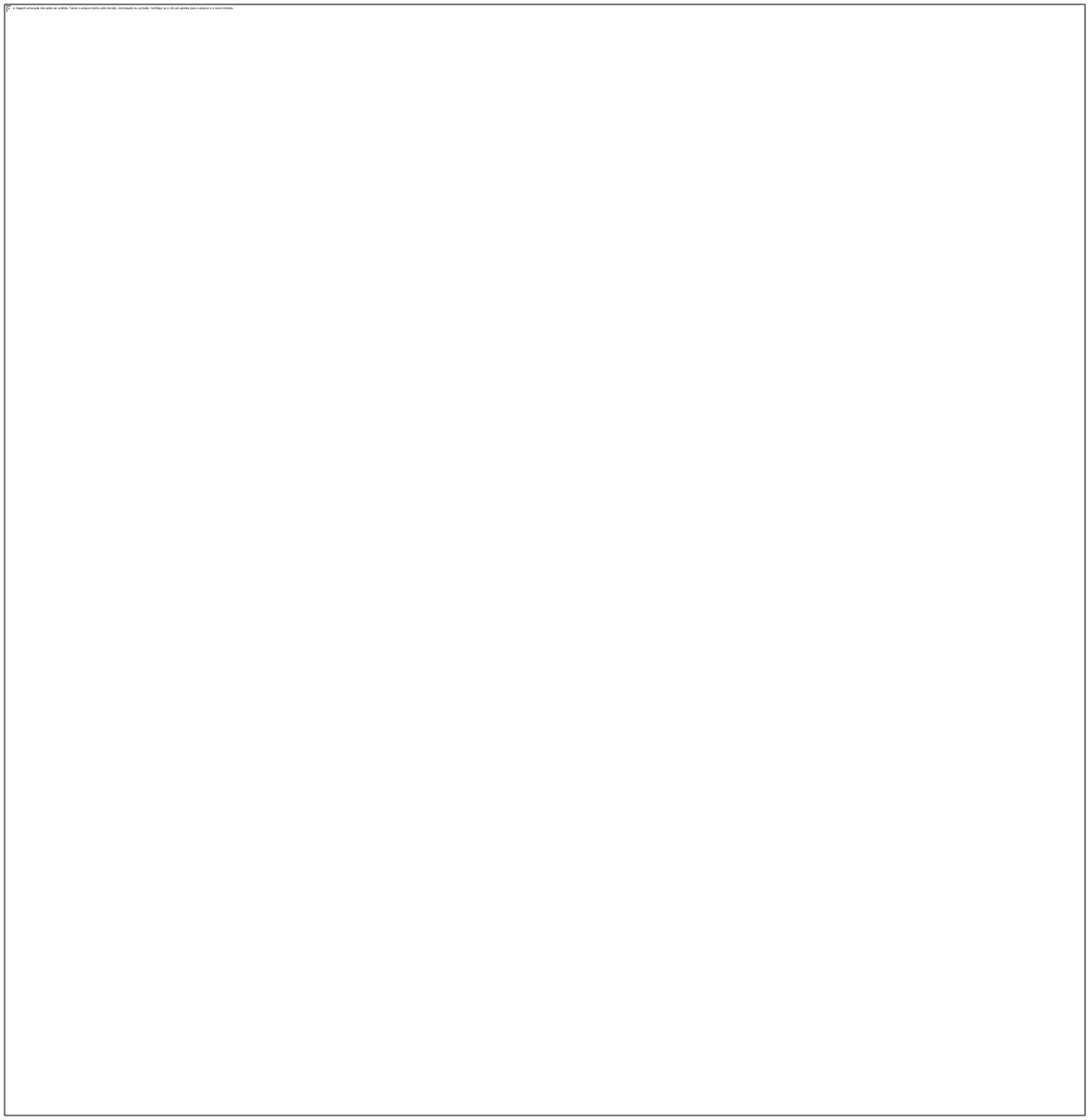


Figura 1: Análise de Componentes Principais (PCA) das características genômicas do locus rfb de Leptospira.



Tabela 5:Classificação de amostras de leptospiras do conjunto de dados de validação

CÓDIGO: SB0807

AUTOR: ANA BEATRIZ CAMILO DA COSTA

ORIENTADOR: JORGE ESTEFANO SANTANA DE SOUZA

TÍTULO: Adaptação e Expansão do Pipeline alignment para Análise de Genomas Eucariotos em Ambiente de Desenvolvimento Local

