



Revista Ibérica de Sistemas e Tecnologias de Informação
Iberian Journal of Information Systems and Technologies

ISSN: 1646-9895

F e v e r e i r o 2 6 • F e b r u a r y 2 6



©ITMA 2026 <http://www.risti.xyz>

Nº E78

Edição / Edition

N.º E78, 02/2026

ISSN: 1646-9895

Indexação / Indexing

Academic Journals Database, Dialnet, DOAJ, DOI, EBSCO, GALE, Google Scholar, IndexCopernicus, Information Systems Journal, Latindex, ProQuest, QUALIS, SciELO, SIS, Ulrich's

Publicação / Publication

ITMA - Information and Technology Management Association

Pc. 9 de Abril, nº 26, 4200-422 Porto, Portugal

Web: <https://itmas.org>

Director

Álvaro Rocha, ISEG, University of Lisbon, PT

Contact: aistic@gmail.com

Coordenadores da Edição / Issue Coordinators

Luis Loo, Universidad Tecnológica de Honduras, HN

Conselho Editorial / Editorial Board

Abel Méndez, Technological Institute of Costa Rica, CR

Alejandro Peña, EAFIT University, CO

Alma Gomez-Rodríguez, University of Vigo, ES

Ana Rita Calvao, ESTGA, University of Aveiro, PT

Ania Cravero, Universidad de La Frontera, CL

António Abreu Silva, ISCAP, Polytechnic Institute of Porto, PT

Antonio Garcia, University of Santiago de Compostela, ES

António Godinho, ISEC, Polytechnic Institute of Coimbra, PT

Antonio Jiménez-Martín, Polytechnic University of Madrid, ES

Arturo J. Méndez, University of Vigo, ES

August Climent, Ramon LLull University, ES

Beatriz Rodríguez, Universidad de la Republica, UY

Borja Bordel, Polytechnic University of Madrid, ES

Brenda L. Flores-Rios, Universidad Autónoma de Baja California, MX

Bruno Miguel Ferreira Gonçalves, Polytechnic Institute of Bragança, PT

Carlos Alexandre Silva, Federal Institute of Minas Gerais, BR

Carlos Carreto, Polytechnic Institute of Guarda, PT

Carlos Moraes, Polytechnic Institute of Bragança, PT

Carlos Rompante Cunha, UNIAG & CeDRI &
Polytechnic Institute Bragança, PT

Carlos Vaz de Carvalho, Polytechnic Institute of Porto, PT

Célio Marques, Polytechnic Institute of Tomar, PT

Ciro Martins, University of Algarve, PT

Concepción Damián-Hernández, Instituto Tecnológico
Superior de Misantla, MX

Cristina Caridade, Polytechnic Institute of Coimbra, PT
Daniel Poland, University of Aveiro, PT
Dante Carrizo, University of Atacama, CL
Edwin Cedeño Herrera, Universidad de Panamá, PA
Fabio Marques, ESTGA, University of Aveiro, PT
Felipe Vasquez, Universidad de La Frontera, CL
Fernando Moreira, REMIT, IJP, Portucalense University &
IEETA, University of Aveiro, PT
Fernando Paulo Belfo, Coimbra Business School –
ISCAC, Polytechnic Institute of Coimbra, PT
Fernando Ribeiro, Polytechnic Institute of Castelo Branco, PT
Fernando Bandeira, PT
Filipe Cardoso, Polytechnic Institute of Viseu, PT
Flor Gomes de María Sánchez Aguirre, Universidad César Vallejo, PE
Francisco Javier Lena-Acebo, University of Cantabria, ES
Gabriel Guerrero-Contreras, University of Cádiz, ES
Gerardo González Filgueira, University of A Coruña, ES
Gloria Valencia, Universidad de las Fuerzas Armadas, EC
Helder Gomes, University of Aveiro, PT
Helder Zagalo, University of Aveiro, PT
Hélia Guerra, University of the Azores, PT
Henrique Gil, Polytechnic Institute of Castelo Branco, PT
Henrique S. Mamede, Open University, Portugal
Inês Domingues, ISEC, Polytechnic Institute of Coimbra, PT
Isaias Bianchi, Federal University of Santa Catarina, BR
Ismael Etxeberria-Agiriano, Universidad del País Vasco, ES
Ivan Garcia, Universidad Tecnológica de la Mixteca, MX
Jeimy Cano, Universidad de los Andes, CO
John Emilio Almeida, ISTECH - Porto, PT
João Paulo Ferreira, Polytechnic Institute of Coimbra, PT
Roberto de Toledo Quadros, CEFET/RJ, BR
Joao Tavares, University of Porto, PT
Joaquim Reis, ISCTE - University Institute of Lisbon, PT

Jorge Eduardo Ibarra Esque, Universidad
Autónoma de Baja California, MX
Jorge Hochstetter, Universidad de La Frontera, CL
José Álvarez-García, University of Extremadura, ES
Jose Felipe Cocon Juarez, Universidad Autónoma del Carmen, MX
José Lousado, Polytechnic Institute of Viseu, PT
José Luis Pastrana Brincones, University of Malaga, ES
Jose M. Molina, Carlos III University of Madrid, ES
José Ribeiro, Polytechnic of Leiria, PT
Jose Silvestre Silva, Military Academy, PT
Juan Angel Contreras Vas, University of Extremadura, ES
Juan M. Santos-Gago, University of Vigo, ES
Juan Pablo DAmato, National University of the
Center of the Province of Buenos Aires, AR
Leila Weitzel, Fluminense Federal University, BR
Leonardo Bermón Angarita, National University of Colombia, CO
Lilia Muñoz, Universidad Tecnológica de Panamá, PA
Lucila Romero, Universidad Nacional del Litoral, AR
Luis Alvarez Sabucedo, University of Vigo, ES
Luis Enrique Sánchez Crespo, University of Castilla-la Mancha, ES
Luisa María Romero-Moreno, Universidad Sevilla, ES
Luz María Hernández Cruz, Universidad Autónoma de Campeche, MX
Luz Sussy Bayona Oré, Universidad Nacional Mayor de San Carlos, PE
Marcelo Marciszack, National Technological University, AR
Marcelo Zambrano Vizuite, Universidad Tecnica Del Norte
Marco Painho, NOVA IMS, PT
Margarita Ramirez Ramirez, Universidad
Autónoma de Baja California, MX
Maria Cristina Marcelino Bento, UNIFATEA, BR
María de la Cruz del Río-Rama, University of Vigo, ES
Maria de los Milagros Gutierrez, National Technological University, AR
Maria do Rosario Bernardo, Open University, PT
Maria João Ferreira, Portucalense University, PT

Maria João Gomes, University of Minho, PT
Maria Sousa, ISCTE - University Institute of Lisbon, PT
Marisol B. Correia, ESGHT - University of Algarve & CiTUR, PT
Maristela Holanda, University of Brasilia, BR
Martin Llamas Nistal, University of Vigo, ES
Miguel Casquilho, University of Lisbon, PT
Miguel Ramón González Castro, Aimen Technological Center, ES
Mirna Ariadna Muñoz Mata, CIMAT, MX
Nelson Rocha, University of Aveiro, PT
Nuno Melão, Polytechnic Institute of Viseu, PT
Nuno Ribeiro, Fernando Pessoa University, Portugal
Patricia Dias, State University of Minas Gerais, BR
Paula Prata, University of Beira Interior, PT
Paulo Martins, University of Trás-os-Montes and Alto Douro, PT
Paulo Pinto, New University of Lisbon, PT
Paulo Rurato, Fernando Pessoa University, Portugal
Pedro Araújo, University of Beira Interior, PT
Pedro R. Palos-Sanchez, Universidad de Sevilla, ES
Pedro Sanz-Angulo, University of Valladolid, ES
Pedro Sobral, Fernando Pessoa University, Portugal
Pedro Sousa, University of Minho, Portugal
René Faruk Garzozzi-Pincay, Santa Elena Peninsula State University, EC
Ruben Pereira, ISCTE – University Institute of Lisbon, PT
Rui Pedro Marques, University of Aveiro, PT
Rui Silva Moreira, Fernando Pessoa University, PT
Samuel Sepúlveda, Universidad de La Frontera, CL
Sandro Carvalho, Polytechnic of Cávado and Ave, PT
Santiago Raul Gonzales Sanches, Universidad Cesar Vallejo, PE
Sara Balderas-Díaz, University of Cádiz, ES
Sergio Araya Guzmán, Universidad del Bío-Bío, CL
Sergio F. Lopes, University of Minho, PT
Solange N Alves-Souza, University of Sao Paulo, BR
Telmo Silva, DigiMedia, University of Aveiro, PT

Thiago Dias, Federal Center for Technological
Education of Minas Gerais, BR

Veronica Vasconcelos, Polytechnic Institute of Coimbra, PT

Vítor Carvalho, Polytechnic Institute of Cávado and Ave, PT

Vitor Santos, New University of Lisbon - NOVA IMS, PT

Wagner Tanaka Botelho, Federal University of ABC, BR

Yamid Hernández Julio, Universidad del Sinú, CO

Índice / Index

ARTIGOS / ARTICLES

Identificação e Mitigação de Ataques DDoS em Redes IoT utilizando HIDPS, NIDS e Aprendizado de Máquina	1
--	---

Awatef A. Yousef R. Fares, Claudio Santoro Wunder, Francisco L. de Caldas Filho, Geraldo P. Rocha Filho, André Luiz M. Serrano, Vinícius P. Gonçalves

Juris Syntax: Automação da Análise Jurídica Brasileira com IA Generativa e PLN	13
--	----

Fábio Lúcio Lopes de Mendonça, Bruno Justino Garcia Praciano, Flávio Garcia Praciano, Thiago Leite de Sousa, Elon Oliveira Albuquerque, José Pérciles Pereira de Sousa

Sistema de apoyo al Diseño Curricular: Extensión de Meta4CBC para la evaluación por competencias	28
--	----

Ninoska Díaz, Ania Cravero

Redes neuronales convolucionales (YOLO) para la detección de megafauna marina y basura marina flotante a partir de imágenes aéreas	40
--	----

Maria V DeLaHoz., Toni Monleón-Getino, Albert Martín, Odei Garcia- Garin, Morgana Vighi, Marina Costa, Valentina De Santis, Caterina Lanfredi, Capt. Daniele Giannelli, Sabina Airoidi

O Futuro da Saúde Preditiva: Uma Análise do Estado da Arte em IoT e Sistemas Embarcados para Identificar Oportunidades de Inovação	55
--	----

Jhonatas Gomes Ribeiro, Keyllane Guedes de Souza, Phaola Paraguai da Paixão Lustosa, Wanderson Jean Conceição Silva, Felipe Gonçalves dos Santos

Implementação da Solução de Gestão de Arquitetura de TI: Estudo em uma Rede Farmacêutica Regional	69
---	----

Murilo Albeest, Rodolfo Miranda de Barros

Detecção de Phishing multicanal (e-mail, sms, sites): uma abordagem com deep learning utilizando DistilBERT, CNN, RNN e MLP	83
---	----

Fredman Wellington Lopes, Vinicius Pereira Goncalves, Leandro Lopes Benfica

A Alta Administração na Linha de Frente da Segurança da Informação: Utopia ou Necessidade?	98
<i>Adriana Marcílio, Edvan Gomes da Silva, Virgínia de Melo Dantas Trinks, Lucas Vinícius Andrade Ferreira, Rafael Rabelo Nunes</i>	
Desafios e Oportunidades na Implantação de Sistemas Embarcados para o Controle do Ambiente em Aviários: Uma Revisão Sistemática de Literatura	114
<i>Daniel de Oliveira Cunha, Leonardo da Silva Lisboa, Marcos do Livramento Amaral, Uenderson Mendes do Nascimento, Felipe Gonçalves dos Santos, Wanderson Jean Conceição Silva</i>	
Sistemas Embarcados e Visão Computacional para Segurança Residencial: Uma Revisão da Literatura.....	129
<i>Joniel Mendes da Silva, João André Borges Santana, Heitor Garcez Martins Louzeiro, Wanderson Jean Conceição Silva, Felipe Gonçalves dos Santos</i>	
Modelo de aprendizagem automático para la recomendación de carreras en jóvenes preuniversitarios	142
<i>Andree Carrillo Guevara, Jose Avila Rabanal, Juan Mansilla-Lopez, Ricardo Loza</i>	
CATCH: A Nova Fronteira dos Chatbots na Gestão de Força de Trabalhos	157
<i>Elon Oliveira Albuquerque, Bruno Justino Garcia Praciano, Paulo Henrique Batista Rodrigues, Flávio Garcia Praciano, Fábio L. Lopes de Mendonça, Geraldo P. Rocha Filho</i>	
Mitigação de Ataques DDoS em IoT: Comparação entre Aprendizado Federado e Modelos Centralizados	170
<i>Mateus Flach Romani, João Kleber M. dos Santos, Gabriel Filipe de O. Cardoso, Giovani B. O. Viana, Francisco L. de Caldas Filho, Fábio L. L. de Mendonça</i>	

Identificação e Mitigação de Ataques DDoS em Redes IoT utilizando HIDPS, NIDS e Aprendizado de Máquina

Awatef A. Yousef R. Fares¹, Claudio Santoro Wunder², Francisco L. de Caldas Filho³,
Geraldo P. Rocha Filho⁴, André Luiz M. Serrano⁵, Vinícius P. Gonçalves⁶

awatefalifares@gmail.com; claudio.santoro.wunder@gmail.com;
franciscolopesaldas@gmail.com; geraldo.rocha@uesb.edu.br;
andrelms@unb.br; vpgvinicius@unb.br

¹ Universidade de Brasília (PPEE/UnB), Brasília–DF, Brasil

² Universidade de Brasília (UnB), Brasília–DF, Brasil

³ Universidade de Brasília (PPEE/UnB), Brasília–DF, Brasil

⁴ Universidade Estadual do Sudoeste da Bahia (UESB), Brasil

⁵ Universidade de Brasília (PPEE/UnB), Brasília–DF, Brasil

⁶ Universidade de Brasília (PPEE/UnB), Brasília–DF, Brasil

Pages: 1-12

Resumo: A rápida expansão da Internet das Coisas (IoT) aumentou as vulnerabilidades das redes, tornando os dispositivos conectados alvos preferenciais para ataques de negação de serviço distribuído (DDoS). Este estudo apresenta uma abordagem híbrida que combina Sistemas de Detecção e Prevenção de Intrusões Baseados em Host (HIDPS) e Sistemas de Detecção de Intrusão Baseados em Rede (NIDS) com aprendizado federado para aprimorar a segurança das redes IoT. O aprendizado federado permite o treinamento descentralizado de modelos, preservando a privacidade dos dados, enquanto o HIDPS atua diretamente nos dispositivos comprometidos para manter a funcionalidade sem interrupção dos serviços. Ao integrar detecção baseada em assinaturas e em anomalias, a solução proposta identifica de forma eficaz tanto ameaças conhecidas quanto emergentes. Os resultados experimentais demonstram uma melhoria significativa na detecção de tráfego malicioso e uma redução na indisponibilidade dos serviços para os dispositivos IoT afetados.

Palavras-chave: Internet das Coisas, HIDPS, NIDS, DDoS, Redução, Aprendizado Federado.

Identificação e Mitigação de Ataques DDoS em Redes IoT utilizando HIDPS, NIDS e Aprendizado de Máquina

Abstract: The rapid expansion of the Internet of Things (IoT) has increased network vulnerabilities, making connected devices prime targets for distributed denial-

of-service (DDoS) attacks. This study presents a hybrid approach that combines Host-Based Intrusion Detection and Prevention Systems (HIDPS) and Network-Based Intrusion Detection Systems (NIDS) with federated learning to enhance IoT network security. Federated learning enables decentralized model training while preserving data privacy, whereas HIDPS operates directly on compromised devices to maintain functionality without service disruption. By integrating signature-based and anomaly-based detection, the proposed solution effectively identifies both known and emerging threats. Experimental results demonstrate a significant improvement in malicious traffic detection and a reduction in service unavailability for affected IoT devices.

Keywords: Internet of Things, HIDPS, NIDS, DDoS, Reduction, Federated Learning.

1. Introdução

A Internet das Coisas (IoT) está revolucionando a comunicação ao integrar dispositivos em cidades inteligentes, saúde e automação residencial. Contudo, o rápido crescimento desses dispositivos amplia as vulnerabilidades das redes, tornando-os alvos atrativos para ataques cibernéticos, especialmente os de negação de serviço distribuído (DDoS). Em redes IoT, os ataques DDoS são uma preocupação crescente devido ao aumento exponencial de dispositivos conectados e à sensibilidade das informações que transportam e armazenam.

Dispositivos IoT possuem limitações de segurança que podem ser exploradas, permitindo a formação de *botnets*, redes de dispositivos comprometidos controlados remotamente por invasores, utilizadas para sobrecarregar serviços essenciais. A atuação dessas *botnets* representa um dos desafios mais críticos da segurança cibernética moderna (Debicha et al., 2023). Falhas em redes de telecomunicações agravam esses riscos na IoT, onde até pequenos dispositivos podem ser comprometidos globalmente. Esse cenário exige soluções de segurança robustas e dinâmicas para proteger dados e serviços.

O aprendizado de máquina (ML) possibilita o desenvolvimento de modelos de inteligência artificial capazes de identificar padrões maliciosos no tráfego de rede, permitindo detecção e mitigação rápidas de anomalias. Dentre suas subdivisões, o aprendizado federado se destaca por permitir que dispositivos treinem modelos localmente, sem compartilhar dados originais. Isso cria modelos robustos para mitigar ataques como o DDoS, preservando a privacidade das informações.

Nos últimos anos, ataques DDoS utilizando dispositivos IoT têm crescido em frequência e volume, apresentando vetores mais sofisticados. Os impactos são amplos, afetando indivíduos, empresas, bancos, provedores de internet, infraestrutura de comunicação e instituições governamentais (Bala & Behal, 2024). A aplicação de algoritmos de aprendizado de máquina, em conjunto com ferramentas tradicionais como HIDPS (Host-based Intrusion Detection and Prevention System) e NIDS (Network-Based Intrusion Detection System), contribui para a identificação de ataques DDoS. HIDPS operam nos dispositivos, detectando e corrigindo problemas no sistema operacional (Tomio, 2021) e falhas que poderiam recrutar dispositivos para *botnets* (Dutra & Martins, 2019).

Apesar dos avanços na capacidade computacional dos dispositivos IoT, eles ainda possuem limitações significativas quando comparados a servidores e equipamentos

de rede mais robustos. Essas restrições podem dificultar a implementação de soluções de segurança mais complexas, como o NIDS (Caldas Filho et al., 2023). No entanto, o NIDS pode ser mais eficiente quando associado ao aprendizado de máquina, tornando a detecção de ataques mais precisa e adaptável a novas ameaças.

O IDS (Intrusion Detection System) é uma ferramenta de administração de redes difundida mundialmente (Verma, Bhandari & Singh, 2022), atuando como linha de defesa contra invasões. O NIDS, um tipo especializado de IDS, analisa o tráfego de rede malicioso, requerendo acesso privilegiado para examinar logs e comparar padrões com bancos de dados de assinaturas conhecidas. Combinado ao aprendizado de máquina, o NIDS pode identificar ameaças desconhecidas, aprimorando significativamente a segurança da rede. A integração do aprendizado de máquina ao NIDS torna a detecção de ataques mais eficaz, pois permite o treinamento de modelos baseados em conjuntos de dados previamente selecionados. Dessa forma, o sistema não apenas identifica ataques já conhecidos, mas também detecta novas ameaças a partir da análise de padrões emergentes, ampliando sua capacidade de resposta diante da constante evolução dos ataques cibernéticos.

O objetivo do aprendizado de máquina é desenvolver programas capazes de melhorar seu desempenho com base em modelos existentes. Para isso, é necessária uma grande quantidade de informações, que permitem ao computador gerar conhecimento a partir da análise de dados e formulação de hipóteses (Izbicki & Santos, 2020).

De acordo com (Polat, Polat & Cetin, 2020), a aplicação de modelos de aprendizado de máquina foi aprimorada por métodos de seleção de características, com o intuito de reduzir custos operacionais e, ao mesmo tempo, aumentar a precisão na detecção de ameaças. Esses algoritmos possuem a capacidade de reconhecer padrões sutis e complexos nos dados, muitas vezes imperceptíveis para métodos tradicionais de detecção ou até mesmo para a observação humana.

Além disso, o aprendizado de máquina se destaca por sua capacidade contínua de aprendizado e adaptação a novos cenários, tornando-se uma abordagem especialmente eficaz para análises em ambientes dinâmicos, onde há grande variabilidade nos padrões de tráfego e na diversidade de ataques (Silva et al., 2024).

Um dos segmentos do aprendizado de máquina é o aprendizado federado, onde um modelo é treinado de acordo com os padrões, métodos e parâmetros necessários para a identificação dos ataques DDoS, sem que sejam informados os dados originais, preservando assim a privacidade das informações. Ou seja, cada dispositivo de rede, através de amostras próprias, realiza treinamento individual do modelo sem compartilhamento de dados privados entre eles (Caldas Filho et al., 2023).

Há diversos estudos que buscam mitigar os ataques de negação de serviço distribuído, explorando técnicas de aprendizado de máquina e sistemas de detecção de intrusão, como HIDS, baseados em host, e NIDS, baseados em rede. No entanto, muitas dessas propostas ainda enfrentam desafios e limitações que precisam ser superados. A pesquisa de (Horchulhack et al., 2024) propõe uma abordagem para detecção de intrusões utilizando redes neurais convolucionais (CNNs) e aprendizado por transferência, transformando o fluxo da rede em imagens, permitindo que as CNNs analisem esses dados com maior eficiência. No entanto, o método não foi testado em redes heterogêneas

nem integrou técnicas não supervisionadas, limitando sua aplicabilidade em ambientes IoT diversificados. Enquanto isso, em (Alhasawi & Alghamdi, 2024) é apresentada uma arquitetura descentralizada baseada em aprendizado federado, empregando CNNs para detectar padrões complexos de tráfego em redes IoT. A abordagem alcançou altas taxas de detecção, superiores a 98%, e garantiu a privacidade dos dados por manter o processamento local.

Adicionalmente, (Abreu, Rothenberg & Abelém, 2025) investigou o uso de aprendizado de máquina quântico para detecção de intrusões, e (Sarmiento, Mota & Villacá, 2025) focou na privacidade e eficiência em aprendizado federado, utilizando estruturas de dados probabilísticas. Embora ambos os estudos sejam tecnicamente promissores, eles ainda enfrentam desafios, como o alto custo computacional e a dificuldade de aplicação prática em dispositivos IoT com recursos limitados.

Este trabalho propõe uma solução inovadora para a identificação e análise de ataques DDoS, utilizando um sistema integrado de HIDPS e NIDS. Essa abordagem é capaz de identificar ataques de maneira eficaz através do uso de aprendizado federado. Além disso, a solução foi projetada para garantir que o dispositivo IoT infectado não fique indisponível, assegurando a continuidade do serviço mesmo em face de ataques cibernéticos.

2. Trabalhos Relacionados

Ataques DDoS em redes de telecomunicações, em principal as que são compostas por equipamentos IoT, têm se tornado uma grande preocupação. Diante deste contexto, este trabalho destaca estudos que exploram HIDPS, ataques DDoS para redes voltadas para dispositivos IoT, seja como alvos de ataques ou como agentes que os promovem. Ao longo desta seção, são discutidas as diferentes abordagens adotadas para lidar com esse problema.

Em (Horchulhack et al., 2024), é feita uma abordagem para detecção de intrusões em redes (NIDS) utilizando redes neurais convolucionais (CNNs) e aprendizado por transferência. O trabalho explora a classificação de fluxos de rede e atualizações periódicas para reduzir o custo computacional dos dispositivos com foco na adaptação às mudanças constantes no seu tráfego. O método foi testado com 2,6 TB de tráfego de rede real, demonstrando melhorias de até 19% no F1-Score, em comparação com métodos tradicionais. A proposta é baseada em dados rotulados e aprendizado supervisionado, não realizando a integração de outros métodos, redes específicas como exemplo da IoT e não tendo sido testado em redes heterogêneas como é proposto neste artigo.