Resumo

A bioinformática tem desempenhado um papel fundamental na análise de dados biológicos em larga escala, impulsionada pelo avanço das tecnologias de sequenciamento de nova geração. Nesse contexto, pipelines automatizadas tornam-se essenciais para garantir eficiência, reprodutibilidade e escalabilidade no processamento de dados genômicos. Este trabalho tem como objetivo estudar, adaptar e desenvolver uma pipeline de alinhamento voltada à análise de genomas de organismos eucariotos, tomando como base o workflow wf-alignment, originalmente desenvolvido para organismos procariontes. A proposta busca adequar essa solução às necessidades do Centro Multiusuário de Bioinformática (BioME) do Instituto Metrópole Digital (IMD/UFRN), ampliando sua aplicabilidade em projetos de pesquisa e atividades didáticas. Para isso, são realizadas análises da arquitetura do pipeline original, adaptação de seus processos para ambientes de execução locais, integração de ferramentas específicas para genomas eucariotos e validação por meio de testes com diferentes conjuntos de dados. Como resultado esperado, pretende-se disponibilizar uma pipeline modular, flexível e escalável, capaz de processar múltiplas amostras de forma eficiente, contribuindo para o fortalecimento da infraestrutura computacional do BioME e para o desenvolvimento de pesquisas na área de bioinformática.

Palavras-chave: Bioinformática; Pipeline de Bioinformática; Genoma;

TITLE: Adaptation and Expansion of the Alignment Pipeline for Eukaryotic Genome Analysis in a Local Development Environment

Abstract

Bioinformatics has played a fundamental role in large-scale biological data analysis, driven by advancements in next-generation sequencing technologies. In this context, automated pipelines are essential for ensuring efficiency, reproducibility, and scalability in genomic data processing. This work aims to study, adapt, and develop an alignment pipeline for the analysis of eukaryotic genomes, based on the *wf-alignment* workflow originally developed for prokaryotes. The proposal seeks to tailor this solution to the needs of the Bioinformatics Multi-user Center (BioME) at the Institute Metrópole Digital (IMD/UFRN), expanding its applicability to research projects and educational activities. To achieve this, the study involves analyzing the original pipeline's architecture, adapting its processes for local execution environments, integrating tools specific to eukaryotic genomes, and validating the system through tests with various datasets. The expected outcome is a modular, flexible, and scalable pipeline capable of efficiently processing multiple samples, thereby contributing to the strengthening of BioME's computational infrastructure and the advancement of bioinformatics research.

Keywords: Bioinformatics; Bioinformatics pipeline; Genome;

Introdução

O crescimento exponencial da produção de dados biológicos, impulsionado pelas tecnologias de sequenciamento de nova geração (GOODWIN, S.; McPherson, J. D.; MCCOMBIE, W. R., 2016), tem transformado a bioinformática em uma área indispensável para a pesquisa científica. O processamento e a interpretação desses dados exigem ferramentas computacionais capazes de executar análises complexas de forma automatizada, reprodutível e eficiente. Nesse cenário, o desenvolvimento e a adaptação de pipelines de análise genômica representam uma estratégia essencial para otimizar fluxos de trabalho e reduzir o tempo necessário para obtenção de resultados confiáveis (DI TOMMASO, P. et al, 2017).

O Centro Multiusuário de Bioinformática (BioME), vinculado ao Instituto Metrópole Digital (IMD) da Universidade Federal do Rio Grande do Norte (UFRN), atua como um importante centro de apoio à pesquisa, oferecendo infraestrutura computacional e suporte técnico para projetos que envolvem bioinformática. Entre suas atividades estão a capacitação de pesquisadores e estudantes, bem como o desenvolvimento e a manutenção de ferramentas voltadas à análise de dados biológicos.

Entre as soluções disponíveis para análise de sequências longas, destaca-se o wf-alignment, um workflow desenvolvido pelo projeto EPI2ME, que automatiza as etapas de alinhamento de leituras genômicas. Entretanto, sua implementação é voltada principalmente para organismos procariontes, como bactérias, limitando sua utilização em estudos envolvendo organismos eucariotos. Esses organismos apresentam genomas significativamente mais complexos, caracterizados por maior tamanho, presença de íntrons, regiões repetitivas e maior diversidade estrutural, exigindo ferramentas e parâmetros específicos para garantir alinhamentos precisos.

Diante dessa limitação, este trabalho propõe o estudo da arquitetura do wf-alignment e sua adaptação para suportar análises de genomas eucariotos. O projeto contempla a compreensão do funcionamento interno da pipeline, a reestruturação de seus módulos, a integração de ferramentas apropriadas para esse tipo de análise e a adequação à infraestrutura computacional do BioME. Além disso, busca-se tornar a solução mais modular, escalável e de fácil manutenção, permitindo sua expansão para diferentes aplicações e facilitando sua utilização por pesquisadores e estudantes.

Espera-se que os resultados deste trabalho contribuam para ampliar a capacidade de processamento de dados genômicos no BioME, promovendo maior autonomia na execução de análises, redução da dependência de ferramentas externas e fortalecimento das pesquisas desenvolvidas no instituto. Adicionalmente, a pipeline desenvolvida poderá servir como base para futuras extensões, incorporando novas ferramentas e metodologias de análise à medida que surgirem avanços na área da bioinformática.