Muitos trabalhos de pesquisa propõem soluções diversas para detectar ataques DDoS em redes de telecomunicações, o artigo (Alhasawi & Alghamdi, 2024), apresenta uma solução de detecção de ataques (DDoS) em redes de IoT por meio de aprendizado federado (FL), assim como este trabalho de pesquisa. É proposta uma arquitetura descentralizada utilizando CNNs. De acordo com os autores, os dispositivos IoT são treinados de forma descentralizada através do aprendizado federado e para a detecção dos ataques é utilizado uma CNN devido à sua capacidade de identificar padrões complexos em tráfego de rede, foi utilizado o *dataset* CICIDS2017 que contém tráfego de redes que incluem ataques

DDoS. A pesquisa comprova a eficácia na taxa de detecção de ataques DDoS superior a 98% e também reduziu significativamente os riscos de vazamento das informações proporcionando privacidade, pois os dados permanecem nos dispositivos locais.

Em (Oliveira, 2024) é o F-NIDS (Sistema Federado de Detecção de Intrusão em Rede), uma solução para detectar intrusões em ambientes distribuídos, através do aprendizado federado são abordadas questões relacionadas à privacidade e ao desempenho em redes IoT, este estudo traz escalabilidade horizontal, segurança e privacidade dos dados. Todavia, diferentemente do presente trabalho, não propõe uma estrutura de mitigação capaz de atuar diretamente no host e em sincronia com o NIDS.

(Narote et al., 2022) utiliza aprendizado de máquina para detecção de ataques de (DDoS), através de treinamento de modelos de aprendizado de máquina por meio de algoritmos como Random Forest, Naïve Bayes e AdaBoost para construir modelos de detecção utilizando *dataset* CICIDS2017 e CSE-CIC-IDS2018 e modelos híbridos que são compostos de algoritmos tradicionais e técnicas avançadas como Stochastic Gradient Boosting (SGB). O modelo híbrido desenvolvido neste estudo alcançou 100% de precisão nos testes realizados. De acordo com o estudo, os testes foram realizados apenas com dados conhecidos o que limita possíveis novos ataques e assim como o anterior, foi testado apenas em ambiente simulado. O que se diferencia do presente trabalho que utilizou também dados reais de malware ativo na rede para validar os modelos propostos.

O estudo (Oliveira, 2024), combina computação quântica e clássica para detectar ataques a redes, em sua proposta é utilizada uma técnica híbrida de aprendizado de máquina quântico, como o Classificador Quântico Variacional (VQC), Máquina de Vetores de Suporte com Kernel Quântico (QSVM) e K-Means Quântico (QKM). Foi utilizado *backends* reais como exemplo da IBM Cairo, De acordo com os resultados expostos no artigo, a metodologia desenvolvida retornou um desempenho superior ou comparável aos métodos tradicionais.

A pesquisa de (Sarmiento, Mota & Villacá, 2025) aborda os desafios de privacidade e eficiência na comunicação utilizando aprendizado federado. A proposta combina o uso de estruturas de dados probabilísticas (*sketches*) e algoritmos de seleção de clientes com base em métricas personalizadas. A estrutura de dados probabilística implementa um mecanismo híbrido que combina compactação com adição de ruído, equilibrando privacidade e utilidade, oferecendo uma redução nos custos da transmissão.

O trabalho feito em, (Oliveira, 2024) propõe um framework para detecção de ataque DoS em dispositivos IoT utilizando aprendizado de máquina. O sistema é capaz de detectar, identificar e classificar ataques quase em tempo real. Faz também a integração com ferramentas de monitoramento, como a suíte ELK e o plugin Wazuh. Utilizando o aprendizado de máquina para identificar anomalias no tráfego IoT, este artigo, desenvolve um protótipo de um framework com base em aprendizado supervisionado (Random Forest). O *dataset* utilizado neste projeto foi NSL-KDD para treinar o modelo e criar cenários reais de ataques através da ferramenta Metasploit (Praciano et al., 2019).

O trabalho de pesquisa de (Vasques, 2021), aborda ataques distribuídos de negação de serviço por reflexão amplificada (AR-DDoS), enfatiza a saturação dos dispositivos IoT, que atuam como amplificadores para ataques DDoS através da utilização dos protocolos

SSDP, SNMP e CoAP. A pesquisa avalia o desempenho de diferentes dispositivos e protocolos, buscando identificar o momento em que esses dispositivos atingem sua capacidade máxima de resposta e como isso afeta a amplificação dos ataques. O trabalho traz como contribuição a análise de como funciona o comportamento dos dispositivos IoT diante da reflexão em ataques AR-DDoS.

Em (Yin et al., 2017) é proposto um sistema de detecção de intrusão baseado em redes neurais recorrentes (RNNs), para aprimorar a identificação de comportamentos anômalos no tráfego de rede em ataques tradicionais e mais usuais e também, em ataques multiclases, ou seja, em diferentes tipos de ataques. As RNNs utilizam unidades ocultas conectadas para memorizar padrões sequenciais do tráfego da rede. O modelo é treinado e testado com o conjunto de dados NSL-KDD, que inclui 41 características de tráfego e traz uma eficiência maior em relação aos modelos convencionais, porém, o treinamento do modelo é mais demorado em comparação com métodos tradicionais.

O trabalho (Caldas Filho et al., 2023) desenvolve um modelo híbrido para detecção e mitigação de botnets em rede IoT utilizando aprendizado federado, aprendizado profundo e computação em neblina. A solução proposta é a criação de arquitetura que integra sistemas HIDPS e NIDS, permitindo a detecção e mitigação de tráfego malicioso diretamente nos dispositivos finais reduzindo a dependência da necessidade de uma infraestrutura centralizada melhorando a privacidade dos dados através do treinamento de modelos por aprendizado federado. Seus testes foram realizados em cenário real.

Em (Oliveira, 2024), é proposto um sistema distribuído utilizando aprendizado federado para detectar intrusões em redes IoT. O aprendizado federado mantém a privacidade dos dados não informando para treinamentos os dados originais dos dispositivos superando limitações de privacidade e robustez. O sistema busca oferecer alta precisão, arquitetura descentralizada para suportar demandas dinâmicas e maior confidencialidade dos dados de treinamento. Por meio de testes extensivos, a solução alcança alta precisão na detecção de tráfego malicioso.

3. Metodologia

A arquitetura de um **HIDPS** (Host-based Intrusion Detection and Prevention System) em sistemas modernos e comerciais de segurança de rede é projetada para proteger dispositivos individuais, como servidores e estações de trabalho, contra ameaças cibernéticas. Sistemas HIDPS tem como principais características:

- **Proteção de Pontos Finais:** O HIDPS é integrado nos endpoints (computadores de usuários finais) para monitorar e proteger contra ameaças.
- **Monitoramento de Servidores:** Em ambientes corporativos, servidores são frequentemente protegidos por HIDPS, garantindo que estejam seguros contra acessos não autorizados e ataques.
- **Integração com Outras Ferramentas de Segurança:** A arquitetura de um HIDPS moderno inclui a integração com outros sistemas de segurança, como firewalls, sistemas de gerenciamento de informações e eventos de segurança (SIEM) e plataformas de proteção de endpoints. Essa integração fornece uma defesa em camadas mais robusta.

- **Resposta Automatizada a Ameaças:** Muitos HIDPS modernos possuem capacidade de resposta automatizada, permitindo bloqueio em tempo real de ameaças, isolamento de dispositivos comprometidos e ajustes nas regras de segurança para prevenir ataques futuros.
- **Conformidade e Auditoria:** Organizações que precisam cumprir regulamentações de segurança, como GDPR ou HIPAA, utilizam HIDPS para monitorar e registrar atividades de rede. Isso auxilia na manutenção da conformidade e fornece auditorias detalhadas.
- **Deteção Baseada em Assinaturas e Anomalias:** A arquitetura do HIDPS moderno combina deteção baseada em assinaturas (para ameaças conhecidas) e deteção baseada em anomalias (para ameaças desconhecidas). Essa combinação proporciona uma proteção mais abrangente.

A arquitetura de um HIDPS em sistemas modernos e comerciais de segurança de rede é projetada para proteger dispositivos individuais, como servidores e estações de trabalho, contra ameaças cibernéticas.

O HIDPS proposto nesta solução tem como objetivo mitigar ameaças em dispositivos de Internet das Coisas (IoT). Um dos desafios para sistemas de HIDPS que operam em dispositivos IoT é manter o serviço operando, evitando apenas o tráfego anômalo. Dispositivos IoT infectados por uma botnet podem ter toda sua comunicação bloqueada por soluções de segurança, deixando o serviço indisponível. Câmeras, sensores e atuadores infectados podem ser totalmente bloqueados, interrompendo o funcionamento do serviço. Dessa forma, o objetivo deste trabalho é utilizar plataformas de inteligência artificial para escolher a forma mais eficaz de mitigar falhas de segurança, sem comprometer a disponibilidade do serviço.

Nos sistemas modernos de segurança de rede, o HIDPS é integrado nos endpoints (computadores de usuários finais) para monitorar e proteger contra ameaças como malware, ransomware e outras formas de ataques cibernéticos. Este monitoramento contínuo permite que o sistema identifique e responda rapidamente a atividades suspeitas, garantindo a segurança dos dados e da operação dos dispositivos. Em ambientes corporativos, servidores são frequentemente protegidos por HIDPS, o que garante que dados críticos e aplicativos essenciais estejam seguros contra acessos não autorizados e contra ataques. Essa proteção é fundamental para manter a integridade e a confidencialidade das informações da organização.

O HIDPS pode ser integrado a outras plataformas de segurança, essa integração cria uma defesa em camadas mais robustas, permitindo que as organizações identifiquem e respondam a ameaças de forma mais eficiente. Além disso, muitos HIDPS modernos possuem capacidades de resposta automatizada, permitindo o bloqueio em tempo real de ameaças, o isolamento de dispositivos comprometidos e ajustes nas regras de segurança para prevenir ataques futuros. Essa resposta automatizada é crucial para minimizar o impacto de um ataque e garantir a continuidade dos serviços.

Além disso, a arquitetura do HIDPS moderno combina deteção baseada em assinaturas (para ameaças conhecidas) e deteção baseada em anomalias (para ameaças desconhecidas). Essa combinação proporciona uma proteção mais abrangente, permitindo que o sistema identifique e responda a um amplo espectro de ameaças. Ao

utilizar plataformas de inteligência artificial, o HIDPS pode analisar grandes volumes de dados e identificar padrões que poderiam passar despercebidos por métodos tradicionais. A inteligência artificial permite uma análise mais sofisticada e uma resposta mais precisa às ameaças, o que é fundamental para proteger dispositivos IoT sem comprometer a disponibilidade do serviço.

A prevenção de falhas em sistemas de segurança, especialmente no contexto de dispositivos IoT, requer soluções específicas para diferentes níveis de monitoramento. O HIDS, que monitora processos e tráfego em um único dispositivo é particularmente útil para detectar atividades suspeitas localmente, ou seja, no próprio dispositivo. No entanto, o HIDS apresenta limitações significativas devido à restrição de recursos, como processamento e memória, típicos de dispositivos IoT. Essa limitação pode impedir a identificação de padrões maliciosos mais complexos que exigem uma visão global do tráfego. Por exemplo, considere uma rede de dispositivos IoT infectados, organizados em uma botnet, tentando realizar um ataque de negação de serviço (DoS) a sites governamentais. Cada dispositivo individualmente, pode interpretar suas requisições constantes como normais, pois não tem a capacidade de correlacionar o tráfego global da rede. No entanto, em um contexto mais amplo, onde múltiplos dispositivos estão enviando inúmeras requisições ao mesmo site, sem finalizá-las, o comportamento coletivo configura claramente um ataque DDoS. Nesse cenário, um NIDS se torna essencial, pois tem a capacidade de analisar o tráfego de rede de forma abrangente, identificando padrões anômalos e detectando ataques por meio do comportamento geral da rede, em vez de focar apenas em atividades individuais. Essa distinção evidencia a importância de integrar sistemas de detecção baseados em host e rede para criar uma defesa mais robusta contra ameaças cibernéticas.

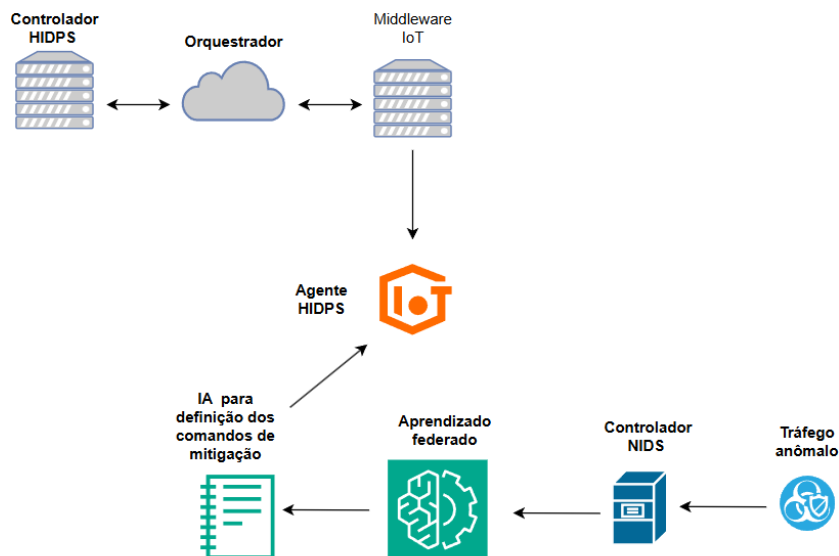


Figura 1 – Arquitetura Proposta do HIDPS

A solução proposta neste trabalho desenvolve um ambiente em que os dispositivos IoT são equipados com HIDPS configurados com um conjunto de regras implementadas para direcionar seu comportamento de acordo com possíveis ataques DDoS. Estas regras monitoram processos no sistema operacional, fazem o gerenciamento de fluxo da rede e realizam o bloqueio dos dispositivos IoT e também de protocolos, desde que este fluxo passa através do servidor central conforme é demonstrado na figura. Os dispositivos IOT são monitorados em tempo real, assimila as regras configuradas no HIDPS e associada ao aprendizado federado identifica os padrões do tráfego da rede produzindo uma solução de cibersegurança robusta contra ataques DDoS.

4. Resultados e Análises

Para validar a eficácia da infraestrutura proposta na mitigação de tráfego anômalo gerado por botnets, realizamos um conjunto de três testes distintos. O objetivo desses experimentos foi demonstrar a capacidade do sistema HIDPS de identificar e bloquear tráfego malicioso de forma seletiva, garantindo a integridade e disponibilidade dos serviços de Internet das Coisas (IoT). A ativação do HIDPS foi conduzida a partir da estrutura do NIDS, que, ao detectar padrões de tráfego anômalos, acionou a aplicação progressiva de regras de filtragem. Essas regras foram implementadas de maneira gradual, restringindo o número de conexões estabelecidas com o servidor de destino. Essa abordagem evitou a sobrecarga do servidor sem interromper completamente o serviço, assegurando a continuidade das operações para dispositivos legítimos da rede.

A Figura 2 ilustra as regras de filtragem aplicadas durante um dos testes realizados. Neste experimento, um ataque conduzido por dois bots operando sob controle do malware Mirai foi direcionado contra a infraestrutura-alvo. O sistema HIDPS detectou o tráfego anômalo em tempo real e, como resposta inicial, aplicou uma regra de limitação da taxa de pacotes enviados ao servidor. Essa medida preventiva reduziu significativamente o risco de sobrecarga dos recursos computacionais, mitigando o impacto do ataque sem comprometer a disponibilidade do serviço.

Para avaliar o desempenho do sistema, foram conduzidas três baterias de testes, focadas na medição do tempo médio necessário para a mitigação do tráfego malicioso. Durante esses experimentos, o HIDPS aplicou diferentes comandos de contenção, como limitação do número de conexões simultâneas e bloqueio progressivo de IPs responsáveis pelo ataque.

Os resultados obtidos indicaram que o HIDPS foi capaz de neutralizar completamente os ataques em um tempo médio de 3 segundos. Essa resposta rápida demonstra a eficiência do modelo proposto na identificação e mitigação de tráfego anômalo sem comprometer a operação normal dos dispositivos IoT protegidos. Além disso, o tempo de resposta do sistema evidencia a vantagem do uso de inteligência artificial, que permitiu a adaptação dinâmica das regras de mitigação conforme a evolução do ataque. Esses testes validam a eficácia do HIDPS em cenários reais, garantindo que dispositivos IoT permaneçam operacionais mesmo sob ataques cibernéticos, minimizando indisponibilidades e preservando a integridade da rede.

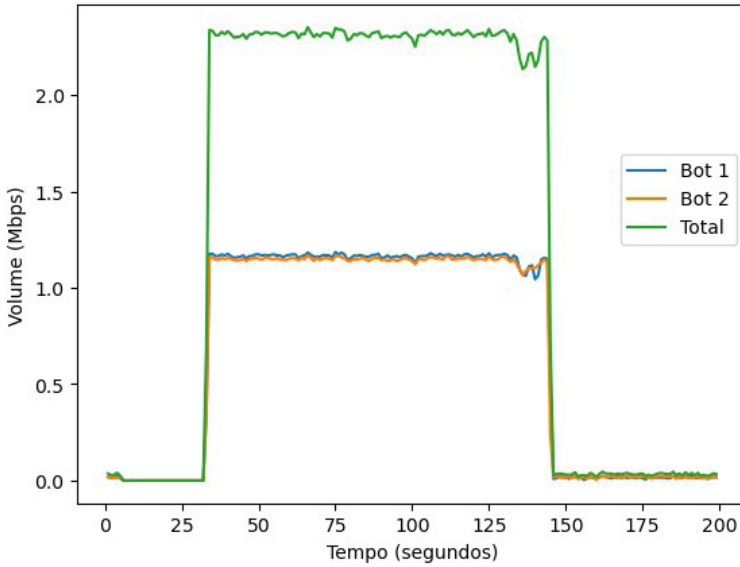


Figura 2 – Tráfego da Simulação

Os resultados demonstraram que a abordagem proposta foi capaz de identificar e conter o tráfego malicioso de forma eficaz, minimizando os impactos negativos para dispositivos IoT legítimos conectados à rede. A integração do NIDS com o modelo de aprendizado de máquina baseado em Random Forest permitiu uma adaptação dinâmica das regras de mitigação, tornando o sistema mais resiliente e eficiente contra ataques de negação de serviço distribuídos.

5. Conclusões e Trabalhos Futuros

Este trabalho apresenta um sistema híbrido de detecção e prevenção de intrusões para mitigar ataques DDoS em redes IoT. A arquitetura é composta por dois componentes principais: HIDPS Agente, implantado em dispositivos IoT para filtrar tráfego malicioso, e HIDPS Controlador, um sistema baseado em nuvem que orquestra a distribuição de regras.

Um Sistema de Detecção de Intrusão em Rede (NIDS) baseado em Aprendizado Federado é integrado para detectar anomalias sem expor dados brutos. O framework Flower é utilizado para treinar modelos distribuídos, aumentando a privacidade e a detecção de ataques. Aplicando regras de filtragem de forma dinâmica e mitigação adaptativa, o sistema neutraliza efetivamente o tráfego anômalo na origem, reduzindo a sobrecarga nos servidores. Os resultados experimentais confirmam que a abordagem proposta bloqueia o tráfego malicioso sem comprometer a disponibilidade dos serviços IoT.

Agradecimentos

Os autores agradecem o apoio do Laboratório LATITUDE, da Universidade de Brasília, ao TED 01/2021 da Secretaria Nacional de Assistência Social – SNAS/DGSUAS/CGRS, ao TED 01/2021 da Procuradoria Geral da Fazenda Nacional – PGFN, ao Projeto SISTER City (Outorga 625/2022 – FAP/DF), ao Projeto SISPRO-DF” (Outorga 497/2023 – FAP/DF), ao Decanato de Pesquisa e Inovação – DPI/UnB e a FAP/DF.

Referencias

- Abreu, D.; Rothenberg, C. R. E.; Abelém, A. QIDS: sistema de detecção de ataques baseado em aprendizado de máquina quântico híbrido. In: SBRC. Universidade Federal do Pará (UFPA) e Universidade Estadual de Campinas (UNICAMP), 2025.
- Ahmad, R.; Alsmadi, I.; Alhamdani, W.; Tawalbeh, L. Zero-day attack detection: a systematic literature review. *Artificial Intelligence Review*, v. 56, n. 10, p. 10733–10811, 2023.
- Alhasawi, Y.; Alghamdi, S. Federated learning for decentralized DDoS attack detection in IoT networks. *IEEE Access*, v. 12, p. 42357–42368, 2024.
- Bala, B.; Behal, S. AI techniques for IoT-based DDoS attack detection: taxonomies, comprehensive review and research challenges. *Computer Science Review*, v. 52, 100631, 2024.
- Caldas Filho, F. L. de; Soares, S. C. M.; Oroski, E.; Albuquerque, R. de O.; Mata, R. Z. A. da; Mendonça, F. L. L. de; Sousa Júnior, R. T. de. Botnet detection and mitigation model for IoT networks using federated learning. *Sensors*, v. 23, n. 14, 6305, 2023.
- Debicha, I.; Cochez, B.; Kenaza, T.; Debatty, T.; Dricot, J.-M.; Mees, W. Advbot: realistic adversarial botnet attacks against network intrusion detection systems. *Computers & Security*, v. 129, 103176, 2023.
- Dutra, B. V.; Martins, L. M. C. E. HIDS by signature for embedded devices in IoT networks. In: ACTAS DE LAS V JORNADAS NACIONALES DE CIBERSEGURIDAD. p. 53–61. Servicio de Publicaciones, 2019.
- Horchulhack, P.; Viegas, E. K.; Santin, A. O.; Simioni, J. A. Fortalecendo a segurança de redes: um olhar profundo na detecção de intrusões com CNN baseada em imagens e aprendizado por transferência. In: SBRC, 2024.
- Izbicki, R.; Santos, T. M. dos. Aprendizado de máquina: uma abordagem estatística. Rafael Izbicki, 2020.
- Narote, A.; Zutshi, V.; Potdar, A.; Vichare, R. Detection of DDoS attacks using concepts of machine learning. *International Journal for Research in Applied Science & Engineering Technology*, v. 10, p. 390–403, 2022.
- Oliveira, J. A. de. F-NIDS: sistema de detecção de intrusão baseado em aprendizado federado, 2024.

- Polat, H.; Polat, O.; Cetin, A. Detecting DDoS attacks in software-defined networks through feature selection methods and machine learning models. *Sustainability*, v. 12, n. 3, 1035, 2020.
- Praciano, B. J.; Caldas Filho, F. L. de; Lucas, M.; Martins, D. F.; Silva, D. A. da; Sousa Júnior, R. T. de. Segurança do ambiente usando dispositivo IoT com processamento distribuído. In: *ATAS DAS CONFERÊNCIAS IBERO-AMERICANAS WWW/INTERNET 2019 E COMPUTAÇÃO APLICADA 2019*. p. 163–170, 2019.
- Sarmiento, E. M. M.; Mota, V. F. S.; Villacá, R. S. Privacidade e comunicação eficiente em aprendizado federado: uma abordagem utilizando estruturas de dados probabilísticas e seleção de clientes. In: *SBRC. Programa de Pós-Graduação em Informática (PPGI), Universidade Federal do Espírito Santo (UFES)*, 2025.
- Silva, R. R.; Henriques, F. R.; Moraes, I. M.; Mascarenhas, D. M. Detecção de ataques de negação de serviço distribuídos com algoritmos de aprendizado de máquina, 2024.
- Tomio, R. L. Um modelo baseado em aprendizado de máquina para detecção autônoma de intrusão. 2021. Tese (Doutorado) – Pontifícia Universidade Católica do Paraná, Brasil.
- Vasques, A. T. Análise de saturação de dispositivos IoT atuando como refletores em ataques distribuídos de negação de serviço por reflexão amplificada, 2021.
- Verma, J.; Bhandari, A.; Singh, G. INIDS: SWOT analysis and TOWS inferences of state-of-the-art NIDS solutions for the development of intelligent network intrusion detection system. *Computer Communications*, v. 195, p. 227–247, 2022.
- Yin, C.; Zhu, Y.; Fei, J.; He, X. A deep learning approach for intrusion detection using recurrent neural networks. *IEEE Access*, v. 5, p. 21954–21961, 2017.