Metodologia

Este trabalho caracteriza-se como uma pesquisa aplicada de natureza experimental, fundamentada em técnicas de engenharia reversa, desenvolvimento incremental e validação computacional. O objetivo inicial consistiu em compreender o funcionamento interno do workflow wf-alignment, desenvolvido pelo projeto EPI2ME, para posteriormente projetar e implementar uma nova pipeline voltada à análise de organismos eucariotos, adequada às necessidades do Centro Multiusuário de Bioinformática (BioME).

Investigação do workflow de referência

A primeira etapa do trabalho consistiu em um estudo aprofundado do workflow wf-alignment, utilizado como referência arquitetural para o desenvolvimento da nova solução. Inicialmente foram analisados sua documentação, parâmetros de execução, organização dos módulos, dependências e fluxo geral de processamento. O objetivo desta etapa não foi modificar diretamente o código-fonte existente, mas compreender sua estrutura interna, identificar suas principais etapas computacionais e avaliar quais componentes poderiam ser reutilizados conceitualmente.

Em seguida foram realizados diversos experimentos executando o workflow com diferentes conjuntos de dados, permitindo observar seu comportamento durante todo o fluxo de execução. A inspeção dos diretórios temporários produzidos pelo Nextflow possibilitou acessar os scripts .command.sh, arquivos responsáveis por armazenar os comandos efetivamente executados em cada processo da pipeline. Essa análise permitiu reconstruir detalhadamente a sequência operacional utilizada pelo workflow, identificando as chamadas ao Minimap2, Samtools, SeqKit, Mosdepth e demais ferramentas auxiliares, além da forma como os canais do Nextflow transportavam informações entre os processos.

Durante essa investigação também foram identificados mecanismos internos para preparação das referências genômicas, geração dos índices MMI, concatenação de múltiplos arquivos FASTA, validação das sequências de referência, organização automática dos diretórios de trabalho e gerenciamento dos recursos computacionais utilizados durante a execução.

A análise detalhada desses componentes permitiu compreender a arquitetura do workflow sem depender exclusivamente de sua documentação, caracterizando uma etapa de engenharia reversa aplicada a pipelines de bioinformática.

Levantamento das limitações

Após a compreensão do funcionamento do workflow original, foram avaliadas suas limitações para utilização em pesquisas envolvendo organismos eucariotos.

Observou-se que sua arquitetura apresenta forte acoplamento com bibliotecas específicas do ecossistema EPI2ME, dificultando modificações estruturais significativas. Além disso, o pipeline foi concebido principalmente para análises de organismos procariontes, utilizando parâmetros, estratégias de organização dos dados e dependências pouco flexíveis para aplicações envolvendo genomas maiores e estruturalmente mais complexos.

Outro fator identificado foi a dificuldade de manutenção e personalização do código, consequência da elevada quantidade de módulos internos e dependências externas. Dessa forma, concluiu-se que uma adaptação direta do workflow não atenderia aos objetivos do projeto.

Projeto da nova arquitetura

Com base nas informações obtidas durante a investigação, foi projetada uma nova arquitetura de pipeline utilizando Nextflow. Ao invés de reutilizar diretamente o código do wf-alignment, optou-se por implementar processos independentes responsáveis por cada etapa do fluxo de processamento. Essa abordagem proporcionou maior modularidade, facilidade de manutenção e controle sobre os parâmetros utilizados em cada ferramenta.

A pipeline desenvolvida contempla etapas de leitura das amostras, preparação das referências, construção do índice genômico utilizando o Minimap2, alinhamento das leituras, ordenação e indexação dos arquivos BAM utilizando o Samtools e organização automática dos resultados em diretórios específicos.

Todos os módulos foram desenvolvidos de forma desacoplada, permitindo futuras substituições de ferramentas ou inclusão de novas etapas analíticas sem necessidade de reestruturação completa do workflow.

Após a validação funcional de cada módulo, os processos passaram a ser integrados gradualmente até formar uma pipeline completamente automatizada. Durante essa etapa foram realizados diversos ciclos de desenvolvimento, testes, identificação de falhas, correções e novas validações.

Também foram implementados mecanismos para parametrização da execução, permitindo ao usuário informar dinamicamente os arquivos FASTQ, as referências genômicas e os diretórios de saída, aumentando a flexibilidade da aplicação.