Juris Syntax: Automação da Análise Jurídica Brasileira com IA Generativa e PLN

Fábio Lúcio Lopes de Mendonça¹, Bruno Justino Garcia Praciano¹,
Flávio Garcia Praciano¹, Thiago Leite de Sousa¹, Elon Oliveira Albuquerque¹,
José Péricles Pereira de Sousa²

**fabio.mendonca@redes.unb.br; bruno.justino@ieee.org; flavio.praciano@redes.unb.br;
thiago.leite@redes.unb.br; elon.albuquerque@gmail.com; jose.pereira-sousa@pgfn.gov.br**

¹ Universidade de Brasília, Faculdade de Tecnologia - FT – 70910-900 – Brasília, Brasil

² Procuradoria Geral da Fazenda Nacional – PGAJUD/PGDAU – 70830-030 – Brasília, Brasil

Pages: 13-27

Resumo: Este artigo apresenta o Juris Syntax, um sistema inteligente para análise e contextualização de documentos jurídicos, desenvolvido com foco no Direito brasileiro. A solução integra Processamento de Linguagem Natural (PLN), modelos de similaridade semântica e Inteligência Artificial generativa, possibilitando extração de informações, fichamento, sumarização, busca por precedentes e geração de relatórios estruturados. Sua arquitetura combina backend em Python 3.11, embeddings com Sentence Transformers e integração ao modelo LLaMA 3.1 70B na Oracle Cloud, além de uma interface web responsiva. Avaliado com documentos reais da Procuradoria-Geral da Fazenda Nacional (PGFN), o Juris Syntax demonstrou redução significativa do tempo de análise e alta precisão na identificação de documentos similares, reforçando sua aplicabilidade prática. Os resultados indicam o potencial da ferramenta para otimizar fluxos jurídicos, apoiar a tomada de decisão baseada em precedentes e contribuir para a transformação digital no setor jurídico.

Palavras-chave: Processamento de Linguagem Natural; Inteligência Artificial Generativa; Direito; Similaridade Semântica; Oracle Cloud; Automação Jurídica.

Juris Syntax: Intelligent System for Legal Analysis and Contextualization with Natural Language Processing Techniques

Abstract: Juris Syntax is an intelligent system designed to streamline the analysis and contextualization of legal documents, with emphasis on Brazilian Law. The system integrates advanced Natural Language Processing (NLP), semantic similarity models, and generative Artificial Intelligence to support tasks such as information extraction, summarization, semantic search, and structured reporting. Its architecture combines a Python 3.11 backend, embeddings with Sentence Transformers, and integration with the LLaMA 3.1 70B model hosted on Oracle Cloud, together with a modern and responsive web interface. Tested with real cases from the Office of the National Treasury Attorney (PGFN), Juris Syntax

demonstrated significant reduction in document analysis time and improved accuracy in identifying similar cases and legal precedents. The results highlight its potential to optimize legal workflows, strengthen decision-making, and contribute to the digital transformation of the legal field.

Keywords: Natural Language Processing; Generative Artificial Intelligence; Law, Semantic Similarity; Oracle Cloud, Legal Automation.

1. Introdução

O sistema jurídico brasileiro enfrenta uma crescente sobrecarga informacional. Apenas em 2024, havia mais de 77 milhões de processos em tramitação, muitos com duração superior a cinco anos em determinadas áreas (Conselho Nacional de Justiça, 2024). Órgãos como a Procuradoria-Geral da Fazenda Nacional (PGFN) lidam diariamente com milhares de petições, recursos e decisões, tornando inviável a análise manual diante da complexidade e do volume documental.

Nesse cenário, soluções de Inteligência Artificial (IA) e Processamento de Linguagem Natural (PLN) têm ganhado espaço em aplicações jurídicas internacionais, tais como sumarização automática, classificação de casos e geração de relatórios. Modelos de linguagem de grande porte (LLMs), tais como GPT e LLaMA, já demonstraram resultados significativos, mas ainda carecem de adaptação às especificidades do Direito brasileiro. Iniciativas nacionais como o BERTimbau e o RoBERTaLexPT representam avanços, porém, muitas vezes não contemplam integração ponta a ponta, segurança adequada ou usabilidade para operadores jurídicos (Garcia et al., 2024; R. et al., 2022).

Apesar desses avanços, ainda persistem lacunas importantes. Muitos sistemas comerciais e acadêmicos não se encontram adaptados às peculiaridades da legislação e da prática processual brasileiras, carecem de integração com fluxos de trabalho reais e, frequentemente, não oferecem garantias de privacidade e segurança adequadas para o tratamento de dados sensíveis. Além disso, poucos oferecem mecanismos de busca por similaridade semântica com foco específico no direito brasileiro, recurso essencial para identificar precedentes relevantes mesmo quando redigidos com terminologias distintas.

O Juris Syntax surge como resposta a essas lacunas trata-se de uma plataforma modular que combina: (i) extração robusta de texto a partir de arquivos PDF, mesmo em documentos extensos ou digitalizados; (ii) geração de embeddings semânticos especializados no domínio jurídico; (iii) busca por similaridade baseada em índices vetoriais de alta performance; e (iv) sumarização contextualizada com o apoio de LLMs e pipelines de *Retrieval-Augmented Generation* (RAG). O sistema foi projetado para operar tanto em ambientes locais quanto em nuvem, com controle rigoroso sobre a confidencialidade e integridade dos dados.

Este artigo apresenta a concepção, implementação e avaliação do *Juris Syntax*, discutindo seus aspectos técnicos, metodológicos e práticos. São descritas a arquitetura do sistema, as tecnologias utilizadas, o pipeline de processamento, as estratégias de segurança, bem como um estudo de caso conduzido na PGFN. Também são analisados

trabalhos relacionados nacionais e internacionais publicados nos últimos quatro anos, situando o *Juris Syntax* no estado da arte e evidenciando suas contribuições originais para a modernização da atividade jurídica no Brasil.

2. Comparativo com Trabalhos Relacionados

2.1. Análise Gráfica: Comparativo entre o Juris Syntax e Trabalhos Correlatos

O gráfico de linhas a seguir apresenta uma visualização comparativa entre o projeto Juris Syntax e os trabalhos brasileiros e estrangeiros, avaliando-os em diferentes dimensões. Essa abordagem permite identificar padrões de desempenho, evidenciar pontos fortes e apontar oportunidades de melhoria em relação ao estado da arte. Além disso, a comparação possibilita observar tendências de evolução tecnológica e metodológica no campo jurídico. Por fim, a análise facilita a compreensão do posicionamento do Juris Syntax no cenário nacional e internacional, oferecendo subsídios para decisões estratégicas de desenvolvimento. Conforme demonstrado na Figura 1 e Tabela I.

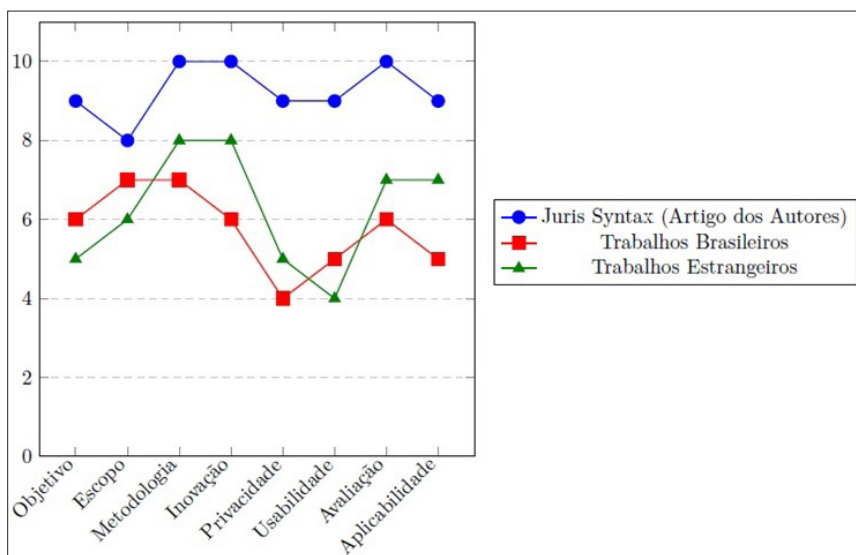


Figura 1 – Comparativo: Juris Syntax vs. Outros Trabalhos

O gráfico destaca a performance superior do Juris Syntax em dimensões críticas como Metodologia e Inovação, atingindo a pontuação máxima de 10. Ele também demonstra um foco robusto em áreas negligenciadas por outros estudos, como Privacidade e Usabilidade, consolidando a sua abordagem como a mais completa e pronta para aplicação prática.

Dimensão	Juris Syntax (Este Artigo)	Trabalhos Brasileiros (10)	Trabalhos Estrangeiros (10)
Objetivo principal	Plataforma modular, integrada e validada para análise, fichamento, sumarização contextualizada e busca semântica de documentos jurídicos no Brasil, com foco no Direito brasileiro e potencial de expansão internacional.	Revisões e estudos de PLN jurídico (Polo et al., 2021; Silva Junior et al., 2025); aplicações institucionais como o sistema Victor/STF (Ferreira, 2022) e Athos/STJ (Figueiredo, 2022); propostas de <i>embeddings</i> jurídicos (Carmo et al., 2023); e integração inicial de IA em tribunais (Pereira et al., 2024), ainda com foco limitado em integração ponta a ponta.	Propostas exploratórias e prototípicas para sumarização, classificação e geração automática de textos (Ariai & Demartini, 2024; Bertalan & Ruiz, 2022), geralmente sem validação em ambientes reais.
Escopo	Voltado ao Direito brasileiro, mas arquitetado de forma modular para adaptação a outros ordenamentos e idiomas.	Cobertura predominantemente restrita ao contexto jurídico nacional (Campos de Carvalho et al., 2022; Melo et al., 2023), com pouca atenção a aplicações multilíngues ou internacionais.	Cobertura ampla de contextos jurídicos e sistemas legais (Silva, 2025; Anonymous, 2025), mas sem ajustes às especificidades brasileiras.
Metodologia técnica	Pipeline completo que abrange extração robusta de texto (PDF), geração de <i>embeddings</i> jurídicos, indexação e busca vetorial, RAG, sumarização por LLaMA 3.1, interface web responsiva e protocolos de segurança e privacidade	Utilização de modelos BERTimbau e RoBERTaLexPT (Polo et al., 2021), <i>embeddings</i> específicos (Carmo et al., 2023) e aplicações pontuais de LLMs (Pereira et al., 2024), geralmente sem integração de todas as etapas em um único sistema.	Emprego de LLMs (GPT, LLaMA, RoBERTa) (Huang & Chang, 2025; Rahman et al., 2025; Goyal et al., 2022), RAG e Knowledge Graphs (Ariai & Demartini, 2024), com forte foco em desempenho técnico, mas menos em orquestração completa de processos.
Inovação	Combina múltiplas funcionalidades críticas em um único sistema validado em órgão público (PGFN), priorizando escalabilidade, modularidade e confidencialidade.	Avanços conceituais e metodológicos (Barros et al., 2024; Silva Junior et al., 2025), mas com pouca evidência de aplicação prática em larga escala.	Introduzem novos métodos e <i>benchmarks</i> (Akter et al., 2025; Silva, 2025), porém com escassa transposição para uso institucional.
Tecnologias usadas	Backend em Python, Flask, Py-MuPDF, Sentence Transformers, integração com LLaMA 3.1 na Oracle Cloud, containerização com Docker e frontend em Tailwind CSS.	Modelos BERTimbau, RoBERTaLexPT e <i>embeddings</i> jurídicos treinados localmente (Carmo et al., 2023); uso de linguagens como R em análises específicas (Barros et al., 2024).	Arquiteturas baseadas em Transformers, técnicas de <i>prompt engineering</i> e infraestrutura em nuvens como AWS e Azure (Ariai & Demartini, 2024; Goyal et al., 2022).

Dimensão	Juris Syntax (Este Artigo)	Trabalhos Brasileiros (10)	Trabalhos Estrangeiros (10)
Privacidade e segurança	Processamento local como prioridade, uso de SDKs seguros para nuvem, variáveis de ambiente para credenciais e remoção automática de arquivos após processamento.	Pouco detalhamento técnico sobre segurança e conformidade com LGPD (Ferreira, 2022; Figueiredo, 2022).	Menções pontuais ao GDPR (Rahman et al., 2025), mas sem aprofundamento técnico operacional.
Usabilidade	Interface web intuitiva e responsiva, com drag-and-drop, exportação em PDF/CSV e histórico de análises.	Interfaces pouco exploradas ou documentadas (Barros et al., 2024; Campos de Carvalho et al., 2022), com foco maior em <i>backend</i> .	Grande parte das soluções acessível apenas via API ou linha de comando (Anonymous, 2025; Goyal et al., 2022), limitando o alcance a usuários técnicos.
Avaliação	Testado com documentos reais e volumosos da PGFN, medindo eficiência, precisão e adequação ao fluxo de trabalho.	Testes com <i>datasets</i> restritos ou simulados (Barros et al., 2024; Silva Junior et al., 2025), sem avaliação operacional contínua.	Avaliações baseadas em <i>corpora</i> públicos como COLIEE (Ariai & Demartini, 2024; Akter et al., 2025), que não refletem necessariamente a realidade jurídica brasileira.
Aplicabilidade	Implementável em fluxo de trabalho de órgãos públicos, escritórios de advocacia e departamentos jurídicos, com ganho mensurável de eficiência.	Potencial teórico identificado, mas com limitações de maturidade para uso em produção (Campos de Carvalho et al., 2022; Melo et al., 2023).	Aplicabilidade ampla em contextos internacionais (Silva, 2025; Silva, 2024), mas sem compatibilidade direta com a legislação e práticas brasileiras.

Tabela 1 – Comparativo entre o Juris Syntax e os trabalhos nacionais e internacionais analisados

3. Visão Geral do Juris Syntax

O projeto *Juris Syntax* foi concebido com a finalidade de atender a um conjunto de necessidades críticas identificadas no contexto do processamento e gestão de documentos jurídicos, particularmente no cenário brasileiro, caracterizado por elevado volume documental e complexidade normativa. A proposta busca combinar eficiência operacional com precisão analítica, explorando o potencial de técnicas avançadas de Processamento de Linguagem Natural (PLN) e Inteligência Artificial (IA) generativa para otimizar fluxos de trabalho em órgãos públicos, escritórios de advocacia e departamentos jurídicos corporativos.

Entre seus objetivos centrais, destaca-se a automatização integral do processo de análise e fichamento de petições, recursos e outros documentos processuais, permitindo a extração sistemática de informações essenciais, tais como identificação das partes, pedidos formulados, fundamentos jurídicos e decisões proferidas. Complementarmente, a ferramenta visa oferecer mecanismos robustos para a geração de resumos estruturados e relatórios detalhados, a partir de arquivos em PDF ou texto puro, favorecendo a rápida compreensão e o tratamento de grandes volumes de informação.

Outro objetivo relevante consiste na implementação de um módulo especializado para busca por similaridade semântica, capaz de identificar, com elevado grau de acurácia, documentos juridicamente correlatos. Essa funcionalidade proporciona suporte valioso à pesquisa jurisprudencial, à análise comparativa de peças processuais e à tomada de decisão baseada em precedentes.

O projeto também contempla o desenvolvimento de uma interface web moderna, responsiva e intuitiva, projetada segundo princípios de usabilidade e acessibilidade, de forma a atender perfis heterogêneos de operadores do Direito, independentemente de seu nível de proficiência tecnológica. Adicionalmente, a integração de recursos avançados de IA garante não apenas a precisão das análises, mas também a sua contextualização, adaptando os resultados ao tipo e à natureza do documento analisado. Conforme demonstrado nas Figuras 2 e 3.

No que se refere à segurança e à privacidade, o *Juris Syntax* adota uma abordagem de proteção por padrão, privilegiando o processamento local sempre que possível, aliado a mecanismos seguros de comunicação com a nuvem por meio de SDKs oficiais e variáveis de ambiente. Com isso, o sistema se posiciona como uma solução escalável, flexível e de alta confiabilidade, apta a contribuir de forma significativa para a redução de custos operacionais, a mitigação de riscos e o incremento da eficiência na gestão de informações jurídicas.

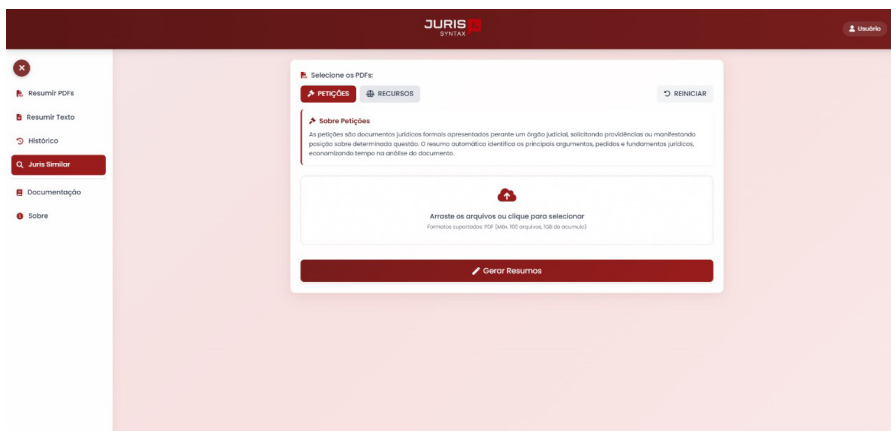


Figura 2 – Juris Similar

4. Metodologia e Arquitetura

A concepção arquitetural do *Juris Syntax* fundamentase em uma abordagem modular e escalável, estruturada em três camadas principais: Backend, Frontend e Infraestrutura. Essa organização visa garantir separação de responsabilidades, flexibilidade para evolução tecnológica e manutenção facilitada, preservando a integridade do sistema em ambientes de alta demanda.

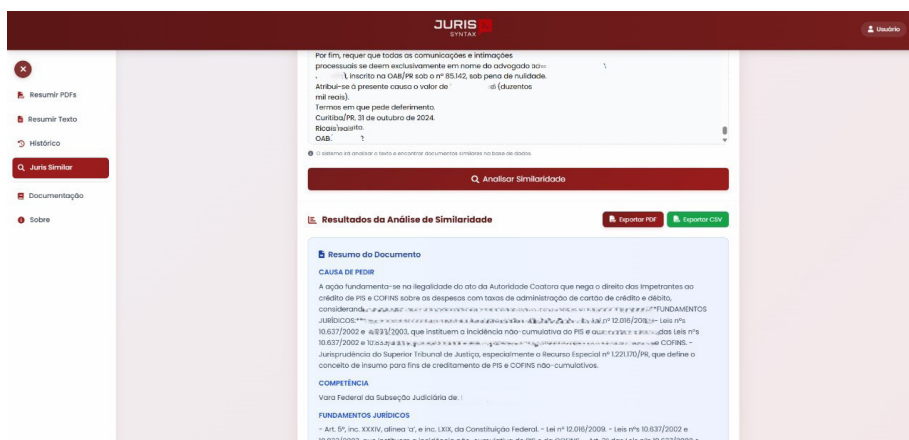


Figura 3 – Análise de similaridade

No Backend, desenvolvido integralmente em Python 3.11, concentra-se a lógica de negócio e o processamento central das funcionalidades. A aplicação é estruturada sobre o *framework* Flask 2.3.2, que fornece a base para a construção da API e para o roteamento das operações internas. A extração de conteúdo textual a partir de arquivos PDF é realizada por meio das bibliotecas *PyPDF2* e *PyMuPDF*, garantindo robustez mesmo diante de documentos extensos ou com formatação heterogênea. A geração de relatórios e a exportação em formato PDF são conduzidas pela biblioteca *FPDF2*, permitindo a produção de documentos profissionais e padronizados. Para a análise de similaridade semântica, o sistema emprega o modelo *Sentence Transformers (all-MiniLM-L6-v2)*, responsável pela criação de *embeddings* vetoriais de alta qualidade. A integração com a *Oracle Cloud Generative AI*, utilizando o modelo Llama 3.1 70B, possibilita a sumarização contextualizada de textos jurídicos e a geração de insights adaptados ao tipo de documento processado. Operações de manipulação de dados e matrizes de *embeddings* são suportadas por *Pandas* e *NumPy*, enquanto a containerização e orquestração do ambiente são realizadas por *Docker* e *Docker Compose*. A comunicação segura com os serviços em nuvem é viabilizada pelo *OCI SDK*, utilizando variáveis de ambiente para proteção de credenciais.

O *Frontend* é responsável pela interação direta com o usuário e foi implementado com tecnologias web modernas, incluindo HTML5 e JavaScript, proporcionando uma estrutura leve e de rápida resposta. A estilização é conduzida com *Tailwind CSS*, o que assegura responsividade, consistência visual e personalização da interface. Elementos gráficos e ícones são incorporados por meio da biblioteca *Font Awesome*, favorecendo a clareza e a atratividade da experiência visual.

A camada de Infraestrutura adota um modelo baseado em containerização, o que permite a execução tanto em ambientes locais quanto em plataformas de nuvem, como *Google Cloud Run* e *Oracle Cloud*. Essa arquitetura garante portabilidade e reprodutibilidade do ambiente de execução, reduzindo riscos de incompatibilidade. Arquivos de configuração, incluindo *.env*, *.yml* e documentação em *.md*, foram incorporados para simplificar o

processo de implantação (*deploy*) e manutenção, além de padronizar o gerenciamento de variáveis e parâmetros operacionais.

Essa arquitetura, além de oferecer um elevado grau de modularidade e segurança, está orientada para o processamento eficiente de documentos jurídicos em larga escala, mantendo a flexibilidade necessária para futuras expansões e adaptações a novos cenários de aplicação. Conforme demonstrado na Arquitetura do sistema representado pela Figura 4.

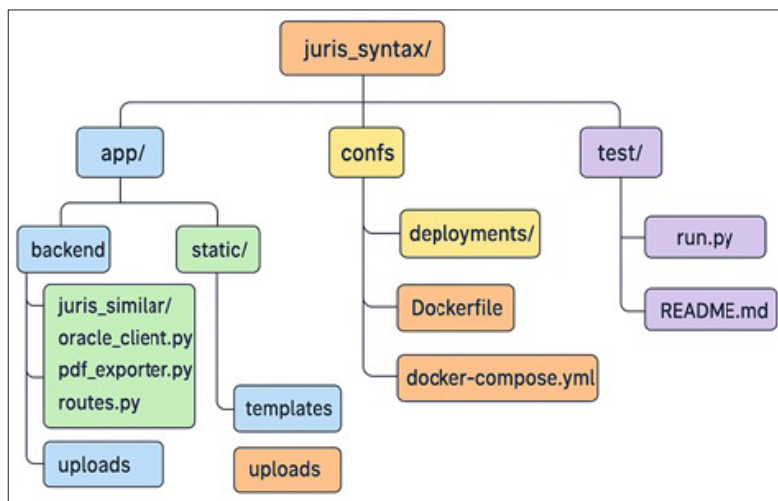


Figura 4 – Arquitetura do Juris Syntax

A Figura 4 apresenta a arquitetura do projeto Juris Syntax, estruturada de forma modular para garantir organização, escalabilidade e facilidade de manutenção. No diretório raiz *juris_syntax/* concentram-se os principais elementos que compõem o sistema. O núcleo da aplicação encontra-se no diretório *app/*, subdividido em componentes essenciais para o funcionamento do sistema.

O subdiretório *backend/* concentra a lógica de processamento e integração, incluindo o módulo *juris_similar/*, responsável pela análise de similaridade textual entre documentos jurídicos; o script *oracle_client.py*, que realiza a integração com os serviços de inteligência artificial da Oracle Cloud AI; o módulo *pdf_exporter.py*, dedicado à exportação de documentos em formato PDF; e o arquivo *routes.py*, que define as rotas principais para comunicação entre o frontend e o backend. Outros módulos utilitários complementam a infraestrutura de processamento de dados.

A camada de apresentação é composta pelos diretórios *static/*, que armazena arquivos estáticos como CSS, JavaScript e imagens; e *templates/*, que contém os modelos HTML utilizados para renderização da interface web. O diretório *uploads/* é destinado ao armazenamento temporário de arquivos enviados pelos usuários.

Em nível de configuração, o diretório *confs/* centraliza parâmetros e prompts de IA utilizados pelo sistema. Já o diretório *deployments/* contém scripts e configurações

voltadas à implantação do sistema em diferentes ambientes. O diretório `test/` abriga os testes automatizados que asseguram a qualidade e a estabilidade do código.

No que diz respeito à infraestrutura, o arquivo `requirements.txt` define as dependências do projeto, enquanto `Dockerfile` e `docker-compose.yml` oferecem suporte à containerização e orquestração, facilitando a replicação e escalabilidade do ambiente. Por fim, o arquivo `run.py` atua como ponto de entrada da aplicação, e o `README.md` fornece documentação introdutória para desenvolvedores e usuários técnicos.

Essa organização hierárquica e modular foi concebida para otimizar o desenvolvimento, manutenção e evolução do Juris Syntax, permitindo que as equipes de pesquisa e tecnologia atuem de forma integrada e eficiente, com uma arquitetura robusta e preparada para futuras expansões.

5. Funcionalidades

O sistema desenvolvido incorpora um conjunto abrangente de funcionalidades voltadas à otimização do tratamento e análise de documentos jurídicos, com foco na automação e na precisão dos resultados. Entre os principais recursos, destaca-se a análise automatizada de documentos, que realiza a extração estruturada de informações relevantes, o fichamento dos conteúdos e a geração de resumos sintéticos, permitindo a rápida compreensão de grandes volumes de texto. Adicionalmente, foi implementada uma busca por similaridade semântica, baseada em técnicas avançadas de Processamento de Linguagem Natural (PLN) e vetorização contextual, possibilitando a comparação entre documentos de forma mais precisa do que métodos tradicionais de busca por palavras-chave.

Outro diferencial é a integração com modelos de inteligência artificial generativa, que viabiliza a produção de contextualizações aprofundadas e a obtenção de insights jurídicos, contribuindo para uma análise mais qualificada. O sistema também oferece exportação de relatórios em múltiplos formatos, assegurando interoperabilidade com outras ferramentas de gestão e análise de dados, além de registrar todas as operações no histórico de análises, garantindo rastreabilidade e reprodutibilidade dos resultados. Por fim, dispõe de uma interface web moderna e responsiva, com navegação por abas, suporte a drag-and-drop e design adaptativo, assegurando usabilidade consistente em diferentes dispositivos.

6. Segurança e Privacidade

A segurança e a privacidade dos dados foram elementos centrais no desenvolvimento do sistema. Sempre que viável, o processamento dos documentos é realizado de forma local, minimizando a exposição de informações sensíveis a ambientes externos. Quando a integração com serviços em nuvem é necessária, esta ocorre por meio de um SDK seguro e do uso de variáveis de ambiente, garantindo a proteção das credenciais e a criptografia das comunicações. Além disso, os arquivos são armazenados apenas temporariamente durante o processamento, sendo automaticamente removidos após a conclusão das operações, em conformidade com as melhores práticas de governança de dados e privacidade.

7. Fluxo de Utilização do Sistema

O processo de interação com o sistema inicia-se quando o usuário acessa a interface web e seleciona o tipo de análise desejada. Em seguida, o usuário realiza o upload de arquivos no formato PDF ou insere o conteúdo manualmente em um campo de texto. O sistema processa o material submetido, executando tarefas de extração, fichamento e sumarização, bem como a busca por documentos semanticamente similares caso essa funcionalidade seja requisitada. Os resultados gerados são apresentados de forma clara na interface, com opções para visualização detalhada, exportação em múltiplos formatos (por exemplo, PDF e CSV) e compartilhamento. Adicionalmente, todo o histórico de análises é registrado com metadados relevantes (data, usuário, parâmetros da análise), permitindo rastreabilidade, reusabilidade e auditoria das operações realizadas.

7.1. Diferenciais Competitivos

O Juris Syntax reúne atributos que o posicionam de forma competitiva no domínio das soluções jurídicas automatizadas. Destaca-se pela especialização em Direito Brasileiro, com ajuste de modelos e prompts às especificidades legislativas e práticas forenses nacionais. Sua arquitetura permite o processamento concorrente de múltiplos documentos extensos, mantendo latências aceitáveis em cenários de alta carga. A flexibilidade e customização são pilares centrais: o sistema admite parametrizações para diferentes órgãos, áreas do direito ou fluxos processuais. A solução também foi projetada com escalabilidade, podendo ser executada localmente, em servidores dedicados ou em provedores de nuvem, com contêineres Docker e orquestração. Por fim, adota privacidade por padrão, priorizando processamento local, armazenamento temporário controlado e comunicação segura via SDKs e variáveis de ambiente, aspectos cruciais para dados jurídicos sensíveis.

8. Resultados e Discussão

Os experimentos conduzidos com um conjunto diversificado de documentos jurídicos reais evidenciaram resultados promissores quanto à eficiência e eficácia do sistema desenvolvido. Observou-se uma redução significativa no tempo médio de análise documental, o que indica uma otimização considerável dos processos que tradicionalmente demandam esforço manual e prolongado. Além disso, o sistema demonstrou elevada precisão na identificação de documentos similares, fator crucial para a recuperação e comparação de jurisprudência e outras referências jurídicas relevantes.

1. Outro aspecto relevante é a consistência e qualidade dos fichamentos e resumos gerados automaticamente, que apresentaram alinhamento com padrões técnicos esperados, proporcionando suporte confiável para profissionais do direito. A adoção do modelo Llama 3.1 70B para a contextualização de textos extensos mostrou-se particularmente eficiente, superando abordagens tradicionais de sumarização jurídica baseadas em métodos convencionais, sobretudo na manutenção da coerência e riqueza semântica dos conteúdos processados. Esses resultados reforçam a aplicabilidade do *Juris Syntax* como ferramenta capaz de auxiliar a automação inteligente em ambientes jurídicos, agregando valor por meio da integração avançada de técnicas de Processamento de Linguagem

Natural (PLN) e Inteligência Artificial generativa.

Similaridade de Documentos O gráfico revela que os casos analisados apresentam um alto índice de similaridade, todos acima de 83%, indicando forte convergência temática e de fundamentos jurídicos. Isso sugere que os precedentes e decisões correlacionadas possuem elevado potencial de influência mútua, permitindo identificar padrões consistentes na aplicação das teses. Conforme demonstrado na Figura 5.

2. **Frequência de Fundamentos Legais** A análise demonstra que a Constituição Federal (CF/88) e o Código Tributário Nacional (CTN) são as normas mais recorrentes, reforçando seu papel central nas disputas tributárias. A presença frequente da Lei nº 9.718/1998 e da Lei nº 11.033/2004 evidencia que a discussão se concentra especialmente na interpretação do regime de PIS/COFINS. Conforme demonstrado na Figura 6.
3. **Tipos de Provas Apresentadas** Notas fiscais representam a maior parcela das provas, seguidas de documentos societários e registros contábeis (EFD). Isso indica que a comprovação material das operações é o elemento central para sustentar o direito pleiteado, sendo complementada por laudos técnicos e extratos financeiros. Conforme demonstrado na Figura 7.
4. **Principais Pedidos Formulados** Os pedidos mais frequentes são o creditamento de PIS/COFINS e a restituição ou compensação de valores pagos indevidamente. Esses resultados apontam para um padrão nas ações judiciais, cujo foco é recuperar valores e manter o direito à dedução tributária. Conforme demonstrado na Figura 8.
5. **Linha do Tempo de Normas Aplicadas** A evolução normativa revela marcos regulatórios desde 1967 até 2022. Apesar da longa linha histórica, observa-se uma concentração de normas pós-2000, o que reflete o aumento de mudanças legislativas e a complexidade crescente do regime tributário aplicado ao setor de combustíveis e outras atividades econômicas. Conforme demonstrado na Figura 9.

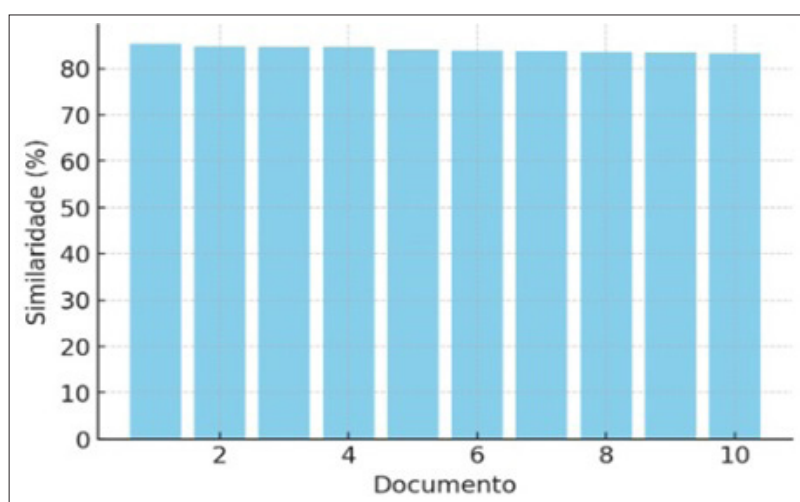


Figura 5 – Similaridade de documentos

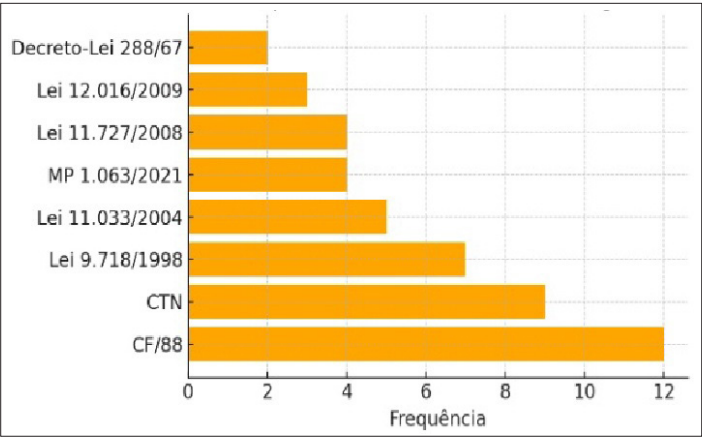


Figura 6 – Fundamentos Legais

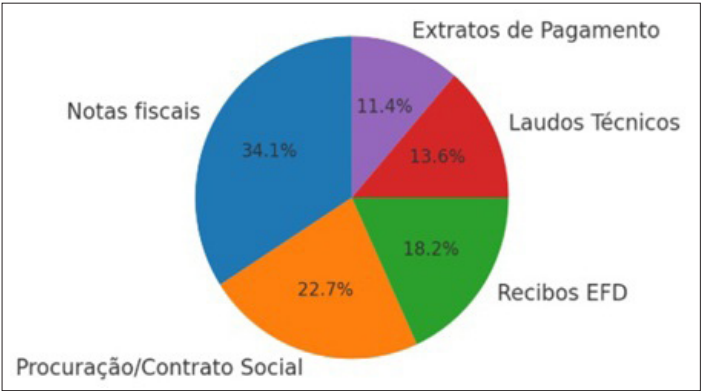


Figura 7 – Provas apresentadas

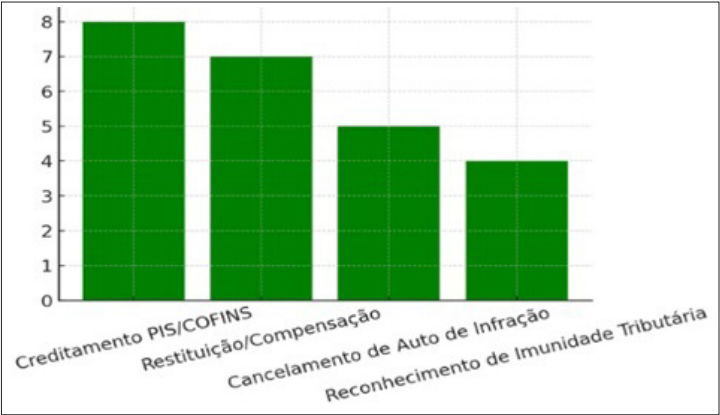


Figura 8 – Pedidos Formulados

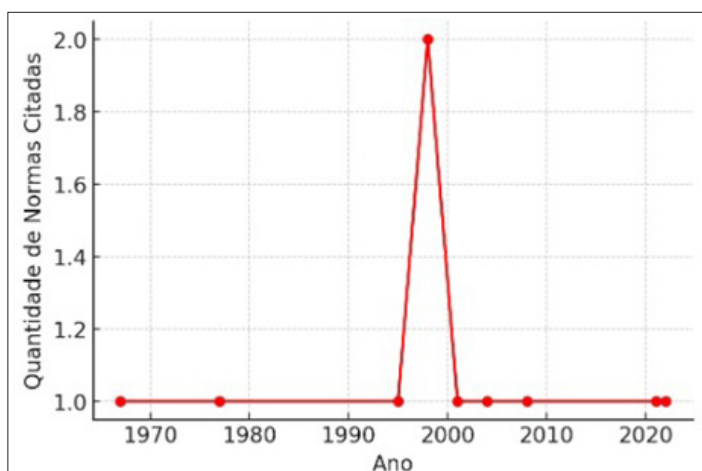


Figura 9 – Normas aplicadas

9. Conclusão e Trabalhos Futuros

O sistema Juris Syntax representa um avanço expressivo na automação e no aprimoramento de processos jurídicos, ao incorporar tecnologias de ponta em Processamento de Linguagem Natural (PLN) e Inteligência Artificial (IA) generativa. A proposta evidenciou viabilidade técnica e impacto positivo na eficiência operacional, demonstrando elevado potencial para transformar a prática jurídica cotidiana, reduzir a carga de trabalho manual e aumentar a precisão na análise documental. As perspectivas de evolução contemplam integração com bases de jurisprudência em tempo real, proporcionando respostas mais precisas e alinhadas aos entendimentos recentes. Pretende-se também o aprimoramento dos modelos de *embedding*, ampliando a capacidade de representação semântica e fortalecendo o suporte multilíngue, possibilitando aplicação em diferentes jurisdições e contextos internacionais.

Adicionalmente, o Juris Syntax poderá incorporar módulos de autenticação e controle de acesso; integração direta com sistemas processuais eletrônicos — como PJe e e-SAJ — promovendo interoperabilidade, extração automatizada de peças e otimização de fluxos processuais; dashboards analíticos para suporte à decisão; e processamento assíncrono para lidar com grandes volumes de documentos. Essas evoluções projetam o Juris Syntax como ferramenta robusta, versátil e estratégica na análise jurídica automatizada em escala nacional e internacional.

Agradecimentos

Este trabalho é apoiado pela Procuradoria Geral da Fazenda Nacional (nº PGFN 23106.148934/2019-67), Em parte pelo CNPq – Conselho Nacional de Pesquisas (Nº PQ-2 312180/2019-5 de Cibersegurança nº 465741/2014-2), em parte pela Advocacia Geral da União (nº AGU 697.935/2019), e em parte pela Fundação de Apoio à Pesquisa do Distrito Federal – FAPDF.

Referências

- Akter, M., Çano, E., Weber, E., Dobler, D., & Habernal, I. (2025). A comprehensive survey on legal summarization: Challenges and future directions. arXiv preprint.
- Anonymous. (2025). Analysing similarities between legal court documents using natural transformers (e.g., BERT, GPT-2, RoBERTa, LLaMA). PLOS ONE.
- Ariai, F., & Demartini, G. (2024). Natural language processing for the legal domain: A survey of tasks, datasets, models, and challenges. arXiv preprint.
- Barros, F. M. de C., Silva, C. D., Silva, I. R. de M., & Martins, V. S. (2024). Processamento de linguagem natural como ferramenta de suporte em documentos jurídicos: uma revisão sistemática. *Revista de Casos e Consultoria*, 15(1), e36701.
- Bertalan, V. G. F., & Ruiz, E. E. S. (2022). Using attention methods to predict judicial outcomes. arXiv preprint.
- Campos de Carvalho, T. P. F., Freitas, S. H. Z., Ribeiro, A. de S., et al. (2022). A inteligência artificial no poder judiciário brasileiro e a gestão de conflitos. *Meritum, Revista de Direito da Universidade FUMEC*.
- Carmo, F. A. de, Serejo, F., Jacob Junior, A. F., et al. (2023). Embeddings jurídico: Representações orientadas à linguagem jurídica brasileira (pp. 188–199).
- Conselho Nacional de Justiça. (2024). Justiça em números 2024. <https://www.cnj.jus.br/wp-content/uploads/conteudo/arquivo/2024/08/relatorio-justica-em-numeros-2024.pdf>
- Ferreira, G. M. (2022). Inteligência artificial como auxiliar do poder judiciário: A experiência do sistema Victor no âmbito do STF [Monografia de graduação, Universidade Federal do Ceará].
- Figueiredo, G. S. (2022). Projeto Athos: Um estudo de caso sobre a inserção do Superior Tribunal de Justiça na era da inteligência artificial [Dissertação de mestrado profissional, Universidade de Brasília].
- Garcia, E., Silva, N., Siqueira, F., et al. (2024). Robertalexpt: A legal RoBERTa model pre-trained with deduplication for Portuguese. In *Proceedings of PROPOR (ACL Anthology) – Legal NLP Workshop*. <https://aclanthology.org/2024.propor-1.38.pdf>
- Goyal, G., et al. (2022). Prompt-based summarization using GPT-3. arXiv preprint.
- Huang, H., & Chang, C. (2025). Advancing prompt-based language models in the legal domain: Adaptive methods. *Journal of Legal AI*.
- Melo, J. S. a. S., Nascente, V. F., & dos Santos, L. E. (2023). Discussão sobre a viabilidade técnica e jurídica para a aplicação de processamento de linguagem natural em decisões vinculantes em processos judiciais.

- Pereira, J., Assumpção, A., Trecenti, J., et al. (2024). Inacia: Integrating large language models in Brazilian audit courts: Opportunities and challenges. arXiv preprint.
- Polo, F. M., Mendonça, G. C. F., Parreira, K. C. J., et al. (2021). LegalNLP – Natural language processing methods for the Brazilian legal language. In Anais do XVIII Encontro Nacional de Inteligência Artificial e Computacional (SBC).
- Rahman, M. M., et al. (2025). Natural language processing in legal document analysis software: A systematic review. *International Journal of Innovative Research in Social Sciences (IJIRSS)*.
- R., et al. (2022). Legal-BERTimbau: BERT for the Brazilian legal domain (model card / Hugging Face). <https://huggingface.co/rufimelo/Legal-BERTimbau-base>
- Silva, D. de J., Oliveira, D. de, & Paes, A. (2025). Evaluating text representations for unsupervised legal semantic textual similarity in Brazilian Portuguese. *Discover Data*, 3, 23.
- Silva, P. A. de L. A. S. (2025). Performance analysis of LLMs for abstractive summarization in legislative documents.
- Silva, A. da. (2024). A legal framework for natural language processing model training in Portugal. arXiv preprint.

Sistema de apoyo al Diseño Curricular: Extensión de Meta4CBC para la evaluación por competencias

Ninoska Díaz¹, Ania Cravero²

n.diaz10@ufromail.cl; ania.cravero@ufrontera.cl

¹ Magister en Ingeniería Informática, Depto. de Cs. de la Computación e Informática, Universidad de La Frontera, Temuco, Chile.

² Depto. de Cs. de la Computación e Informática, CEMCC, Universidad de La Frontera, Temuco, Chile.

Pages: 28-39

Resumen: La educación superior enfrenta el reto de asegurar coherencia entre objetivos formativos, currículo y evaluación. Aunque el currículo basado en competencias se ha consolidado como enfoque dominante, persisten vacíos en la representación formal de los sistemas de evaluación, especialmente a nivel meso. Este trabajo tiene como objetivo proponer un sistema de apoyo para el diseño curricular, mediante la extensión del metamodelo Meta4CBC. La propuesta incorpora un paquete de “Mecanismos de Evaluación” –competencias, indicadores de desempeño, reglas de agregación, referentes, ciclos de revisión y roles de gobernanza– que no estaban contemplados en el modelo original. Además, se integra un nivel supra que conecta estos elementos con marcos internacionales, fortaleciendo la trazabilidad y legitimidad del currículo.

Palabras-clave: Sistema, Evaluación, Diseño Curricular, Competencias.

Curriculum Design Support System: Meta4CBC Extension for Competency-Based Assessment

Abstract: Higher education faces the challenge of ensuring coherence between learning objectives, curriculum, and assessment processes. Although competency-based curriculum has become a dominant approach, significant gaps remain in the formal representation of evaluation systems, particularly at the meso level. The objective of this work is to propose a support system for curriculum design by extending the Meta4CBC metamodel. The proposal incorporates an “Evaluation Mechanisms” package—competencies, performance indicators, aggregation rules, benchmarks, review cycles, and governance roles—that were not considered in the original model. In addition, a supra-level is integrated to link these elements to international frameworks, strengthening curriculum traceability and legitimacy.

Keywords: Meta4CBC, Evaluation, Curriculum Design, Competencies.

1. Introducción

La Educación Superior (ES) enfrenta el desafío de asegurar la coherencia entre intenciones formativas, planificación curricular y procesos de evaluación. El currículo basado en competencias se ha consolidado como un marco de referencia ampliamente utilizado, al integrar conocimientos, habilidades y actitudes orientadas a contextos profesionales (Fahimirad et al., 2019; Bakhshinategh et al., 2017). No obstante, persiste el problema del desalineamiento curricular entre competencias, Resultados de Aprendizaje (RA), estrategias de enseñanza y mecanismos de evaluación, lo cual afecta la trazabilidad y dificulta el aseguramiento de la calidad educativa (Dobb, 2021; Muñoz & Bruna, 2021; Cravero et al., 2024).

Para enfrentar este reto, han sido promovidos diversos estándares internacionales como referentes normativos que guían el diseño y la evaluación curricular en múltiples niveles. Marcos como ABET, CDIO, Tuning y el EQF establecen taxonomías, descriptores de resultados y criterios de calidad que facilitan el alineamiento desde el nivel supra (normativo) hasta el nivel micro (curso) (Bay, 2016; EQF, 2017; Wagenaar, 2014). Estos marcos refuerzan la comparabilidad y la transparencia del currículo, y constituyen un soporte fundamental para procesos de acreditación y movilidad académica.

Sin embargo, la literatura muestra que aún no existen sistemas informáticos capaces de integrar todos los componentes curriculares en los distintos niveles. La alineación suele resolverse de manera fragmentada, especialmente en el nivel micro, sin garantizar que las competencias de egreso definidas en el nivel meso sean evaluadas y monitoreadas de forma sistemática (Mistamiruddin et al., 2024; Dubé et al., 2023; Dobb, 2021). Esta ausencia de soporte tecnológico limita la trazabilidad y la generación de evidencia para procesos de mejora continua.