Validação externa e participação em eventos científicos

Como parte complementar da metodologia, o processo de desenvolvimento e investigação metodológica foi submetido à validação externa por meio da participação em eventos científicos, fundamentais para debater desafios técnicos complexos — como o gerenciamento de memória e o tratamento de gargalos de Out Of Memory (OOM). Entre os dias 5 e 7 de novembro de 2025, o projeto foi apresentado em formato de pôster no Natal Bioinformatics Forum, evento bienal que favoreceu a comutabilidade de pesquisas e a troca de experiências com especialistas sobre o ecossistema Nextflow. Posteriormente, entre os dias 17 e 21 de agosto de 2026, a participação no IX Simpósio Norte e Nordeste de Bioinformática permitiu refinar ainda mais as estratégias de otimização, auxiliando na definição do escopo ideal para a solução dos problemas computacionais enfrentados.

Documentação

Durante todas as etapas do projeto foi mantida uma documentação contínua do processo de desenvolvimento. Foram registrados os experimentos realizados, as dificuldades encontradas durante a investigação do workflow de referência, as decisões arquiteturais adotadas, os testes executados e as soluções implementadas.

Resultados e Discussões

Os resultados obtidos refletem as conquistas técnicas e a validação da nova pipeline desenvolvida para o Centro Multiusuário de Bioinformática (BioME). A engenharia reversa do wf-alignment evidenciou que seu acoplamento ao ecossistema EPI2ME inviabilizava sua adaptação direta para organismos eucariotos, motivando o desenvolvimento de uma arquitetura própria, limpa, desacoplada e modular em Nextflow, capaz de atender às demandas específicas do BioME.

A integração dos diferentes módulos, incluindo Minimap2, Samtools, SeqKit e Mosdepth, proporcionou maior flexibilidade operacional e permitiu a automatização das principais etapas do fluxo de análise. Durante os testes de desempenho com conjuntos de dados superiores a 10 GB, entretanto, foi identificado um gargalo crítico relacionado ao consumo excessivo de memória (Out Of Memory – OOM), especialmente nas etapas de indexação e ordenação de genomas com maior complexidade estrutural. A identificação desse problema possibilitou a realização de ajustes na configuração do fluxo de execução, no gerenciamento de canais e diretórios de trabalho e na distribuição dos recursos computacionais, resultando na mitigação dos erros de OOM e na maior estabilidade da pipeline.

Como resultado desse processo de otimização, a pipeline desenvolvida apresentou estabilidade e desempenho compatíveis com a infraestrutura computacional disponível no BioME. A automação eliminou etapas manuais e repetitivas, enquanto a parametrização dinâmica possibilitou o processamento de diferentes genomas de referência e amostras biológicas, ampliando a flexibilidade e a aplicabilidade da ferramenta. Além disso, a estrutura modular em Nextflow favorece a manutenção, a expansão e a reutilização do fluxo em diferentes projetos de pesquisa.

Por fim, todo o conhecimento adquirido ao longo do desenvolvimento e da etapa de otimização foi sistematizado em uma documentação técnica abrangente, contemplando guias de instalação, configuração dos ambientes computacionais, descrição dos parâmetros e exemplos de utilização. Essa documentação contribui para a reprodutibilidade dos experimentos, reduz a curva de aprendizado de novos usuários e consolida a pipeline como uma ferramenta sustentável para as atividades de pesquisa e ensino desenvolvidas no BioME.

Como consequência do desenvolvimento e da consolidação da pipeline, o projeto também possibilitou sua apresentação e discussão em eventos científicos da área. Entre os dias 5 e 7 de novembro de 2025, o trabalho foi apresentado em formato de pôster na edição bienal do Natal Bioinformatics Forum, proporcionando oportunidades de divulgação e troca de experiências com outros pesquisadores. Posteriormente, entre 17 e 21 de agosto de 2026, a participação no IX Simpósio Norte e Nordeste de Bioinformática permitiu ampliar a disseminação dos resultados e compartilhar as estratégias de otimização adotadas durante o desenvolvimento da ferramenta. Dessa forma, a participação nos eventos constituiu um desdobramento natural do trabalho realizado, contribuindo para sua divulgação, validação junto à comunidade científica e intercâmbio de conhecimentos na área de bioinformática.

Conclusão

O desenvolvimento deste trabalho possibilitou compreender em profundidade o funcionamento de pipelines de bioinformática baseadas em Nextflow, especialmente o workflow wf-alignment, utilizado inicialmente como referência para investigação. A análise detalhada de sua arquitetura, de suas dependências e dos processos executados permitiu identificar os principais componentes responsáveis pelo alinhamento de sequências genômicas, bem como as limitações da solução para aplicações envolvendo organismos eucariotos.

Durante o processo de pesquisa, verificou-se que a elevada complexidade estrutural e o forte acoplamento do wf-alignment com ferramentas específicas do ecossistema EPI2ME dificultavam sua adaptação às necessidades do Centro Multiusuário de Bioinformática (BioME). Em razão disso, optou-se pelo desenvolvimento de uma pipeline independente, mantendo apenas os conceitos arquiteturais obtidos durante a etapa de engenharia reversa e implementando uma solução própria, mais simples, modular e flexível.