Para superar esta brecha, se requieren modelos formales que identifiquen los componentes curriculares y expliciten sus relaciones, de manera que los equipos de desarrollo curricular –usualmente multidisciplinarios y con distintos niveles de experiencia pedagógica y tecnológica– compartan un lenguaje común que facilite la coordinación (Thian et al., 2018). Estos modelos permiten estructurar el diseño curricular, además de sentar las bases para desarrollar herramientas digitales que automaticen la validación del alineamiento.

En este sentido, los sistemas informáticos de apoyo al diseño curricular son importantes para garantizar la consistencia entre competencias, RA y mecanismos de evaluación, así como para proveer datos accionables que fortalezcan la gobernanza académica (Shaltry, 2020; Dera, 2021). No obstante, la oferta de soluciones de este tipo sigue siendo escasa y fragmentaria, lo cual constituye un desafío para la consolidación del CBC.

Una de las propuestas más relevantes es el metamodelo Meta4CBC, que ofrece una base conceptual para representar el currículo a través de la *Model Driven Architecture* (MDA), utilizando UML para modelar la estructura y SBVR para definir reglas semánticas (Cravero et al., 2024). Este metamodelo ha demostrado su potencial para alinear competencias, RA y objetivos de programa; sin embargo, carece de un bloque específico para la representación de los mecanismos de evaluación, particularmente en el nivel meso, donde se definen las competencias de egreso.

El presente trabajo amplía Meta4CBC mediante la incorporación del paquete “Mecanismos de Evaluación”, que vincula competencias con indicadores de desempeño, reglas de agregación, referentes, ciclos de revisión y roles de gobernanza. Esta extensión habilita el desarrollo de sistemas informáticos que apoyen el diseño curricular y aseguren el alineamiento entre lo declarado, lo enseñado y lo evaluado. Su relevancia es especialmente alta en programas STEM, donde la medición de competencias es importante para garantizar la pertinencia con estándares industriales y la empleabilidad de los egresados (Kirwan et al., 2023; Ali et al., 2023).

En consecuencia, este trabajo se orienta por la siguiente pregunta de investigación: ¿Cómo extender el metamodelo Meta4CBC para desarrollar sistemas informáticos que integren los mecanismos de evaluación en el nivel meso y fortalezcan el alineamiento curricular en programas de ES?. El objetivo es proponer un sistema informático de apoyo al diseño curricular, basado en la extensión de Meta4CBC, que permita operacionalizar la evaluación por competencias de manera formal, trazable y validada.

2. Estado del Arte

2.1. Alineamiento curricular en ES

El alineamiento curricular se refiere a la coherencia entre los distintos componentes del currículo: competencias, RA, estrategias de enseñanza y mecanismos de evaluación. Un currículo bien alineado asegura que lo que se enseña, lo que se espera que los estudiantes aprendan y la manera en que se evalúa dicho aprendizaje estén en correspondencia directa. Como señala Biggs (1996), esta coherencia es condición fundamental para favorecer aprendizajes significativos y profundos.

En la ES, esta coherencia debe sostenerse de manera articulada en los distintos niveles curriculares (Álvarez et al., 2023). El nivel supra corresponde a los marcos normativos externos e institucionales que orientan la formación; el nivel macro se centra en las políticas educativas y modelos institucionales; el nivel meso se relaciona con los programas de estudio, donde se definen competencias y perfiles de egreso; y el nivel micro se vincula a las asignaturas, con RA específicos, contenidos y métodos de evaluación (Cravero et al., 2024; Squires, 2009). La falta de articulación entre estos niveles impide que los propósitos formativos se traduzcan en prácticas pedagógicas consistentes, debilitando la trazabilidad del currículo.

Este problema se hace más evidente en los programas STEM, donde los conocimientos son altamente especializados y su evaluación exige un alto grado de precisión metodológica (Wijngaards-de Meij & Merx, 2018). Uno de los principales desafíos del currículo basado en competencias es precisamente el diseño de sistemas de evaluación que midan de forma válida y confiable las competencias declaradas. En muchos casos, los RA se formulan de manera ambigua o sin una clara vinculación con las competencias, lo que genera dificultades para medir el logro estudiantil. A ello se suma que los instrumentos de evaluación suelen no corresponder al nivel cognitivo requerido por los RA, produciendo un desfase entre lo declarado y lo efectivamente evaluado (Shallal, 2018; Bay, 2016).

Estas brechas resultan particularmente críticas en el nivel meso –donde se establecen las competencias de egreso y los perfiles profesionales– y en el nivel micro, donde se definen los RA y los métodos de enseñanza-aprendizaje. Cuando no existe un alineamiento entre ambos niveles, se dificulta el seguimiento del progreso del estudiante y la implementación de procesos de mejora continua en el currículo (Cravero et al., 2024).

2.2. Educación basada en competencias

En la ES, este enfoque busca que los estudiantes desarrollen no sólo conocimientos, sino también habilidades y actitudes aplicables a contextos reales y multidisciplinarios. Particularmente en carreras STEM, permite conectar la formación académica con los estándares profesionales. No obstante, su implementación efectiva requiere modelos y herramientas que garanticen la coherencia entre competencias esperadas, enseñanza impartida y evaluación aplicada (Ali et al., 2023).

2.3 Metamodelo Meta4CBC

El Meta4CBC (Metamodel for Competency-Based Curriculum) es un metamodelo diseñado para ofrecer una representación formal del currículo basado en competencias. Fue desarrollado tomando como referencia el Business Motivation Model (BMM) y emplea el lenguaje SBVR (Semantics of Business Vocabulary and Rules) para definir los conceptos curriculares y las reglas que gobiernan sus relaciones (Cravero et al., 2024). Su enfoque se inscribe en el paradigma de la Model Driven Architecture (MDA), lo que permite la definición conceptual en un nivel abstracto y su posterior implementación en sistemas informáticos para apoyar el diseño y la gestión curricular.

Meta4CBC organiza el currículo en niveles jerárquicos: Nivel Macro: define políticas institucionales, modelos educativos y orientaciones curriculares. Nivel Meso: corresponde a los programas de estudio, donde se establecen perfiles de egreso, competencias de programa y objetivos formativos. Nivel Micro: abarca asignaturas y módulos, con resultados de aprendizaje, contenidos y metodologías de enseñanza-evaluación.

Cada nivel se estructura en bloques principales que agrupan los componentes curriculares clave. Por ejemplo, en el nivel meso se incluyen el Perfil de Egreso, las Competencias de Programa y los Objetivos de Aprendizaje, los cuales se vinculan con los RA definidos en el nivel micro. Estas relaciones se modelan formalmente con UML, que permite representar clases, atributos y asociaciones, facilitando la construcción de sistemas informáticos para el análisis y seguimiento del currículo.

Una de las fortalezas de Meta4CBC es su formalización técnica: los componentes curriculares se definen como clases y atributos, además se regulan mediante reglas SBVR. Estas reglas expresan de manera declarativa la lógica que asegura la coherencia interna del currículo. Ejemplos de reglas son: *“toda competencia debe derivar en uno o más RA”* o *“todo RA debe estar alineado con al menos un objetivo del programa”*. Este enfoque garantiza consistencia semántica, posibilita la verificación automática y habilita la trazabilidad entre lo declarado, lo enseñado y lo evaluado.

No obstante, el metamodelo original presenta limitaciones. En particular, carece de un bloque formal para representar los mecanismos de evaluación, lo cual es crítico en

el nivel meso, donde se monitorea el logro de las competencias de egreso. Asimismo, no contemplaba explícitamente el nivel supra como espacio para articular el currículo con marcos normativos externos. Estas carencias reducen su capacidad para apoyar procesos de aseguramiento de la calidad y para demostrar alineamiento con estándares internacionales.

3. Metodología

Este trabajo adopta un enfoque de diseño basado en modelado formal, orientado a representar de forma estructurada, el subsistema de evaluación por competencias en el nivel meso curricular, articulado con los niveles micro, macro y supra. La propuesta se enmarca en los principios del enfoque Model-Driven Architecture (MDA) (Object Management Group [OMG], 2014), empleando lenguajes estandarizados como UML para la definición estructural del modelo y SBVR para la representación de sus reglas lógicas y semánticas.

Esta aproximación permite que el modelo sea interpretable tanto por humanos como por sistemas digitales, facilitando su interoperabilidad con plataformas de gestión académica, su validación lógica y su integración con herramientas de visualización de datos institucionales. Además, habilita la trazabilidad curricular y la documentación formal del alineamiento entre lo declarado, lo enseñado y lo evaluado (Cravero et al., 2024).

3.1. Fases del desarrollo

a) Revisión de modelos conceptuales y estructuras curriculares

Se realizó una revisión crítica de literatura científica y modelos estructurales relacionados con la planificación y evaluación por competencias. Se analizaron enfoques tradicionales como ADDIE (Molenda, 2003) y técnicas como el mapeo curricular (Harden, 2001), así como propuestas de estandarización como XCRI-CAP (Wilkerson & Wilson, 2009), que si bien aportan a la representación de programas, carecen de una semántica formal robusta.

Paralelamente, se estudió en profundidad la arquitectura del metamodelo Meta4CBC (Cravero et al., 2024), identificando un vacío en la representación del bloque de evaluación en el nivel meso, así como la ausencia de un mecanismo explícito para vincular los componentes curriculares con referentes normativos institucionales o internacionales (nivel supra). Este diagnóstico motivó la extensión del modelo mediante nuevos paquetes estructurales y relaciones formales que permiten integrar estándares como CDIO, Tuning o EQF como parte del diseño curricular.

b) Modelado formal en UML y SBVR

A partir del análisis de literatura científica sobre evaluación por competencias, se identificó la ausencia de un bloque específico en Meta4CBC que representara los procesos evaluativos a nivel meso, es decir, en el nivel de los programas de estudio. Para subsanar esta brecha, se propuso la incorporación de un nuevo paquete denominado **“Mecanismos de Evaluación”**, integrado al nivel meso.

El proceso de modelado formal siguió cuatro etapas:

1. Selección de componentes curriculares: a partir de revisiones de modelos conceptuales y marcos de referencia (p. ej. COPA, ABET, Tuning, EQF), se definieron los elementos esenciales que un sistema de evaluación por competencias debe incluir: *indicadores de desempeño, reglas de agregación, referentes, ciclos de revisión y roles de gobernanza*.
2. Ubicación y relaciones en el nivel meso: los nuevos componentes se integraron al paquete meso, de modo que cada CompetenciaPrograma quedó vinculada explícitamente a uno o más Indicadores de Desempeño, cerrando la brecha entre competencias declaradas y su evaluación formal.
3. Definición semántica con SBVR: se formularon reglas de negocio y consistencia que aseguran la validez del modelo. Por ejemplo:
 - *Regla 1 (Competencia–Indicador)*: “Toda CompetenciaPrograma debe estar asociada a al menos un IndicadorDesempeño”.
 - *Regla 2 (Indicador–Referente)*: “Todo IndicadorDesempeño debe ser comparado con al menos un Referente definido institucional o externamente”.
 - *Regla 3 (Ciclo–Gobernanza)*: “Cada CicloRevision debe ser ejecutado por un RolGobernanza claramente establecido en el programa”.
4. Validación del modelo: la propuesta fue sometida a una revisión conceptual con expertos en currículo y gestión académica, quienes confirmaron la pertinencia de los componentes y la utilidad de las reglas para asegurar trazabilidad y coherencia en los procesos de evaluación.

c) Validación conceptual en entorno real

Para evaluar la viabilidad del modelo, se realizó una instanciación en la carrera de Ingeniería Informática de la Universidad de La Frontera. Se seleccionó una competencia de egreso representativa, y se construyó un sistema evaluativo completo, incluyendo: “Indicadores de desempeño concretos”, “Reglas de cálculo”, “Metas institucionales como referentes”, “Fechas y frecuencias de revisión”, “Responsables de evaluación y toma de decisiones”. Este ejercicio demostró la aplicabilidad práctica del modelo y su utilidad como herramienta operativa en entornos reales.

4. Resultados

4.1 Formalización UML y reglas SBVR

La Figura 1 muestra el modelo curricular extendido codificado en PlantUML, utilizando una sintaxis formal para representar clases, atributos y relaciones. Este diagrama organiza el currículo en cuatro niveles (supra, macro, meso y micro) e incorpora un nuevo paquete denominado “**Mecanismos de Evaluación**” (ver Figura 2 con más detalles), que constituye el principal aporte de la propuesta.

Dentro del nivel meso, el paquete Mecanismos de Evaluación articula los componentes necesarios para operacionalizar la evaluación de competencias de manera formal y trazable. En este paquete, la clase central es **ModeloEvaluacion**, que actúa como contenedor lógico del sistema evaluativo del programa.

Los componentes principales y sus relaciones se describen a continuación:

- **IndicadorDesempeño**: define qué aspecto específico de la competencia se mide y cuál es la evidencia observable asociada. Cada **CompetenciaPrograma** debe estar vinculada a uno o más indicadores, asegurando que toda competencia declarada tenga mecanismos formales de medición.
- **ReglaAgregacion**: establece cómo se integran los indicadores de desempeño, ya sea mediante ponderaciones, umbrales o fórmulas de cálculo. Se relaciona directamente con **IndicadorDesempeño**, definiendo la lógica que transforma evidencias en resultados globales.
- **Referente**: representa los valores de referencia o estándares comparativos (internos o externos) contra los cuales se juzga el logro de un indicador. Todo **IndicadorDesempeño** debe estar asociado a un referente, lo que asegura comparabilidad y coherencia.
- **CicloRevision**: determina la periodicidad con que los indicadores son analizados (semestral, anual, por cohorte). Se relaciona con **IndicadorDesempeño** y con **RolGobernanza**, estableciendo cuándo y cómo se revisan los resultados.
- **RolGobernanza**: identifica a los actores institucionales responsables de ejecutar y supervisar la evaluación (comités de calidad, coordinadores, evaluadores externos). Su relación con **CicloRevision** asegura que toda revisión esté asociada a responsables claramente definidos.

Las relaciones entre estos componentes fueron codificadas mediante conectores UML de asociación y composición, y reguladas a través de reglas **SBVR**, como:

- *“Toda CompetenciaPrograma debe estar vinculada a al menos un IndicadorDesempeño”.*
- *“Todo IndicadorDesempeño debe ser evaluado en referencia a un estándar comparativo”.*
- *“Cada CicloRevision debe estar a cargo de un RolGobernanza definido institucionalmente”.*

Finalmente, la Figura 1 muestra la integración explícita del **nivel supra**, a través de la clase **Marco**, que representa los marcos normativos y estándares externos (p. ej. ABET, CDIO, Tuning, EQF). Este elemento se vincula con:

- el **ModeloEducativo**, que se diseña con base en dichos marcos;
- las **CompetenciasPrograma**, que se formulan en coherencia con referentes normativos;
- el **ModeloEvaluacion**, que se fundamenta en principios y criterios reconocidos externamente.

De esta manera, el modelo asegura que tanto los fines formativos como los mecanismos evaluativos respondan a principios pedagógicos reconocidos y verificables, fortaleciendo la legitimidad, la trazabilidad y la comparabilidad del currículo en procesos de acreditación, aseguramiento de la calidad y movilidad académica.

4.2. Ejemplo aplicado: Ingeniería Informática (UFRO)

Para validar conceptualmente la aplicabilidad del modelo propuesto, se realizó una instancia de su implementación en el contexto de la carrera de Ingeniería Informática de la Universidad de La Frontera (UFRO). El objetivo fue modelar un sistema formal de evaluación para una competencia de egreso representativa del perfil profesional del programa, utilizando los componentes del paquete “Mecanismos de Evaluación”.

a) Competencia seleccionada

“Diseñar soluciones de software eficientes, seguras y sostenibles, aplicando principios de ingeniería de software, estructuras de datos y estándares de calidad.”

Esta competencia se clasifica como disciplinar específica, de nivel avanzado, y está asociada a las asignaturas del eje formativo “Desarrollo de Software Profesional”.

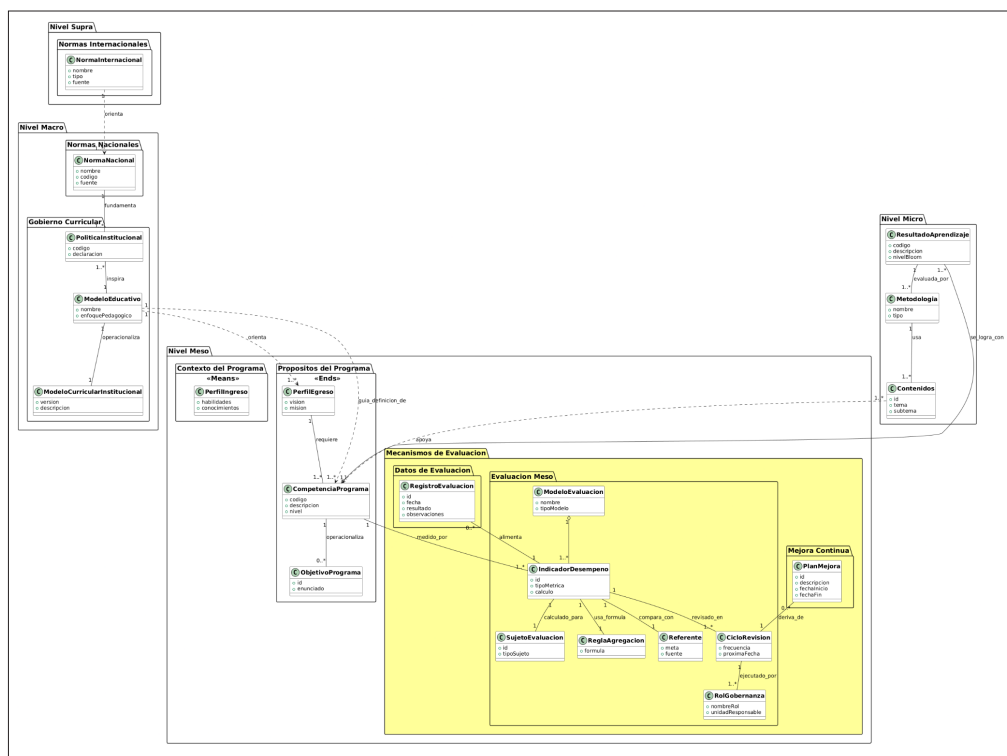


Figura 1 – Extensión del metamodelo Meta4CBC en PlantUML.

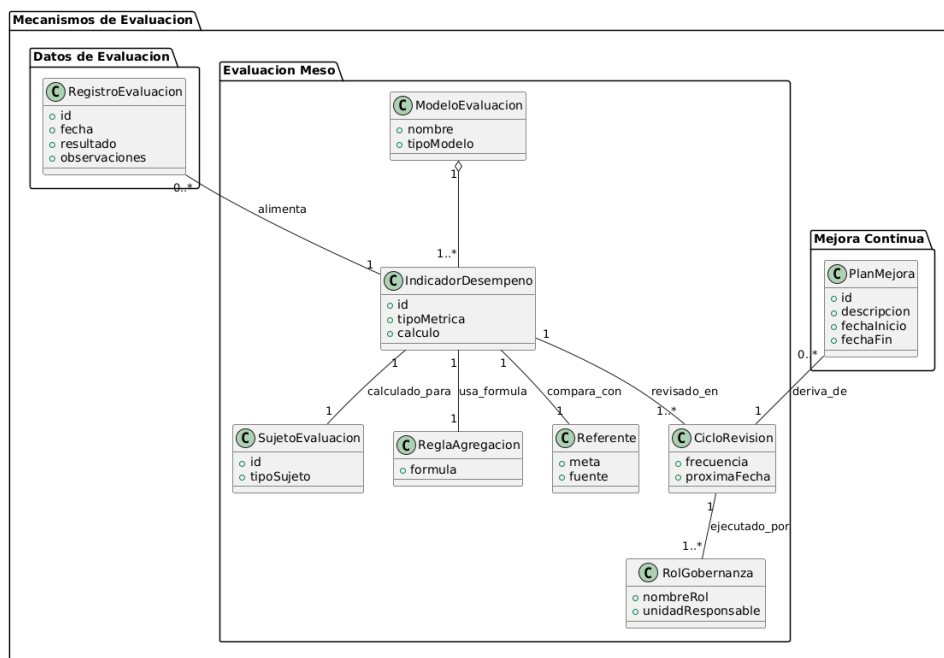


Figura 2 – Mecanismo de Evaluación para Meta4CBC en PlantUML.

b) Componentes del Modelo Evaluación

A partir de esta competencia, se construyó un ModeloEvaluacion que agrupa de manera coherente los elementos siguientes:

- **Indicadores de Desempeño:** **ID1:** Porcentaje de estudiantes que implementan algoritmos con eficiencia computacional conforme al perfil esperado (medido mediante rúbrica de complejidad). **ID2:** Nivel de cumplimiento de estándares OWASP en prácticas de codificación segura (evaluado con lista de cotejo). **ID3:** Calidad de documentación técnica en entregas finales (medida mediante rúbrica específica).
- **Reglas de Agregación:** **ID1:** Media ponderada 70% evaluación individual + 30% revisión cruzada. **ID2:** Cumplimiento binario (Sí/No) de al menos 80% de los ítems críticos. **ID3:** Evaluación con escala 1 a 4, promedio superior a 3 requerido.
- **Referentes:** **ID1:** 75% de estudiantes deben alcanzar al menos nivel 3 **ID2:** 90% de cumplimiento mínimo. **ID3:** Media ≥ 3.0 como estándar de calidad documental.
- **Ciclo de Revisión:** Periodicidad: Semestral (julio y diciembre). Datos comparativos: Últimos tres semestres. Revisión conjunta por: Coordinador de eje, docente responsable, asesor curricular.

- **Rol de Gobernanza: Comité de Calidad del Programa**, compuesto por el jefe de carrera, académicos del área de software y un representante del centro de innovación docente.

c) Operación práctica del modelo

Durante el segundo semestre de 2024, el análisis del indicador ID1 reveló un cumplimiento promedio de 68%, por debajo del referente establecido. Esta desviación activó un proceso de mejora que incluyó: Revisión del material de apoyo en algoritmos, Aumento del tiempo dedicado a prácticas guiadas, Reformulación parcial de los criterios de evaluación para la rúbrica de eficiencia.

El sistema, al estar formalmente modelado, permitió documentar el ciclo completo: Competencia → Indicadores → Evidencia → Análisis → Decisión, facilitando su trazabilidad en el sistema de aseguramiento de la calidad y el rediseño curricular.

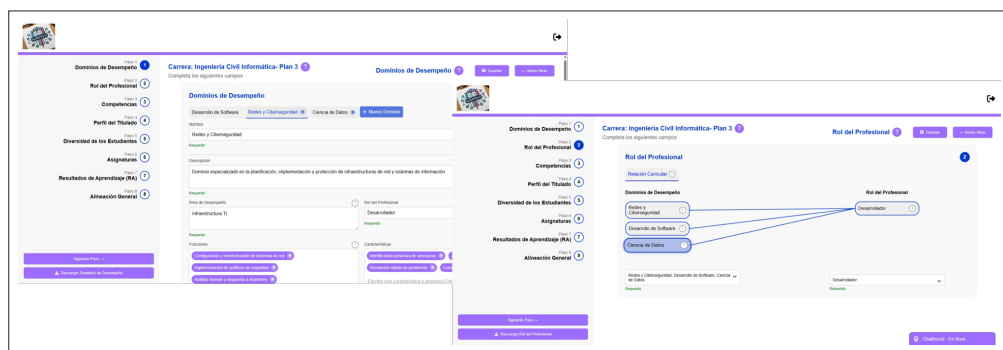


Figura 3 – Sistema de apoyo para el diseño curricular.