A pipeline desenvolvida foi construída utilizando Nextflow e integra ferramentas consolidadas da bioinformática, como Minimap2 e Samtools, organizadas em processos independentes que facilitam a manutenção, a reutilização e futuras expansões do sistema. Além disso, a solução permite a parametrização da execução, o processamento automatizado de múltiplas amostras e a utilização de diferentes referências genômicas, tornando-a adequada para diferentes cenários de pesquisa.

Os testes realizados demonstraram a estabilidade da implementação e sua capacidade de processar conjuntos de dados de grande porte, incluindo arquivos de entrada superiores a 10 GB, mantendo desempenho compatível com a infraestrutura computacional disponível no BioME. Esses resultados evidenciam que a arquitetura proposta atende aos objetivos estabelecidos para o projeto, oferecendo uma base sólida para análises de organismos eucariotos e para futuras otimizações relacionadas ao paralelismo, utilização de recursos computacionais e incorporação de novas ferramentas.

Além dos resultados práticos obtidos, este trabalho proporcionou uma experiência significativa na aplicação de técnicas de engenharia reversa, desenvolvimento de workflows científicos, computação paralela e gerenciamento de pipelines em bioinformática. O conhecimento adquirido durante a investigação do workflow original foi fundamental para o projeto da nova solução e contribuiu para uma compreensão mais aprofundada das estratégias utilizadas na construção de pipelines de análise genômica.

Portanto, espera-se que a solução desenvolvida possa contribuir tanto para as atividades de pesquisa quanto para ações de ensino e capacitação promovidas pelo BioME, consolidando-se como uma ferramenta de apoio para estudos em bioinformática.

Referências

1- Epi2me Labs (2024). wf-alignment. Disponível em: https://github.com/epi2me-labs/wf-alignment

2- Di Tommaso, P. et al. (2017). Nextflow enables reproducible computational workflows. Nature Biotechnology.

3- Kim, D. et al. (2015). HISAT: a fast spliced aligner with low memory requirements. Nature Methods.

4- Dobin, A. et al. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics.

5- Merkel, D. (2014). Docker: lightweight Linux containers for consistent development and deployment. Linux Journal.

6- Kurtzer, G. M., Sochat, V., Bauer, M. W. (2017). Singularity: Scientific containers for mobility of compute. PLoS ONE.

7- Cock, P. J. A., et al. (2010). The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic Acids Research.

Anexos



Pôster Natal Bioinformatics Forum

CÓDIGO: SB1224

AUTOR: MARIA EDUARDA LOURENÇO

COAUTOR: GABRIELLE GERMEK COELHO SANTOS

COAUTOR: JULIO CESAR DA SILVA FILHO

COAUTOR: ANA BEATRIZ CAMILO DA COSTA

ORIENTADOR: JORGE ESTEFANO SANTANA DE SOUZA

TÍTULO: Estudo e Customização do Pipeline genomeassembler para Montagem e Análise de Genomas Eucariotos no ambiente computacional do BioME - IMD/UFRN

Resumo

O aumento expressivo no volume de dados genômicos, impulsionado pelos avanços das tecnologias de sequenciamento de nova geração (NGS) tornou fundamental o desenvolvimento e o uso de pipelines eficientes para a montagem e análise genômica. Nesse cenário, o Centro Multiusuário de Bioinformática (BioME) do Instituto Metrópole Digital (IMD/UFRN) demanda fluxos de trabalho otimizados para processar conjuntos de dados complexos, especialmente em projetos voltados a organismos eucariotos. Diante disso, este trabalho teve como objetivo analisar e otimizar o pipeline nf-core/genomeassembler, visando aprimorar o desempenho na montagem de genomas eucariotos na infraestrutura computacional do BioME. Inicialmente, o pipeline foi reproduzido localmente utilizando o genoma da espécie-modelo Arabidopsis thaliana, para a análise de parâmetros de montagem e identificação de possíveis gargalos computacionais nos ambientes Conda e Docker. Em seguida, desenvolveu-se o protótipo de pipeline por meio de uma estratégia de engenharia reversa da ferramenta original. O pipeline construído demonstrou resultados prévios satisfatórios, integrando com sucesso as etapas de controle de qualidade, montagem, patching e anotação no perfil Conda. Por fim, o pipeline permitiu otimizar a execução paralela das amostras e a customização de parâmetros, abrangendo as diversas análises provenientes do ambiente computacional do BioME.

Palavras-chave: bioinformática. montagem. nextflow. nf-core. eucarioto. pipeline.