Como resultado del caso, en el segundo semestre de 2024, el Indicador Desempeño ID1 no alcanzó su Referente (68% vs. esperado 75%). El Comité de Calidad del Programa activó acciones de mejora: actualización de rúbrica, refuerzo de contenidos y modificación del tiempo de práctica guiada. Esta acción se registró formalmente como parte del proceso de aseguramiento de calidad del programa. La Figura 3 presenta un ejemplo del sistema informático desarrollado a partir de la extensión de Meta4CBC.

5. Conclusiones

Este trabajo presenta una extensión formal y funcional del metamodelo Meta4CBC mediante la incorporación del paquete “Mecanismos de Evaluación” en el nivel meso curricular. Esta ampliación permite representar explícitamente las relaciones entre competencias de programa y los elementos clave de su evaluación: indicadores de desempeño, reglas de agregación, referentes, ciclos de revisión y roles de gobernanza. Con ello, se resuelve un vacío estructural previamente identificado en Meta4CBC y se fortalece la capacidad del modelo para garantizar coherencia, trazabilidad y mejora continua dentro de los programas de formación por competencias.

La propuesta fue construida utilizando lenguaje de modelado UML y reglas SBVR, lo que garantiza precisión semántica, reutilización formal y posibilidad de integración en sistemas tecnológicos de diseño y gestión curricular. A través de su validación en el contexto de la carrera de Ingeniería Informática, se demostró que el modelo es técnicamente viable, conceptualmente coherente y operativamente útil. La generación del sistema informático evidenció la utilidad práctica del modelo como soporte para la gestión basada en evidencia.

Un aporte significativo de esta propuesta es que permite articular componentes curriculares en todos los niveles del sistema educativo: desde los referentes normativos del nivel supra (representados por la clase Marco), pasando por los elementos del modelo educativo y competencias (macro y meso), hasta los RA y actividades evaluativas (micro). Esta cobertura multinivel favorece una visión sistémica e integrada del currículo, asegurando que cada decisión de diseño y evaluación esté alineada con marcos institucionales y estándares internacionales. El sistema informático desarrollado, mejora la trazabilidad vertical, fortaleciendo la gobernanza curricular y la capacidad institucional para demostrar calidad.

Referencias

- Álvarez, D., Melillán, A., Cravero, A., & Sepúlveda, S. (2023). Propuesta de Modelo para el Alineamiento entre el Diseño Curricular y las TI. *Revista Ibérica de Sistemas e Tecnologías de Informação*, (E59), 531-543.
- Ali, M., Hafeez, M., & Tariq, R. (2023). Competency-based curriculum design in engineering education: Challenges and solutions. *International Journal of Engineering Pedagogy*, 13(1), 45–58. <https://doi.org/10.3991/ijep.v13i1.34213>
- Bay, E. (2016). Ensuring curriculum quality through international frameworks: The role of CDIO and Tuning. *Journal of Engineering Education*, 105(3), 478–495. <https://doi.org/10.1002/jee.2016.1053>
- Cravero, A., Álvarez, D., Sepúlveda, S., & Valdivieso, M. A. (2024). Meta4CBC: A model-driven metamodel for curriculum design based on competencies. *Applied Sciences*, 14(22), 10110. <https://doi.org/10.3390/app142210110>
- EQF (European Qualifications Framework). (2017). *Recommendation of the European Parliament and of the Council on the establishment of a European Qualifications Framework for lifelong learning*. Official Journal of the European Union. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32017H0615%2801%29>
- Gysling, J., Muñoz, J., & Ávalos, B. (2021). Desafíos del alineamiento curricular en la educación basada en competencias. *Revista Calidad en la Educación Superior*, 10(2), 37–54. <https://doi.org/10.22370/caledus.2021.10.2.2834>
- López Rupérez, F. (2022). El enfoque del currículo por competencias: Un análisis de la LOMLOE. *Revista Española de Pedagogía*, 80(281), 5–27. <https://doi.org/10.22550/REP80-1-2022-01>

- Molenda, M. (2003). In search of the elusive ADDIE model. *Performance Improvement*, 42(5), 34–37. <https://doi.org/10.1002/pfi.4930420508>
- Muñoz-Artacho, J., Gysling, J., & Rivera, M. (2020). Marcos, modelos y herramientas para medir competencias en la educación superior. *Estudios Pedagógicos*, 46(1), 111–130. <https://doi.org/10.4067/S0718-07052020000100111>
- Muñoz, J., & Bruna, D. (2021). Alineamiento entre competencias y resultados de aprendizaje: Desafíos para su evaluación efectiva. *Revista Docencia Universitaria*, 19(3), 88–106. <https://doi.org/10.4995/redu.2021.15359>
- Object Management Group (OMG). (2014). *MDA guide version 2.0*. OMG. <https://www.omg.org/cgi-bin/doc?orm/14-06-01.pdf>
- Squires, D. A. (2009). *Curriculum alignment: Research-based strategies for increasing student achievement*. Corwin Press.
- Shallal, M. (2018). Linking learning outcomes with assessment strategies in higher education. *Journal of Educational Measurement and Evaluation*, 9(1), 12–21. <https://doi.org/10.1007/s11092-018-09274-1>
- Wilkerson, D., & Wilson, J. (2009). XCRI-CAP: A UK standard for course marketing data. *Journal of Computing and Information Technology*, 17(2), 111–118. <https://doi.org/10.2498/cit.1001306>
- Wijngaards-de Meij, L., & Merx, S. (2018). Improving curriculum alignment and achieving learning goals by making the curriculum visible. *International Journal for Academic Development*, 23(3), 219–231. <https://doi.org/10.1080/1360144X.2018.1462187>

Redes neuronales convolucionales (YOLO) para la detección de megafauna marina y basura marina flotante a partir de imágenes aéreas

Maria V DeLaHoz.^{1,3}, Toni Monleón-Getino^{1,3,4}, Albert Martín^{1,3}, Odei Garcia- Garin^{4,5}, Morgana Vighi^{2,6}, Marina Costa⁶, Valentina De Santis⁶, Caterina Lanfredi⁶, Capt. Daniele Giannelli⁷, Sabina Airoidi⁶.

mvdelahoz@hotmail.com; amonleong@ub.edu; odei.garcia@ub.edu;
morgana.vighi@gmail.com; marinza.costa@gmail.com; vale.desa@tiscali.it;
lanfredicaterina@gmail.com; sabina.airoidi@iol.it

¹ Departamento de Genética, Microbiología y Estadística, Universidad de Barcelona, Diagonal 643 08028 Barcelona, España

² Departamento de Biología Evolutiva, Ecología y Ciencias Ambientales, e Institut de Recerca de la Biodiversitat (IRBio), Universidad de Barcelona, Diagonal 643 08028 Barcelona, España

³ BIST3: Grupo de investigación en Estadística, Ciencia de Datos y Bioinformática. Universidad de Barcelona, Diagonal 643 08028 Barcelona, España

⁴ Instituto de Ecología Acuática (IEA), Universitat de Girona; Gerona (España)

⁵ Instituto de Investigación de la Biodiversidad (IRBio) y Departamento de Biología Evolutiva, Ecología y Ciencias Ambientales, Universitat de Barcelona, Diagonal 643 08028; Barcelona (España)

⁶ Tethys Research Institute, Viale G.B. Gadio 2, 20121 Milano, Italia

⁷ Italian Coast Guard Headquarters. Roma, Italia

Pages: 40-54

Resumen: Este estudio evalúa el potencial de la red neuronal convolucional YOLOv8 para detección y clasificación de megafauna y basura flotante en el mar a partir de imágenes aéreas. Se realizaron distintos ensayos con imágenes de cetáceos, obteniendo un rendimiento satisfactorio en casi todas las especies evaluadas, con probabilidades de detección correcta, valores de mAP-50 y valores de F1-score alrededor de 0.9. El mejor compromiso entre *overfitting* y función de pérdida a lo largo de los ensayos, se observó al ejecutar 500 iteraciones, más allá de las cuales aumentaba fuertemente dicho sobreajuste. A modo comparativo, un ensayo con distintos ítems de basura marina mostró proporciones relativamente altas de detecciones incorrectas y métricas muy fluctuantes durante la validación, ilustrando signos claros de *underfitting* y sugiriendo posibilidad de mejoras gracias a la flexibilidad del modelo. Se proponen los posibles alcances de YOLOv8 como herramienta de apoyo a actividades de monitoreo para conservación marina.

Palabras-clave: Redes neuronales convolucionales; YOLO; Megafauna marina; Basura marina flotante.

Convolutional Neural Networks (YOLO) for detection of marine organisms and floating marine litter from aerial images

Abstract: This study evaluates the potential of the convolutional neural network YOLOv8 for detection and classification of megafauna and floating marine litter, from aerial images. Several runs were conducted with images of various cetacean species, achieving satisfactory performance in most of them, with probabilities of correct detection, mAP-50 values, and F1-scores around 0.9. The best trade-off between overfitting and loss function was observed when running the algorithm for 500 epochs, beyond which a sharp increase of overfitting occurred. A comparative analysis with different items of marine litter, showed high proportions of incorrect detections and highly fluctuating metrics during validation, illustrating clear signs of underfitting. Varying contrasts between object colorations and the background, may be a factor influencing the CNN's efficiency between classes. However, the flexibility of the model leaves possibilities of improvements. Further application of the use of YOLOv8 is proposed as a tool to support monitoring activities for marine conservation.

Keywords: Convolutional neural networks; YOLO; Marine megafauna; floating marine litter.

1. Introducción

El aprendizaje profundo o *Deep Learning* (DL) constituye un área del aprendizaje automático o *machine learning* (ML), cubriendo ámbitos tan diversos como las ciencias económicas (Wang et al., 2023) o de la salud (Bocanegra-Valencia et al., 2024), entre otros, y representa un importante avance en el campo de la detección y clasificación de objetos (Justel-Pizarro, 2022, Aldaya et al., 2023). El DL incluye las redes neuronales profundas, cuya arquitectura consta de una sucesión de capas con estructura jerárquica inspirada en la corteza visual humana e incorpora algoritmos capaces de procesar grandes cantidades de material digital (Ren et al., 2015; Guirado et al., 2019). Las redes neuronales convolucionales (CNN) se especializan en procesamiento de imágenes para reconocimiento de objetos; a diferencia del ML tradicional, pueden extraer elementos de imágenes de forma autónoma, aprender rasgos y optimizar sus pesos para determinar la clase de objeto. Además, poseen mayor eficiencia computacional y capacidad para capturar la estructura espacial de los datos (Khan et al., 2020; Jiang, 2024).

En el ámbito ecológico, las redes neuronales se han aplicado a detección de objetos como basura marina flotante (García-Garin et al., 2021) y cetáceos. En estos últimos se han utilizado redes como Random Forest, RESNET (Lee et al., 2021), Faster R-CNN (Guirado et al., 2019), YOLOv7 (Aisaldi et al., 2024) y YOLOv5 (Green et al., 2023), mayormente en ballenas grises y belugas. También se ha aplicado YOLOv5 en delfines (Al-Abbar et al., 2024) y U-Net-YOLO-ResNet101a en rorcales comunes (Sanikomm et al., 2025). Sin embargo, el uso de estas redes aún no es extensivo ni se ha aplicado a varias especies simultáneamente en una misma zona. La mayor parte de estudios de megafauna se han basado en monitoreos visuales, de ahí que la incorporación de un sistema automatizado que incluya aplicación de CNNs, sería un excelente soporte al monitoreo visual tradicional para programas de conservación de variedad de cetáceos vulnerables al creciente impacto antrópico en el medio marino (Tethys Research Institute – Cetacean Sanctuary Research, 2024).

YOLO (“*You Only Look Once*”) es una CNN de última generación, lanzada en 2015 y basada en el enfoque de detector de disparo único, que visualiza la imagen completa, localiza y clasifica el objeto en una sola pasada, siendo capaz de procesar múltiples objetos simultáneamente en un único y eficiente flujo de tareas (Redmon et al., 2016; Reis et al., 2024). En comparación con otros sistemas de detección de vanguardia, YOLO presenta menor tasa de errores de fondo, menor probabilidad de detecciones falsas y alta robustez o capacidad de generalización. La versión 8 de YOLO (YOLOv8) demuestra un rendimiento mejorado en objetos pequeños, altamente móviles y con orientaciones e iluminaciones variables. Su mayor precisión de detección radica en su capacidad para centrarse en áreas importantes de la imagen en diferentes escalas y niveles de detalle, empleando un menor tiempo de procesamiento. Además, es considerada una versión altamente estable y su flexibilidad ofrece más opciones para personalizar la arquitectura de la red (Solawets, 2023; Jocher, 2023; Reis et al., 2024). Por dichas razones fue la versión escogida para el presente trabajo.

En este estudio se evalúa el potencial de YOLOv8 bajo la hipótesis de su eficiencia en la detección y clasificación de distintos tipos de objetos (megafauna y basura marina) y de su viabilidad en programas de monitoreo para la conservación del ambiente marino y sus especies. Los resultados de este trabajo, así como posteriores actualizaciones, se compartirán en repositorios de libre acceso.

2. Materiales y métodos

2.1. Adquisición del material

Para el análisis de megafauna se utilizaron 26 vídeos de alta definición cada uno de un minuto de duración aprox. obtenidos a 240-300m de altitud en el área marina protegida del Santuario Pelagos (Golfo de Liguria, Mediterráneo Occidental). Para el análisis de basura marina se utilizaron imágenes obtenidas a 20-120m de altitud en un área de la costa catalana. El material digital se obtuvo mediante drones a las altitudes apropiadas para la correcta detección del objeto (García-Garin et al., 2021; Tethys Research Institute – Cetacean Sanctuary Research, 2024).

2.2. Preprocesamiento de datos

Se extrajeron fotogramas de alta resolución (1080p) de los vídeos mediante un algoritmo de Python en google.colab, registrando seis especies de delfines (*Odontoceti*) y ballenas (*Mysticeti*) (Tabla 1). Las imágenes de cetáceos y basura, fueron etiquetadas en la plataforma Makesense.AI, basándose en rasgos para identificación de cetáceos en censos visuales tradicionales (ej. aletas, forma del cuerpo, color, tamaño), en formas básicas de los diferentes ítems de basura marina con flotabilidad neutra o positiva e incluyendo medusas vivas, por su abundancia en las imágenes de basura (Tabla 2). Se analizaron cerca de 700 imágenes de cetáceos y 300 de basura marina, generándose las etiquetas con el número de clase (0 a n) y coordenadas (x, y) de las esquinas superior izquierda e inferior derecha del objeto, especificadas en porcentaje del ancho y alto de la imagen. Las imágenes de entrenamiento y validación se dividieron aleatoriamente en proporción 80%- 20%, respectivamente.

Clase (nombre científico)	Etiqueta en formato Python	Nombre común	Tiempo de observación en video (%)	Proporción de imagenes (%)
<i>Balaenoptera physalus</i>	0	Rorcual común	33.2	34
<i>Physeter macrocephalus</i>	1	Cachalote	32.8	12.8
<i>Globicephala melas</i>	2	Calderón común	10.3	22.2
<i>Ziphius cavirostris</i>	3	Zifio de Cuvier	8.0	18.3
<i>Grampus griseus</i>	4	Delfín de Risso	2.3	7.8
<i>Stenella coeruleoalba</i>	5	Delfín listado	13.5	4.0

Tabla 1 – Categorías y porcentaje de imágenes de los cetáceos evaluadas con YOLOv8.

Clase (tipo de ítem)	Etiqueta en formato Python	Descripción	Proporción de imagenes (%)
Plástico objeto	0	Material inorgánico de aspecto sólido de distintas formas (ej caja, bolsa, botella, cilindro...)	29.8
Medusas	1	Medusas vivas de color blancuzco	23.9
Otros	2	Material inorgánico o vegetal de aspecto irregular	8.0
Plástico agregados	3	Franjas amplias o nubes de puntos blancos a traslúcidos	36.2
Boyas	4	Objetos de aspecto sólido, forma redondeada o rectangular	2.1

Tabla 2 – Descripción de las categorías y porcentaje de imágenes de los objetos evaluadas con YOLOv8 en el set de datos de basura marina.

2.3. Funcionamiento del modelo

Los modelos YOLO se basan en la arquitectura básica de capas interconectadas de las CNN (fig. 1): Por la capa de entrada, ingresa la imagen transformada a matriz de píxeles, a modo de tensor RxGxB en el caso de las imágenes a color. En las capas convolucionales, varias ventanas deslizantes, filtros o *kernels* realizan una operación de superposición de píxeles a lo largo de la matriz de datos de la imagen, para extraer rasgos de complejidad creciente (ej. bordes: perfiles, coloraciones, robustez) a medida que se profundizan las capas. Las capas no lineales aplican funciones de activación (ej. ReLU) para capturar la no-linealidad de las interacciones entre características. Las capas de agrupamiento reducen las dimensiones del mapa de características (compresión de la imagen), reduciendo los parámetros para aprender, pero manteniendo la información necesaria para reconocer el objeto. La capa totalmente conectada transforma el mapa de características en un vector unidimensional, donde cada entrada se conecta a cada

salida por un peso y aplicando la función de activación *softmax*, calcula la probabilidad de cada clase (Fergus, 2015; Redmon et al., 2016; Khan et al., 2020; Jiang, 2024). La arquitectura de YOLOv8 se deriva de dicha arquitectura básica, estructurándose en tres componentes: Columna vertebral, donde las capas convolucionales extraen características visuales a diferentes escalas y que opera bajo la CSPDarknet53 (una CNN profunda eficiente con 53 capas convolucionales), cuello, que incorpora estructuras de agregación de características (FPN, PAN), para mejorar la robustez a la escala y a la forma, y cabeza, que predice las clases de objetos y las coordenadas del *bounding box*, para producir el resultado de salida (Solawets, 2023; Reis et al., 2024; Yaseen, 2024).

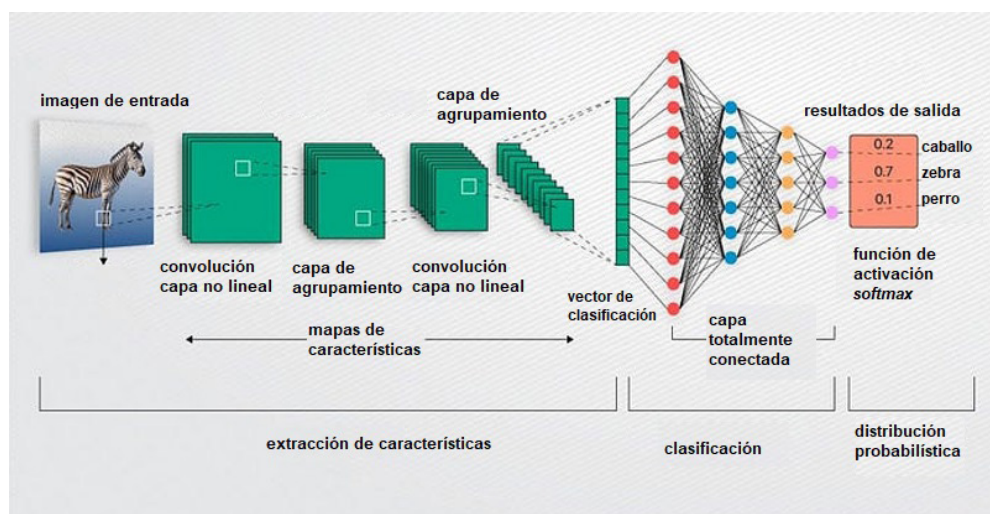


Figura 1 – Representación simplificada de una CNN.

Fuente: AnalytixLabs <https://www.analytixlabs.co.in/blog/convolutional-neural-network/>

El entrenamiento de una red neuronal busca obtener valores de parámetros que a partir de unos datos de entrada y pasando a través de las distintas capas, produzcan una salida o *output* lo más cercano posible al valor real esperado (Justel-Pizarro, 2022). Este proceso se realiza en dos pasos:

1. **Propagación hacia adelante (*forward propagation*):** Los valores de entrada de píxeles por cada neurona van pasando a lo largo de las capas ocultas, y van multiplicándose por un peso o importancia de dicha entrada. Luego se aplica la suma ponderada de dichas multiplicaciones (*dot product*) y el resultado pasa por una función de activación para añadir la no-linealidad.
2. **Retropropagación (*backpropagation*):** Partiendo de la capa de salida hacia atrás, la red calcula los errores (*bias*) para cada entrada y va ajustando los pesos de las conexiones entre nodos mediante el método de *gradient descent*, calculando la tasa de cambio en gradiente negativo del coste o función de pérdida (diferencia entre el valor calculado y el valor real), hasta alcanzar un mínimo (aplicando la regla de la cadena para la derivación). El proceso se repite durante el número

de iteraciones que se haya especificado en el código. De esta manera se van actualizando los valores del mapa de características (Redmond et al., 2016; Reis et al., 2024).

El sobreajuste (*overfitting*) ocurre cuando el modelo tiene alto rendimiento en el conjunto de datos de entrenamiento, pero que empeora en la validación con datos no vistos anteriormente (Smith, 2018; Zhang et al., 2021, Justel-Pizarro, 2022). Por el contrario, el *underfitting* puede ocurrir por bajo rendimiento del modelo, refleja baja capacidad de aprendizaje, donde la función de pérdida decrece continuamente pero no logra reducir el error de predicción al mínimo (Smith, 2018).

2.4. Ejecución de YOLO

YOLOv8 se ejecuta bajo el *framework* de aprendizaje profundo de código abierto PyTorch, muy eficiente en matrices multidimensionales o tensores. El programa se instaló desde la interfaz CLI de Ultralytics y accediendo a una GPU. El modelo es entrenado previamente con el set de datos MS-COCO, un amplio conjunto de imágenes de objetos comunes, que permite aprender características fundamentales presentes en cualquier tipo de objeto (Lin et al., 2014). El código de YOLOv8 se escribió basándose en el repositorio de Ultralytics. Se definió el modelo (estructura de la red y la variante de ésta), usando la versión pequeña, YOLOv8n, más simplificada, (con 168 capas y 3006818 parámetros), es más pequeña pero más rápida y conveniente al contar con recursos computacionales limitados. Luego, el parámetro “data” indica la ruta del archivo “.yaml” con la estructura de los datos, incluyendo los nombres de las clases.

El tipo de tarea (*task*), se especificó para el objetivo de reconocimiento de objetos:

- *detect*: para localización del objeto mediante el *bounding box* o cuadro delimitador.
- *segment*: para detección del contorno del objeto.
- *classify*: para clasificación grupos o lotes de imágenes introducidos como datos de entrenamiento dentro de la clase probable.

El modo se especificó con los argumentos train, val y predict y con distintos hiperparámetros (Smith, 2018; Justel-Pizarro, 2022), tales como:

- *epochs*: Numero de iteraciones, en cada una de las cuales la red utiliza todo el conjunto de datos de entrada para hacer las predicciones.
- *Patience* (o detención temprana): Número de épocas para esperar mejoras en las métricas de validación, en caso contrario, detención del proceso.
- *Batch*: Tamaño de lote. Especifica el número de submuestras de entrenamiento procesadas por cada iteración. Se utilizó un tamaño de 16.
- *Imgsz*: Tamaño de la imagen que ingresa a CNN. Se utilizaron imágenes de alta resolución (1080p) para mayor precisión del modelo.
- *Lr* (*learning rate* o tasa de aprendizaje): mide el avance entre iteraciones. Se utilizó el modo cíclico, desde 0.001 gradualmente hasta 0.0001, para un aprendizaje más regulado y avance más controlado hasta el mínimo de la función de pérdida, evitando sobreajuste.
- *Momentum*: Se utilizó un valor relativamente alto (0.937) para que la tasa de cambio de gradiente de la función de pérdida conduzca al mínimo global.

- *Weight decay*: Regula los pesos, minimizando la función de pérdida y el sobreajuste. Se utilizó 0.0005 recomendado para sets de datos pequeños.
- *Optimizer*: Se utilizó el algoritmo optimizador Adam, por su robustez y eficiencia como mecanismo regulador durante el descenso de gradiente.

Se ejecutaron ocho ensayos del modelo con los cetáceos, variando los argumentos “*epochs*” y “*patience*”. Adicionalmente, se efectuó un ensayo con la basura marina para evaluar el modelo con una nueva categorización de ítems para comparar con resultados anteriores (Garcia-Garin et al., 2021).

2.5. Métricas de evaluación

A lo largo del entrenamiento y la validación se espera que la función de pérdida varíe inversamente con la precisión media o *mean average precision* (mAP), indicada por:

- mAP50: precisión de la predicción que alcanza una superposición del 50 % con el *bounding box* real (umbral IoU de 50).
- mAP50-95: considera detecciones con mayor grado de dificultad a diez umbrales de IoU, de 0,5 a 0,95, en pasos de 0,05 (Reis et al., 2024).

Con base en las probabilidades de clase predichas, los verdaderos positivos (VP) corresponden a clasificaciones correctas, los falsos positivos (FP) corresponden a objetos asignados a una clase incorrecta o confundidos con fondo (o *background*) y los falsos negativos (FN) corresponden a objetos reales, no detectados, sino confundidos con fondo. Se aplicaron las siguientes métricas de rendimiento de acuerdo con las condiciones de equilibrio de las muestras durante el entrenamiento (Garcia-Garin et al., 2021):

- a. Precisión: proporción de todas las muestras positivas correctas sobre el total de muestras. $\text{Precisión} = \text{VP} / (\text{VP} + \text{FP})$
- b. *Recall* (recuperación): proporción de muestras positivas correctas sobre el total de muestras positivas. $\text{Recall} = \text{VP} / (\text{VP} + \text{FN})$.
- c. *F1-score*: promedio ponderado entre precisión y *recall*, representa el punto de equilibrio óptimo entre éstas. Refleja el nivel de confianza del modelo al clasificar una muestra como verdadera positiva.
$$F1 = 2 * ((\text{precisión} * \text{recall}) / (\text{precisión} + \text{recall}))$$

3. Resultados

3.1. Evolución del rendimiento del modelo

La tendencia general a la reducción progresiva en la función de pérdida coincide con un aumento progresivo de la mAP. Con las imágenes de cetáceos, el mejor rendimiento se obtuvo ejecutando el algoritmo durante 500 épocas, alcanzando valores sobre 0.9. El patrón ascendente en la función de pérdida, indica un ligero sobreajuste durante la validación, aunque viéndose compensado por una mayor mAP50, sobre todo en el mejor ensayo (Train9). En cambio, a más de 500 épocas (Train 12) el fuerte incremento

del sobreajuste reduce la precisión, mientras que pocas iteraciones evitan alcanzar el aprendizaje necesario (*underfitting*), como se observa especialmente en la basura marina, donde la función de pérdida no se estabiliza y la precisión y *recall* son muy fluctuantes (fig. 2).

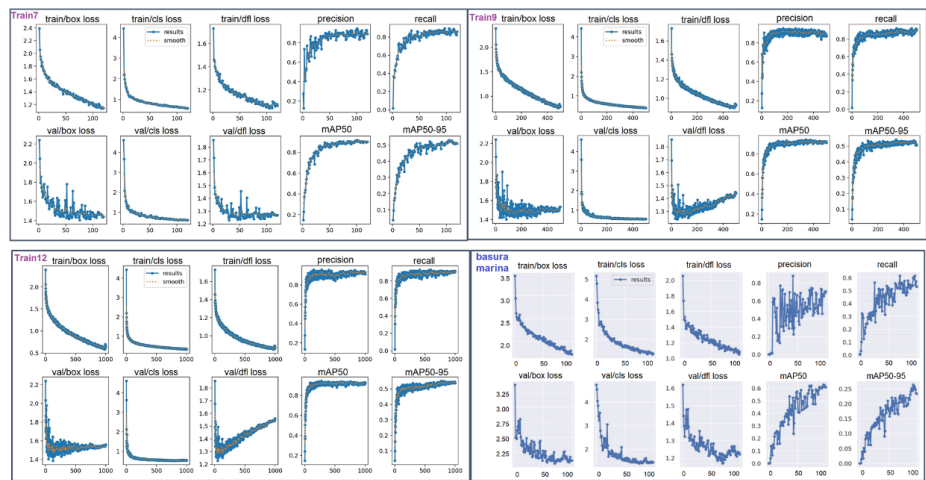


Figura 2 – Función de pérdida Vs precisión de YOLOv8, ilustrando mejor desempeño con los conjuntos de datos de cetáceos (ensayos Train7, Train9, Train12) que con el de basura marina. CLoU (Pérdida de localización): error de coincidencia entre coordenadas de los *bounding boxes* predichos y del objeto real. DFL (pérdida de distribución): diferencia de área de intersección/área de unión (IoU). Pérdida de clasificación (CLS): error entre probabilidad de clase predicha y real (Reis et al. 2024).

3.2. Indicadores de desempeño

Se observaron probabilidades de detección correcta sobre 90% en todas las clases de cetáceos. En la basura marina, las categorías “medusas”, “boyas” y “otros” alcanzaron mayores valores en comparación con las clases de plástico “objetos” y “agregados” y las proporciones de FP y FN fueron relativamente altas (Tabla 3).

Clase	VP (verdaderos positivos)	FP (Falsos Positivos)	FN (Falsos Negativos)	F1-score	Recall
Cetáceos					
Rorcual común	1.00	0.00	0.12	0.94	0.89
Cachalote	0.95	0.05	0.10	0.78	0.68
Zifio de Cuvier	0.98	0.02	0.05	0.96	0.97
Calderón común	0.95	0.05	0.47	0.90	0.91
Delfín de Risso	1.00	0.00	0.00	0.99	1.00
Delfín listado	0.90	0.10	0.25	0.74	0.69

Clase	VP (verdaderos positivos)	FP (Falsos Positivos)	FN (Falsos Negativos)	F1-score	Recall
Basura marina					
Plástico objeto	0.02	0.77	0.00	0.03	0.03
Medusas	0.94	0.06	0.87	0.66	0.50
Otros	0.95	0.05	0.05	0.93	0.91
Plástico agregados	0.38	0.60	0.06	0.38	0.37
Boyas	1.00	0.07	0.02	0.93	0.92

Tabla 3 – Tasas de predicción (%) obtenidas de la matriz de confusión normalizada (basada en número de instancias detectadas), F1-scores y Recall para las clases de cetáceos y de basura marina. Los mejores resultados se obtuvieron con cetáceos (ensayo Train9).

La mAP50 global en los cetáceos osciló alrededor de 0.9 en la mayoría de ensayos, y por clases fue mayor en el delfín de Risso, en el rorcual común y en el zifio de Cuvier, y menor en el delfín listado (Tabla 4). El *underfitting* que se aprecia en el ensayo con basura marina, coincide con las métricas muy fluctuantes en el conjunto de validación (fig. 2), lo cual se reflejó en una mAP-50 global de 0.6, y valores sobre 0.8 solamente en dos de las cinco categorías (“medusas” y “otros”).

Especie	Train	Train3	Train7	Train8	Train9	Train10	Train11	Train12
Rorcual común	0.990	0.983	0.982	0.972	0.985	0.972	0.985	0.973
Cachalote	0.878	0.844	0.828	0.800	0.912	0.800	0.912	0.925
Calderón común	0.968	0.979	0.976	0.973	0.959	0.973	0.959	0.974
Zifio de Cuvier	0.937	0.895	0.873	0.904	0.895	0.904	0.895	0.844
Delfín de Risso	0.995	0.995	0.995	0.995	0.995	0.995	0.995	0.995
Delfín listado	0.731	0.744	0.748	0.804	0.788	0.804	0.788	0.777
mAP Global	<i>0.916</i>	0.907	0.900	0.908	0.922	0.908	0.922	0.915
Hiperparámetros de referencia								
<i>epochs</i>	150	150	120	500	500	500	500	1000
<i>patience</i>	150	50	50	50	500	150	250	350

Tabla 4 – mAP-50 global y por clases de cetáceos en los diferentes ensayos, indicando hiperparámetros de referencia (*epochs* y *patience*). Mejores resultados: ensayo Train9.

Las curvas de F1-score más planas indican mayor precisión, mientras que las más redondas o las irregulares reflejan *underfitting* en algunos ensayos con cetáceos (ej. train8) y en especial en la basura marina (fig 3). Los mayores F1 se obtuvieron con las imágenes de delfín de Risso (0,997), del zifio (0,961) y del rorcual común (0,936).

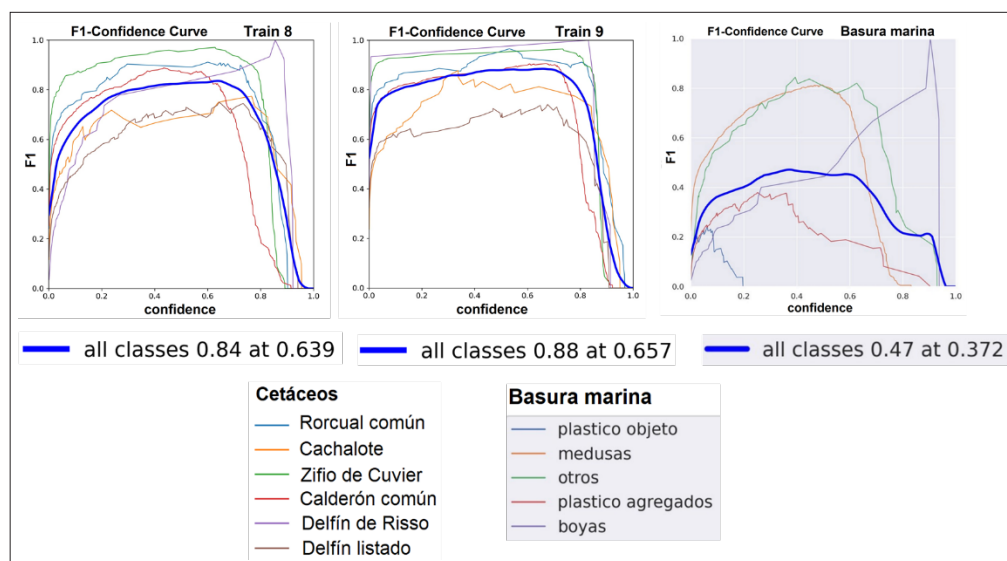


Figura 3 – Curvas de F1-score y valores de F1 global (*all classes*). Mejor ensayo: Train9.

4. Discusión

4.1. indicadores de rendimiento del modelo

La reducción consistente de la función de pérdida, alcanzando un mínimo estable, sumado al aumento de la mAP, indican mejora progresiva en la capacidad del algoritmo para clasificar correctamente los objetos y generalizar en datos no observados (Yu & Liu, 2023). La precisión máxima de nuestro modelo ($>0,95$) se considera satisfactoria para detección de cetáceos con redes neuronales (ej. Guirado et al., 2019; Gray et al., 2019). El *overfitting* presente en algunos ensayos, indica alto rendimiento del modelo durante el entrenamiento, con fuerte reducción de la función de pérdida, pero que se incrementa en la validación, a partir de cierto número de iteraciones, como en el caso del ensayo Train12 (fig 2). Esta dificultad del modelo para reajustarse a datos no vistos anteriormente puede deberse a factores como el uso de más iteraciones de las necesarias, el tamaño reducido del conjunto de datos y la representación no equilibrada de clases. A su vez, el *underfitting*, más acusado en los ítems de basura marina, indica dificultad para reducir el error en la predicción y podría deberse a la ejecución de menos iteraciones de las necesarias, o a la complejidad baja del modelo (ej. menos capas de las necesarias), lo que limita una buena generalización en la interpretación de rasgos de los objetos (Smith 2018; Zheng et al., 2021; Justel-Pizarro, 2022).

Otros estudios de megafauna marina, han encontrado que los F1-scores suelen ser más altos en cetáceos con nado más lento o con largos periodos de reposo en superficie, lo que lleva a mayor precisión en la detección (Guirado et al., 2019). Esto fue cierto en el

rorcual común, que presentó una elevada proporción de detecciones correctas, un alto F1-score y una alta mAP (Tabla 3), lo cual coincide con otras evaluaciones con YOLO en esta especie (Sanikommu et al., 2025). Precisiones de clasificación bajas y número elevado de FP, suelen presentarse en el delfín listado, debido a su talla pequeña en comparación con las otras especies, nado rápido y con saltos, pudiendo confundirse con olas y ocasionando perturbaciones en el agua, como más olas, espuma y salpicaduras, que constituyen rasgos del fondo que reducen el contraste con el animal (Borowicz et al., 2019; Redmon et al., 2016; Reis et al., 2024). Así mismo, el calderón común, si bien es una especie grande, usualmente se observa en grupos densos y su coloración oscura reduce el contraste con el fondo, lo que limita la capacidad del modelo incrementando los FNs. En cambio, el delfín de Risso, a pesar de su forma de nado que también perturba el agua, es de mayor tamaño que el delfín listado y su color gris con rasgos blancuzcos destaca sobre el color del mar favoreciendo el contraste con el fondo.

La *recall* está estrechamente relacionada al f1-score, y aumenta a mayor número de casos positivos reales y a menor número de FN (Yu & Liu, 2023). Así, la mayor capacidad de identificar VP se traduce en mayores F1-scores, reflejando mayor precisión. Sin embargo, a medida que aumenta el número de objetos (umbrales de confianza muy altos), el modelo se vuelve más selectivo, incrementándose la probabilidad de introducción de FP y reflejándose en una menor *recall* (Guirado et al., 2019; Yu & Liu, 2023). La estabilidad del F1-score (curvas más planas) indica robustez del modelo, mientras que curvas más redondeadas o irregulares indican menor estabilidad (fig 3). Menores valores de F1-score y de *recall* en el cachalote y en el delfín listado, concuerdan con mayor incidencia de FNs y menor de VPs, a diferencia del rorcual común y el zifio de Cuvier (Tabla 3). El cachalote es una especie de talla grande que suele permanecer casi inmóvil en superficie, lo que pudo incidir, no obstante, en su valor de mAP relativamente alto (Tabla 4). Sin embargo, su color grisáceo crea un bajo contraste con el fondo, lo que reduce el F1-score. En cambio, el zifio de Cuvier y el delfín de Risso, cuyos colores más claros contrastan bien con el *background*, mostraron mejores F1-scores. Por su parte, en basura marina se han reportado F1-scores del orden de 0.40, mencionando la limitación de contrastes debida a los contornos indefinidos y al aspecto traslúcido de los objetos (García-Garin et al., 2021), como ocurrió en este estudio con la clase “plástico agregados”.

La mAP-0,5 superior a 0,9 global y en la mayoría de clases de cetáceos (tabla 4), refleja un equilibrio satisfactorio entre precisión y *recall*, y concuerda con otros estudios de detección de ballenas (Alsaidi et al., 2024; Green et al., 2022; Sanikommu et al., 2025). La mAP50-95 global estuvo alrededor de 0,54 en todos los ensayos de cetáceos. Al ser una medida para detecciones más complejas, este valor se considera el estado actual del arte para esta métrica (Alsaidi et al., 2024).

La capacidad predictiva positiva de una CNN se considera elevada cuando las detecciones reales superan ampliamente a las falsas (Guirado et al., 2019), tal como ocurrió con los cetáceos, a diferencia de la basura marina. Por ejemplo, en las clases de plásticos “objeto” y “agregados”, la proporción de FP fue mayor a 50% respecto a VP. Esto sumado al patrón de *underfitting* en la función de pérdida (fig 2) señalan menor estabilidad del modelo. La presente evaluación de YOLOv8 en basura marina utiliza una categorización de ítems diferente a evaluaciones previas con otras CNNs en material del mismo tipo y que resultaron más precisas (García-Garin et al., 2021). No obstante, con nuestro

reducido tamaño muestral y la $mAP > 0.6$, el rendimiento puede considerarse aceptable y es susceptible de mejoras, como por ejemplo utilizando una cantidad de imágenes considerablemente mayor y aplicando al menos 400 iteraciones (García-Garin et al., 2021). En comparación, un estudio de detección de microalgas marinas ($< 500 \mu m$) con YOLO en un reducido set de muestras obtuvo una $mAP = 0.61$ para la clase “diatomeas” confirmando el buen desempeño de esta CNN (Aldaya et al., 2021, Justel-Pizarro, 2022).

4.2. Factores que inciden en el rendimiento del modelo

El sesgo de las métricas de rendimiento entre diferentes categorías de objetos, ocurre frecuentemente en modelos de CNNs. La disponibilidad variable de las especies para el método de detección, refleja sus amplias diferencias en abundancias en el medio marino, limitando la representación equilibrada de las clases en el material de estudio. Incrementar el tamaño muestral ayuda a balancear la representación de las clases, mejorando significativamente el rendimiento de las CNNs. Así mismo, ajustar hiperparámetros durante el entrenamiento, tales como el número de iteraciones para minimizar la función de pérdida y el *overfitting* en el caso de los cetáceos, lleva a resultados satisfactorios. Otras variaciones, tales como ajustes en la tasa de aprendizaje, momentum y tamaño de *batches*, entre otros, son recomendables para ensayos posteriores (Smith, 2018; Justel-Pizarro, 2022) y gracias a la flexibilidad de YOLOv8 podrán permitir una continuidad al presente estudio, para obtener mejoras en los resultados.

Por otro lado, la escala de categorías responde al nivel de detalle en el que se plantean los objetivos de trabajo. Si bien una escala muy general contribuye a elevar la precisión del modelo al tener menos probabilidad de equivocarse, una mayor variedad de clases puede ofrecer al modelo más características para aprender y es útil cuando se requiere diferenciar entre categorías como múltiples especies de fauna marina que ocurren simultáneamente (Borowicz et al., 2019; Gray et al., 2019, Green et al., 2022).

Dado que la identificación de las clases se lleva a cabo en un entorno fluido, los límites del objeto cambian constantemente o se difuminan, lo que representa un reto para el modelo, viéndose favorecido su rendimiento con rasgos que contrastan con el fondo, como la coloración y el tamaño relativo.

Las mejoras documentadas de YOLOv8 en la detección de objetos de tamaños diferentes, en movimiento, en ángulos más amplios, en diferentes acercamientos y en condiciones de iluminación más difíciles, incrementan su eficacia en una amplia variedad de escenarios de detección de objetos, siendo incluso capaz de diferenciar entre clases muy similares detectando sólo parte del objeto (Reis et al., 2024). Futuros estudios podrán dirigirse a constatar el alcance de su capacidad en este aspecto, ya que en especies de megafauna en medio acuático es común observar sólo una porción emergida, por ejemplo, cuando los cetáceos reposan en superficie.

5. Conclusiones

- Este trabajo aporta como contribuciones originales, una primera implementación de YOLOv8 en la detección y clasificación de especies de megafauna marina, en particular en un conjunto de datos multiespecífico de seis especies de cetáceos. Por otro lado, constituye también un primer aporte a la detección y clasificación

de basura marina flotante mediante este tipo de CNN. Esto sugiere el amplio alcance de YOLOv8 para el reconocimiento de objetos con características muy dispares. En cuanto al área de estudio, este trabajo también constituye una primera implementación de YOLO8 en el Mar Mediterráneo.

- Los resultados obtenidos son prometedores, considerando el tamaño reducido de los sets de datos analizados. Se muestra el rendimiento satisfactorio de YOLOv8 en un medio fluido, como es el mar, sujeto a las diversas perturbaciones de fondo (ej. oleaje turbidez, luz solar etc.) y a los distintos grados de contraste con el color del objeto.
- La flexibilidad de YOLOv8 para hacer ajustes adicionales en distintos hiperparámetros, permitirá mejorar los resultados y proyectarse para alcances potenciales subsiguientes, que podrían incluir la evaluación más sistematizada de la distribución de megafauna y basura marina a lo largo de áreas más extensas.
- YOLOv8 es un sistema de detección automatizado que puede ser utilizado para apoyar programas de monitoreo de megafauna, ya que, a diferencia de los censos visuales tradicionales, ofrece la posibilidad de análisis posteriores y revisión de los datos obtenidos. El método de CNN presentado podrá ser útil igualmente para automatizar información de *hotspots* de basura marina en áreas geográficas con ambientes susceptibles a contaminación por residuos sólidos.

Agradecimientos

El material de video de megafauna marina fue compartido por el proyecto “Eye in the Sky”, llevado a cabo por la Guardia Costera Italiana y el Tethys Research Institute. El vehículo aéreo autónomo (TEKEVER AR5 Evolution) empleado para la obtención de los videos, fue facilitado por la Agencia Europea de Seguridad Marítima (EMSA) y operado por el consorcio REACT. Las imágenes aéreas de basura marina flotante, fueron suministradas por el proyecto “MEDSEALITTER” (García-Garin et al., 2021).

Referencias

- Al Abbar, F., Eikelboom, J., Costelloe, B.R., Lorenzo, F., Ramon, H., Visser, F., González-García, L., Aarts, G. & van-Langevelde, F. (2024). Automated tracking algorithm for groups of common dolphins (*Delphinus delphis*). In: *Abstract Book 35th European Cetacean Society (ECS) Conference*. Catania, Italy. (pp-171).Marecamp-Association. www.europeancetaceansociety.eu/conference-program-abstract-book
- Aldaya, A., Monleón, T., Flor, N. & Justel-Pizarro, J. (2023). *Object Detection with Machine Learning, a YOLO review*. In: https://www.researchgate.net/publication/372419433_Object_Detection_with_Machine_Learning_a_YOLO_Review
- Alsaïdi, M., Al-Jassani, M. G., Bang, C., O’Corry-Crowe, G., Watt, C., Ghazal, M., & Zhuang, H. (2024). Localization and tracking of beluga whales in aerial video using deep learning. *Front. Mar. Sci.* 11:1445698. doi: 10.3389/fmars.2024.1 445698.