TITLE: Study and Customization of the GenomeAssembler Pipeline for Eukaryotic Genome Assembly Analysis in the BioME – IMD/UFRN Computational Environment

Abstract

The significant increase in the volume of genomic data, driven by advances in next-generation sequencing (NGS) technologies, has made the development and use of efficient pipelines essential for genome assembly and analysis. In this context, the Bioinformatics Multidisciplinary Environment - BioME of the Digital Metropolis Institute (IMD/UFRN) requires optimized workflows to process complex datasets, especially in projects focused on eukaryotic organisms. Given this, the present study aimed to analyze and optimize the nf-core/genomeassembler pipeline to enhance the performance of eukaryotic genome assembly within BioME's computational infrastructure. Initially, the pipeline was reproduced locally using the genome of the model species Arabidopsis thaliana to analyze assembly parameters and identify potential computational bottlenecks in both Conda and Docker environments. Subsequently, a pipeline prototype was developed through a reverse engineering strategy of the original tool. The constructed pipeline demonstrated satisfactory preliminary results, successfully integrating the quality control, assembly, patching, and annotation steps within the Conda profile. Finally, the pipeline enabled the optimization of parallel sample execution and parameter customization, accommodating the diverse analyses originating from BioME's computational environment.

Keywords: bioinformatics. assembly. Nextflow. nf-core. eukaryote. pipeline.

Introdução

Com a evolução das tecnologias de sequenciamento de nova geração (Next-Generation Sequencing - NGS), houve uma explosão na geração de dados genômicos; essa expansão tornou indispensável o uso de pipelines automatizados, eficientes e reprodutíveis para o processamento e a análise de dados. Desde o processo de controle de qualidade (Quality Control - QC) das leituras resultantes do sequenciamento, a montagem genômica representa uma etapa fundamental da análise bioinformática, ela é responsável pela reconstrução e organização das sequências genômicas de um organismo a partir das leituras obtidas por tecnologias de sequenciamento; obtidas principalmente por meio das plataformas Oxford Nanopore Technologies (ONT) ou PacBio HiFi. Após esse processo, é implementado o patching para o refinamento dos contigs, que são sequências contínuas de DNA montadas a partir de leituras de sequenciamento. Finalizando com a anotação para identificar e descrever as informações biológicas.

Dentro do ambiente do Centro Multiusuário de Bioinformática (BioME) do Instituto Metrópole Digital (IMD/UFRN), é observada uma necessidade de ferramentas que sejam capazes de enfrentar desafios do processamento de dados, especialmente de genomas eucariotos, que têm como características a alta complexidade estrutural ligada à heterozigosidade e a forte presença de regiões repetitivas. Esses fatores exigem maior poder computacional, levando em consideração dependências, performance em High-Performance Computing (HPCs) e a ausência de suporte direto a alguns formatos de entrada comuns, especialmente em trabalhos realizados com organismos não modelo, que dispõem de poucos recursos genéticos disponíveis.

Diante disso, a criação de pipelines que fazem a integração de diferentes estratégias e ferramentas de montagem, como Hifiasm e Flye, que podem oferecer perfis pré-configurados permitindo a customização de parâmetros conforme as características dos dados de entrada, se mostra como uma abordagem promissora para a automatização do processamento de dados. O objetivo deste estudo é analisar e otimizar o pipeline genomeassembler (versão 1.0.1) desenvolvido pelo Niklas Schandry da Ludwig-Maximilians University (LMU) (NF-CORE, 2023) utilizando o Nextflow como sistema de fluxo de trabalho, pois esse sistema permite a prototipagem rápida e a implantação de pipelines que combinam pacotes de software complementares (EWELS et al., 2020). Disponibilizado na comunidade global nf-core, para aprimorar o desempenho na montagem de genomas eucariotos na infraestrutura computacional do BioME.

Metodologia

Inicialmente, o pipeline original nf-core/genomeassembler foi reproduzido localmente dentro dos perfis de execução Docker e Conda utilizando o genoma da amostra teste Arabidopsis thaliana. A partir disso, foi realizada uma análise funcional e técnica detalhada do pipeline original, incluindo a revisão da documentação, fluxo de execução das etapas, estrutura e inspeção dos arquivos de configuração.

Durante a execução desses testes, arquivos temporários foram gerados pelo Nextflow, contendo resultados e scripts associados a cada processo, o que possibilitou a inspeção dos comandos executados durante o processamento, como o arquivo .command.sh e outros registros produzidos. A análise desses arquivos tornou possível identificar as ferramentas utilizadas em cada etapa, as dependências entre os processos e os arquivos de entrada e saída. A partir dessas informações, foi possível entender a estrutura operacional do workflow, permitindo a reconstrução do pipeline mantendo o fluxo de dados entre as etapas.

Para o desenvolvimento do protótipo, foram utilizadas como arquivos de entrada, leituras longas Oxford Nanopore Technologies (ONT) e PacBio HiFi, cujo os

fluxos de processamento foram executados paralelamente. As leituras ONT têm como característica seu maior comprimento, por isso contribuem para dar uma maior continuidade da montagem tendo uma melhor resolução para regiões de alta repetição, já as PacBio HiFi apresentam uma elevada acurácia, contribuindo para uma maior precisão e cobertura das Reads. Além das leituras, o pipeline recebe um genoma de referência e um arquivo de anotação genômica, que apresenta informações sobre as características genômicas previamente estabelecidas para o genoma utilizado como base, estes são aplicados na etapa final do processamento, após a finalização da montagem, como dados de referência para as anotações da ferramenta Liftoff, uma ferramenta que mapeia anotações descritas em General Feature Format (GFF) ou General Transfer Format (GTF) entre montagens da mesma espécie, ou espécies relacionadas (SHUMATE; SALZBERG, 2021) .

O protótipo foi implementado utilizando o Nextflow como gerenciador de workflow, possibilitando a automatização do processo e a paralelização das etapas do fluxo de trabalho, bem como da criação de arquivos Logs para auditoria posterior dos dados e acompanhamento do fluxo *step-by-step*, com suporte ao perfil de execução Conda para o gerenciamento do ambiente computacional e das dependências necessárias.(EWELS et al., 2020)

O fluxo de processamento foi organizado em etapas paralelas e subsequentes. Paralelamente, os dados PacBio Hifi são processados pelo Hifiasm (0.25.0), responsável pela montagem, uma vez que a ferramenta demonstrou que se esforça para preservar a contiguidade de todos os haplótipos, diferente de outros montadores baseados em grafos que visam apenas manter a contiguidade de um haplótipo (Cheng et al., 2021). Como resultado, é gerado um arquivo GFA. Este arquivo posteriormente é levado para a ferramenta GFA2FA e convertido para o formato FASTA, enquanto as leituras ONT são submetidas ao Nanoq (0.10.0), para realizar o controle de qualidade (QC), gerando um relatório em formato JSON, após esse módulo a leitura ONT é encaminhada para o Flye (2.9.6) onde também ocorre a montagem, uma vez que o algoritmo da ferramenta foi desenvolvido especificamente para trabalhar com leituras longas e com altas taxas de erro (Kolmogorov et al., 2019). Ao final do processo de montagem e conversão, são gerados dois arquivos “.Fasta”, estes são levados ao RagTag Patch (2.1.0), que é responsável pela etapa de patching, no qual as duas montagens são integradas. Posteriormente, esta montagem resultante se junta ao genoma de referência e o arquivo de anotação para o processo final utilizando o Liftoff (1.6.3)onde é comparada a nova montagem com as referências e é feita anotações por transferência para o resultado finalizado pelo protótipo.

Resultados e Discussões

A reconstrução e otimização do pipeline **nf-core/genomeassembler** demonstraram resultados prévios satisfatórios na integração das etapas de controle de qualidade, montagem, *patching* e anotação. A utilização do

gerenciador de workflow Nextflow permitiu não apenas automatização e a paralelização eficiente dos processos, mas também o acompanhamento por etapa do fluxo de dados, com suporte robusto ao perfil de execução Conda para o gerenciamento de dependências no ambiente computacional do Centro Multiusuário de Bioinformática (BioME). A reprodução inicial foi conduzida localmente utilizando o genoma da espécie-modelo Arabidopsis thaliana, o que permitiu inspecionar arquivos temporários e de configuração, validando o fluxo de dados entre as etapas operacionais.

3.1. Processamento de Leituras Longas e Controle de Qualidade

Durante o processamento paralelo das leituras longas provenientes das plataformas Oxford Nanopore Technologies (ONT) e PacBio HiFi, a etapa inicial responsável pelo QC direcionada aos dados ONT foi executada com sucesso pela ferramenta Nanoq resultando na geração de um relatório em formato JSON. A análise deste relatório evidenciou um conjunto de dados composto por 6.202 leituras, totalizando 26.752.141 bases (~26,75 Mb). Observou-se um tamanho médio de leitura de 4.313 pb, com um N50 de 6.459 pb e o fragmento mais longo atingindo 97.030 pb, além de um escore de qualidade (Q-score) médio de 18,23. O controle de qualidade é considerado indispensável para leituras ONT, pois apesar de possuírem maior comprimento, contribuindo para a continuidade da montagem, elas apresentam taxas de erro de sequenciamento superiores às das leituras PacBio HiFi.

3.2. Métricas de montagem e Patching

Para avaliar o desempenho do pipeline, foram obtidas métricas das montagens geradas pelas diferentes ferramentas. A análise individual dos resultados evidenciou diferenças entre as abordagens utilizadas. Entre as principais métricas utilizadas, o número de contigs permite avaliar o nível de fragmentação da montagem. Nesse contexto, o processamento das leituras PacBio HiFi pelo Hifiasm apresentou elevada contiguidade, resultando em um único contig de 2.045.287 pb e L50 igual a 1. Já a montagem das leituras ONT, obtida pelo Flye, apresentou 2 contigs, com maior contig e N50 de 1.680.905 pb, além de comprimento total de 2.038.938 pb. Dessa forma, a montagem baseada nas leituras PacBio HiFi apresentou menor fragmentação, enquanto a abordagem baseada em ONT também produziu uma montagem de elevada continuidade, embora tenha apresentado dois contigs.

O aspecto mais expressivo dos resultados, no entanto, é o impacto da ferramenta RagTag Patch, responsável pelo patching. Ao utilizar a montagem HiFi, que tem como característica ser altamente contígua e acurada, e as informações da montagem ONT, o processo de patching manteve a integridade estrutural, resultando em um único contig e L50 igual a 1, e aumentou o comprimento total do genoma montado para 2.050.524 pb. Esse aumento no

comprimento, em pares de bases, evidencia o sucesso do RagTag em incorporar informações adicionais provenientes da montagem ONT a ultra-contiguidade das leituras Nanopore contribuindo para a extensão de regiões que não haviam sido completamente resolvidas na montagem inicial pelo Hifiasm.

3.3. Anotação de referência

A anotação genômica foi realizada utilizando a ferramenta Liftoff, responsável por mapear as características genômicas, como genes, éxons e transcritos, do genoma de referência de Arabidopsis thaliana para a montagem híbrida obtida após a etapa de patching realizada pelo RagTag. A utilização de uma anotação previamente estabelecida viabilizou a associação de informações funcionais à sequência reconstruída, complementando a avaliação das métricas de montagem.

A execução do Liftoff foi realizada com êxito por meio do gerenciamento do Nextflow e do ambiente Conda. A anotação por transferência para a montagem híbrida demonstrou a compatibilidade da sequência obtida com as características genômicas presentes no genoma de referência. Dessa forma, essa etapa ampliou a aplicação do pipeline, possibilitando a geração de informações funcionais que podem ser utilizadas em análises posteriores de genômica.

3.4. Documentação

Durante todo o período da pesquisa, foi realizada uma documentação contínua das etapas de testes e desenvolvimento do pipeline. Foram registrados os procedimentos executados, as dificuldades encontradas e os erros operacionais identificados ao longo do processo. Além disso, a versão desenvolvida do pipeline foi armazenada em um repositório no GitHub, contribuindo para o versionamento e a organização (https://github.com/eduardalou/prototipo_assembly).

Conclusão

Por fim, a reconstrução e otimização do pipeline nf-core/genomeassembler permitiu integrar, em um fluxo automatizado, as etapas de controle de qualidade, montagem genômica, patching e anotação, utilizando o Nextflow como gerenciador do fluxo de trabalho e o Conda para o gerenciamento das dependências. A reprodução do pipeline original e os testes realizados com

Arabidopsis thaliana possibilitaram identificar e compreender o funcionamento das diferentes etapas, contribuindo para a implementação do protótipo.

As métricas analisadas demonstraram que a utilização de diferentes leituras e estratégias de montagem produziu resultados complementares. O Hifiasm, aplicado às leituras PacBio HiFi, apresentou elevada contiguidade, enquanto o Flye, aplicado às leituras ONT, também produziu uma montagem com elevada continuidade. A utilização do RagTag Patch possibilitou integrar essas montagens, mantendo um único contig e ampliando o comprimento total da sequência. Posteriormente, a execução do Liftoff permitiu transferir as anotações do genoma de referência para a montagem obtida, agregando informações funcionais ao resultado final.

Além dos resultados das etapas de processamento, a documentação contínua dos testes, dificuldades e modificações realizadas, juntamente com o armazenamento do pipeline em um repositório GitHub, contribuiu para a organização e o versionamento do projeto. Dessa forma, o protótipo desenvolvido demonstrou a viabilidade da integração das estratégias e ferramentas usadas, constituindo uma base para o aprimoramento do processamento e da análise de genomas eucariotos no ambiente computacional do BioME.

Referências

KOLMOGOROV, M. et al. Assembly of long, error-prone reads using repeat graphs. Nature Biotechnology, v. 37, p. 540-546, 2019. Disponível em: https://doi.org/10.1038/s41587-019-0072-8.

EWELS, Philip A. et al. The nf-core framework for community-curated bioinformatics pipelines. Nature Biotechnology, v. 38, p. 276–278, 2020.

CHENG, H. et al. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nature Methods, v. 18, p. 170–175, 2021. DOI: https://doi.org/10.1038/s41592-020-01056-5.

SHUMATE, Alaina; SALZBERG, Steven L. Liftoff: accurate mapping of gene annotations. Bioinformatics, v. 37, n. 12, p. 1639–1643, 2021. DOI: https://doi.org/10.1093/bioinformatics/btaa1016.

nf-core (2023). nf-core/genomeassembler: A Nextflow pipeline for genome assembly. https://nf-co.re/genomeassembler