- Bocanegra-Valencia, J.P., Urrelo-Huiman, L.V. & Cieza-Mostacero, S.V. (2024). Deep Learning para la detección de Cáncer de Pulmón: Una Revisión Sistemática de la Literatura. *RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação*, (N.º E75), 160-174.
- Borowicz, A., Le, H., Humphries, G., Nehls, G., H"oschle, C.,... & Kosarev, V. (2019). Aerial-trained deep learning networks for surveying cetaceans from satellite imagery. *PLoS One*, 14, e0212532.
- Boulent, J., Charry, B., Kennedy, M. M., Tissier, E., Fan, R., Marcoux, M., ... & Gagné-Turcotte, A. (2023). Scaling whale monitoring using deep learning: A human-in-the-loop solution for analysing aerial datasets. *Front. Mar. Sci.* 10, 1099479.
- Garcia-Garin, O., Monleon-Getino, A., Lopez-Brosa, P., Borrell, A., Aguilar, A., Borja-Robalino, R., Cardona, L. & Vighi M. (2021). Automatic detection and quantification of floating marine macro-litter in aerial images: Introducing a novel deep learning approach connected to a web application in R*. *Environmental Pollution* 273: 1-11.
- Gray, P.C., Bierlich, K.C., Mantell, S.A., Friedlaender, A.S., Goldbogen, J.A., Johnston, D.W. & Ye, H. (2019). Drones and convolutional neural networks facilitate automated and accurate cetacean species identification and photogrammetry. *Methods in Ecology and Evolution* 10(9):1490-1500
- Green, K. M., Virdee, M. K., Cubaynes, H. C., Aviles-Rivero, A. I., Fretwell, P. T., Gray, P. C., ... & Jackson, J. A. (2023). Gray whale detection in satellite imagery using deep learning. *Remote Sensing in Ecology and Conservation*, 9(6), 829-840.
- Guirado, E., Tabik, S., Rivas, M. L., Alcaraz-Segura, D., & Herrera, F. (2019). Whale counting in satellite and aerial images with deep learning. *Scientific reports*, 9(1), 14259.
- Jiang J. (2024). The eye of artificial intelligence-Convolutional Neural Networks. *Proc. 2nd International Conference on Software Engineering and Machine Learning*. 76(1):273 – 279
- Jocher, G. 2023. Ultralytics YOLOv8. <https://github.com/ultralytics/ultralytics> (accessed Nov 10, 2025).
- Justel-Pizarro, J.T. (2022). *Clasificación automática de formas biológicas mediante deep learning*. Tesis Máster Interuniversitario en Estadística e Investigación Operativa. Universidad Politécnica de Catalunya. Departamento de Estadística e Investigación Operativa. <https://upcommons.upc.edu/bitstream/handle/2117/371950/memoria.pdf?sequence=1&isAllowed=y>
- Khan, A., Sohail, A., Zahoor, U. & Qureshi, A.S. (2020) A survey of the recent architectures of deep convolutional neural networks. *Artificial Intelligence Review*, 53, 5455–5516
- Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P.,... & Ramanan, D. (2014). Microsoft COCO: common objects in context. *Computer vision – ECCV 2014. Cham*, pp. 740–755. doi: 10.1007/978-3-319-10602-1

- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. arXiv:2305.09972v2 [cs.CV]. 10 p. <http://pjreddie.com/yolo/>
- Reis, D., Kupec, J., Hong, J., & Daoudi, A. (2024). Real-Time Flying Object Detection with YOLOv8. arXiv:2305.09972v2 [cs.CV]. <https://doi.org/10.48550/arXiv.2305.09972>
- Ren, S., He, K., Girshick, R. & Sun, J. (2015). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. in *Advances in Neural Information Processing Systems* 28 (eds Cortes, C., Lawrence, N. D., Lee, D. D., Sugiyama, M. & Garnett, R.) 91–99 (Curran Associates, Inc., 2015).
- Sanikommu, V., Sura, A., & Chimirala, P. (2025). Deep Learning-Based Detection and Tracking of Fin Whales Using High-Resolution Space-Borne Remote Sensing Data. *Remote Sensing Applications: Society and Environment*, 101580.
- Solawetz, J., Francesco. (2023). *What is yolov8? the ultimate guide*. <https://blog.roboflow.com/what-is-yolov8/>
- Smith, L. N. (2018). A disciplined approach to neural network hyper-parameters: Part 1 -- learning rate, batch size, momentum, and weight decay, arXiv:1803.09820 [cs.LG]. <http://arxiv.org/abs/1803.09820>
- Tethys Research Institute–Cetacean Sanctuary Research. (2024). *Eye in the Sky*, Technical activities Report, Phase I. By Costa, M., De Santis, V., Lanfredi, C., & Airoldi, S. Tethys Research Institute. Italy. 75 p.
- Wang, X., Zuraidah, R., Binti, R. & Rasi, M. (2024). Consumer Demand Prediction Algorithm Based on Deep Learning in C2M Mode Supply Chain. RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação, (N.º E67), 372-382.
- Yaseen, M., (2024). What is YOLOv8: An in-depth exploration of the internal features of the next generation object detector. ArXiv:2408.15857v1 [cs.CV] <https://doi.org/10.48550/arXiv.2408.15857>
- Yu, L. & Liu, A. (2023). Single-Stage Deep Learning-based Approach for Real-Time License Plate Recognition in Smart Parking System. *International Journal of Advanced Computer Science and Applications* 14(9). DOI: 10.14569/IJACSA.2023.01409119
- Zhang A., Lipton, Z. C. Li M. & Smola A. J. (2021). Dive into Deep Learning. arXiv preprint arXiv:2106.11342, 2021.

O Futuro da Saúde Preditiva: Uma Análise do Estado da Arte em IoT e Sistemas Embarcados para Identificar Oportunidades de Inovação

Jhonatas Gomes Ribeiro¹, Keyllane Guedes de Souza²,
Phaola Paraguai da Paixão Lustosa³, Wanderson Jean Conceição Silva⁴,
Felipe Gonçalves dos Santos⁵

jhonatasgomes2003@gmail.com; keyllaneguede@gmail.com; paixaphaola@gmail.com;
wanderson.jean@ifpi.edu.br; felipe.santos@ifpi.edu.br

^{1,2,3,4,5} Instituto Federal do Piauí - IFPI, Corrente, 64980-000, Piauí, Brasil

Pages: 55-68

Resumo: A convergência entre Sistemas Embarcados e a Internet das Coisas (IoT) está catalisando uma mudança de paradigma na saúde, transitando de um modelo reativo para uma abordagem proativa e preditiva. Este artigo apresenta uma revisão sistemática da literatura para mapear este cenário, analisando 30 artigos de alto impacto publicados nos últimos cinco anos. Os resultados revelam um crescimento exponencial de publicações, com pico em 2024, e uma concentração geográfica da produção científica na Índia e Arábia Saudita. A principal contribuição tecnológica identificada é a fusão de ecossistemas de sensores vestíveis e sem contato com Inteligência Artificial para viabilizar o monitoramento contínuo, permitindo a detecção precoce de riscos e a criação de ambientes inteligentes de cuidado. A análise sintetiza como essas inovações estão pavimentando o caminho para uma medicina personalizada e remota, ao mesmo tempo que aponta para desafios críticos de segurança de dados e usabilidade que precisam ser superados.

Palavras-chave: Sistemas embarcados; Internet das Coisas; assistência médica; tecnologias assistivas; monitoramento de sinais vitais.

The Future of Predictive Healthcare: A State-of-the-Art Analysis in IoT and Embedded Systems to Identify Innovation Opportunities

Abstract: The convergence of Embedded Systems and the Internet of Things (IoT) is catalyzing a paradigm shift in healthcare, moving from a reactive model to a proactive and predictive approach. This paper presents a systematic literature review to map this landscape, analyzing 30 high-impact articles published within the last five years. The results reveal exponential growth in publications, peaking in 2024, with a geographical concentration of scientific production in India and Saudi Arabia. The main technological contribution identified is the fusion of sensor ecosystems—both wearable and contactless—with Artificial Intelligence to enable continuous monitoring, allowing for early risk detection and the creation of smart care environments. This analysis synthesizes how these innovations are paving

the way for personalized and remote medicine, while also highlighting critical challenges in data security and usability that must be overcome.

Keywords: Embedded systems; Internet of Things; healthcare; assistive technologies; monitoring of connected signals.

1. Introdução

A intersecção entre Sistemas Embarcados e a Internet das Coisas (IoT) está catalisando uma revolução silenciosa na área da saúde, prometendo transformar o cuidado reativo em um modelo proativo e preditivo. O volume de inovações é notável, com sistemas portáteis capazes de medir simultaneamente múltiplos sinais vitais [Siam et al. 2021] e plataformas de baixo custo que viabilizam a captura de dados cardíacos em tempo real [Asif et al. 2022]. No entanto, a rápida expansão deste campo gerou um cenário de pesquisa vasto e fragmentado, tornando crucial a existência de um mapa que organize o conhecimento, identifique as fronteiras da inovação e, mais importante, aponte as lacunas críticas que ainda precisam ser preenchidas.

A relevância dessas tecnologias é inquestionável, especialmente em aplicações de alto impacto como o monitoramento contínuo de idosos e acamados [Kumar and Krishnamoorthi 2021], onde a tecnologia pode identificar padrões que antecedem eventos fatais ou detectar quedas com alta precisão [Tahir et al. 2022]. Ao promover a autonomia dos usuários [Javed et al. 2021], essas soluções já demonstram seu potencial para uma melhoria significativa na qualidade de vida. Contudo, sem uma análise estratégica, a comunidade científica corre o risco de concentrar esforços em problemas já solucionados, negligenciando áreas que necessitam de maior desenvolvimento.

Diante deste desafio, o presente artigo vai além de uma revisão tradicional. Propomos um mapeamento crítico da literatura com o objetivo de sintetizar o estado da arte e fornecer um direcionamento claro para futuras pesquisas. A nossa análise não busca apenas catalogar o que já foi feito, mas sim responder a uma pergunta fundamental: Onde estão as fronteiras do conhecimento em IoT para a saúde e quais são as oportunidades mais promissoras para a próxima geração de inovações? Este trabalho, portanto, serve como um guia para pesquisadores e desenvolvedores, iluminando os caminhos para a criação de soluções que definirão o futuro da medicina.

2. Questões de Pesquisa

Este trabalho apresenta uma revisão de literatura que mapeia as principais soluções, tecnologias, tendências e limitações no uso de sistemas embarcados e Internet das Coisas (IoT) na saúde. A condução do estudo foi fundamentada na metodologia de [Kitchenham and Charters 2007] para definir as questões de pesquisa e guiar a seleção, análise e síntese dos trabalhos relevantes.

Para nortear a investigação, a pesquisa parte de uma questão principal (QP): O que a literatura relata sobre o uso de sistemas embarcados e IoT no desenvolvimento de soluções voltadas à saúde?

A partir deste questionamento central, foram criadas outras perguntas secundárias para aprofundar a análise e estruturar os resultados apresentados neste artigo.

- **PP1:** Qual é o número de artigos publicados por ano relacionados ao uso de sistemas embarcados e IoT na saúde?
- **PP2:** Quais países têm se destacado na produção científica sobre o uso de sistemas embarcados e IoT em soluções de saúde?
- **PP3:** Quais eventos científicos e periódicos concentram a maior parte das publicações sobre sistemas embarcados e IoT aplicados à área da saúde
- **PP4:** Como os sistemas embarcados, quando utilizados em residências, podem contribuir com a qualidade de vida e promover o bem-estar dos seus usuários?
- **PP5:** Como os sistemas embarcados podem ser utilizados para prever fatalidades em pacientes críticos através do monitoramento dos sinais vitais?
- **PP6:** Como os sistemas embarcados integrados a sensores vitais podem identificar ocorrências antes de eventos fatais?

3. Seleção dos Artigos Relevantes

A seleção de artigos relevantes iniciou-se com a elaboração de uma string de busca, etapa fundamental no mapeamento da literatura. A sua construção foi baseada na identificação das palavras-chave centrais ao tema: “Sistemas Embarcados”(Embedded Systems), “Internet das Coisas”(IoT) e “Saúde”(Healthcare). Com o objetivo de otimizar a consulta, foram cuidadosamente incorporados sinônimos, buscando um equilíbrio metodológico para não ampliar os resultados com artigos irrelevantes, nem restringi-la a ponto de excluir estudos pertinentes.

Para refinar a busca e garantir a alta relevância dos resultados, foram aplicados critérios adicionais. A pesquisa foi delimitada a publicações dos últimos cinco anos (entre 2021 e 2025), escritas em inglês e pertencentes às áreas de Computação, Engenharia, Medicina ou Neurociência. Com a adição dos critérios acima, resultou em uma string de busca final: (embed- ded AND system AND healthcare) OR (IoT AND Internet of Things) AND PUBYEAR > 2020 AND PUBYEAR < 2026 AND (LIMIT-TO (SUBJAREA, “COMP”) OR LIMIT-TO (SUBJAREA, “ENGI”) OR LIMIT-TO (SUBJAREA, “MEDI”) OR LIMIT-TO (SUB- JAREA, “NEUR”)) AND (LIMIT-TO (LANGUAGE, “English”)) AND (LIMIT-TO (EXACTKEYWORD , “Embedded Systems”))

As bases de dados acadêmicas selecionadas para este mapeamento são Scopus, ACM e Web of Science. Essas bases foram escolhidas devido à sua abrangência de con- ferências e periódicos mais importantes nas áreas relevantes. Critérios de Inclusão e Exclusão Para garantir a relevância e a qualidade dos estudos, a seleção foi guiada por um conjunto de critérios de inclusão e exclusão pré-definidos. Os critérios de inclusão (CI) foram:

- **CI1:** Artigos escritos em inglês.
- **CI2:** Artigos com data de publicação nos últimos 5 anos (de 2021 a 2025).
- **CI3:** Artigos pertencentes às subáreas de computação, engenharia ou medicina.

Os critérios de exclusão (CE) foram:

- **CE1:** Artigos não relacionados diretamente ao tópico da pesquisa.
- **CE2:** Artigos sem aplicação prática (e.g., apenas revisões bibliográficas ou trabalhos teóricos).

- **CE3:** Artigos que não atendiam aos critérios de inclusão (e.g., duplicatas, idioma incorreto, fora do período de publicação).

A Figura 1 apresenta uma representação dos ciclos da pesquisa. A aplicação da string de busca principal e dos critérios de inclusão/exclusão ocorreu entre 18/03 e 01/08 de 2025.

Inicialmente, a busca retornou os seguintes resultados:

FONTE	QUANTIDADE DE RESULTADOS
SCOPUS	918
ACM	947
WEB OF SCIENCE	861
TOTAL	2726

Tabela 1 – Quantidade de Resultados por Fonte

3.1. Processo de Seleção

A busca inicial retornou 2.726 artigos. O processo de seleção foi realizado em dois ciclos para refinar este conjunto.

No **primeiro ciclo**, após a remoção de duplicatas, foi aplicado o critério CE1 para excluir artigos que não tratavam do uso de microcontroladores na área da saúde. Essa filtragem resultou em 2.311 artigos.

No **segundo ciclo**, foram aplicados os critérios CE2 e CE3. Foram removidos os trabalhos sem aplicação prática e aqueles que não se alinhavam aos critérios de inclusão.

Esta decisão metodológica (CE2), que exclui revisões de literatura, foi deliberada para focar a análise diretamente em estudos primários e aplicações práticas de ponta. Embora as meta-revisões agreguem valor, o foco deste mapeamento foi identificar lacunas e oportunidades de implementação direta, alinhando-se com o objetivo central desta investigação de servir como um guia para o desenvolvimento de novas soluções. Ao final do processo, restaram 30 artigos para a análise final.

4. Resultados

Esta seção responde às seis perguntas de pesquisa com base nos 30 artigos selecionados.

4.1. RQ1: Qual é o número de artigos publicados por ano?

A análise da distribuição anual das publicações revela um interesse crescente na área, apesar de algumas flutuações. Foram identificados 5 artigos em 2021, 6 em 2022 e 5 em 2023, com um pico notável de 12 publicações em 2024, o que demonstra um forte engajamento da comunidade de pesquisa. Os 2 artigos registrados em 2025 refletem uma contagem parcial, uma vez que a coleta de dados foi realizada com o ano ainda em

curso. Desconsiderando este dado preliminar, a tendência geral confirma que o tema é um campo de pesquisa ativo e de relevância crescente.

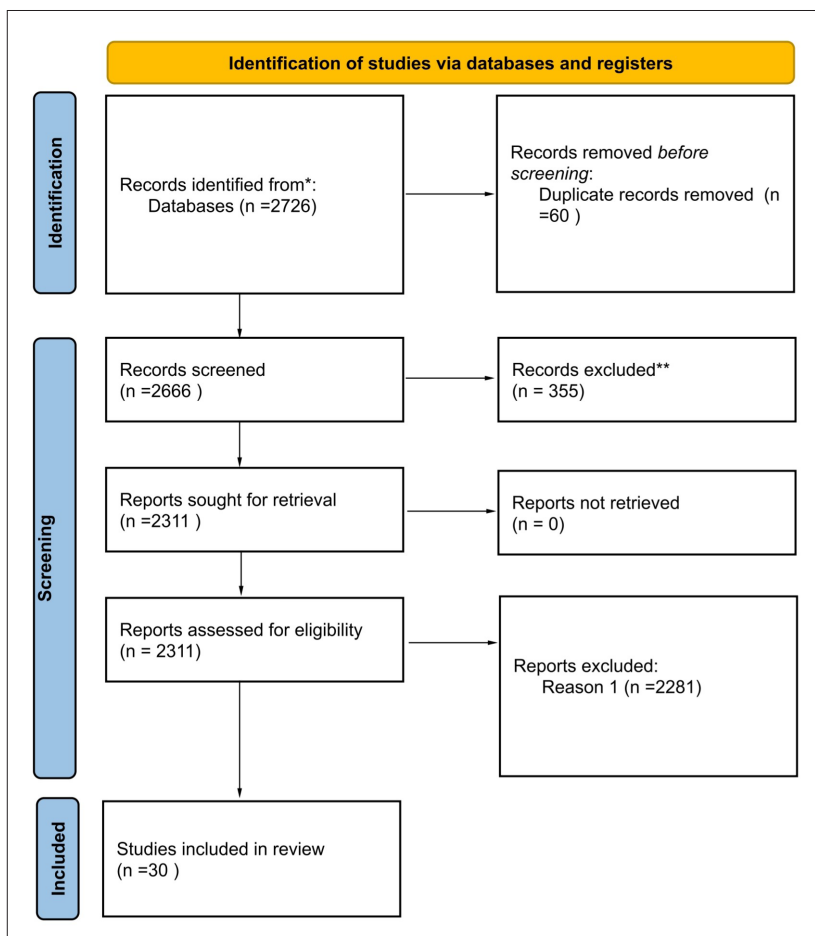


Figura 1 – Fluxograma do processo de seleção dos artigos.

4.2. RQ2: Quais países têm se destacado na produção científica?

Durante o processo de revisão, foram identificados 8 países e 3 continentes: Ásia, Europa e América do Norte. A Figura 3 mostra a distribuição dos países presentes nos resultados da seleção.

A Tabela II apresenta os 8 países com maior produção de artigos sobre o tema. Observa-se que Índia e Arábia Saudita se destacam como os principais produtores. Outros países com número significativo de publicações incluem China e Estados Unidos.

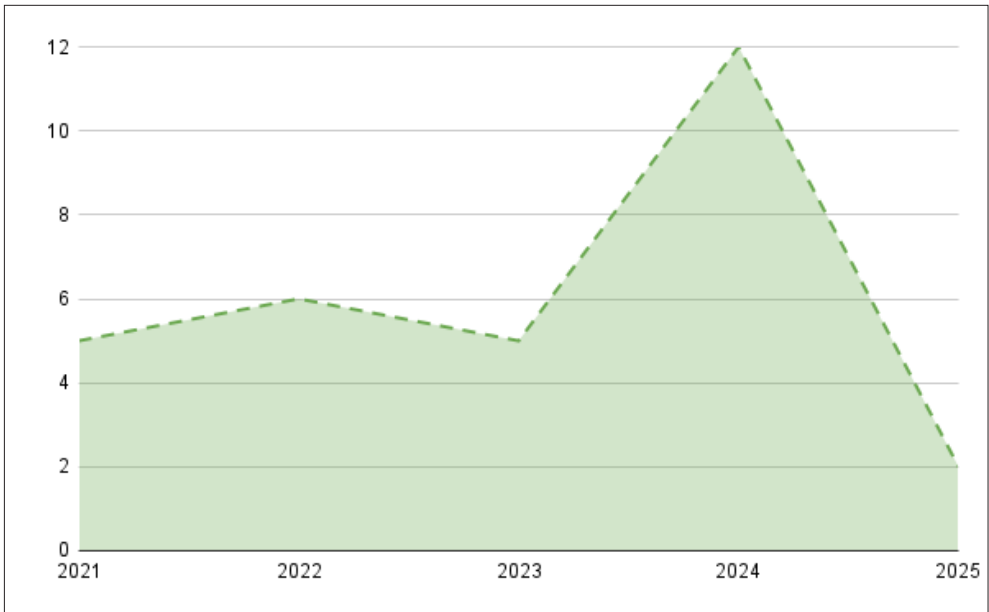


Figura 2 – Quantidade de artigos publicados por ano.

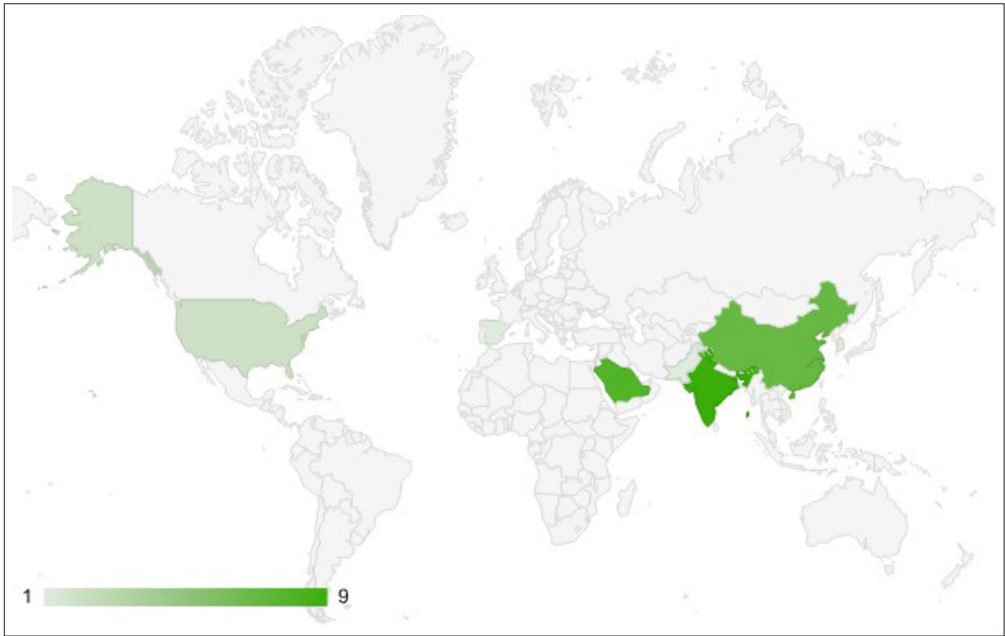


Figura 3 – Distribuição dos países presentes nos resultados da seleção.

Países	Nº	Artigos
Índia	9	Kumar and Krishnamoorthi (2021); Sravanthi et al. (2024); Jain et al. (2024); Abubeker et al. (2024); Yashudas et al. (2024); Prasad et al. (2025); Verma et al. (2022); Kumar and Rajaram (2024); Kaur and Sharma (2024).
Arábia Saudita	8	Siam et al. (2021); Al-Rakhami et al. (2021); Yadav et al. (2022); Arunachalam et al. (2023); Asif et al. (2022); Alkanhel et al. (2024); Almujaid et al. (2023); Vaiyapuri et al. (2024).
China	7	Xu et al. (2023); Shoukat et al. (2024); Zhang et al. (2023); Nekui et al. (2024); Yang (2023); Qian et al. (2022); Umar et al. (2025).

Fonte: Elaboração própria (2025).

Tabela 2 – Países com mais publicações

Outros países entre o número de produção de artigos incluem: Estados Unidos, com 2 artigos [Javed et al. 2021] e [Manogaran et al. 2021]; Coreia do Sul [Astillo et al. 2022]; Paquistão [Tahir et al. 2022]; Espanha [Escobedo et al. 2024]; e Portugal [Rivadeneira et al. 2024], cada um com 1 artigo.

4.3.RQ3: Quais fontes de informação (conferências ou jornais) publicaram mais artigos sobre este tópico?

A Tabela III sistematiza a distribuição dos artigos selecionados por periódico científico, permitindo identificar os veículos de publicação mais relevantes na área de monitoramento de saúde baseado em IoT, sendo todos do tipo jornal. A análise evidencia uma predominância de publicações em periódicos de alto impacto, com destaque para aqueles indexados nas editoras IEEE, Springer, MDPI e Elsevier.

Periódico	Nº	Artigos
IEEE Access	6	Jain et al. (2024); Kumar and Rajaram (2024); Vaiyapuri et al. (2024); Manogaran et al. (2021); Rivadeneira et al. (2024); Al-Rakhami et al. (2021).
Sensors (MDPI)	4	Tahir et al. (2022); Sravanthi et al. (2024); Nekui et al. (2024); Almujaid et al. (2023).
Multimedia Tools and Applications	3	Kaur and Sharma (2024); Shoukat et al. (2024); Yadav et al. (2022).
IEEE Internet of Things Journal	3	Xu et al. (2023); Qian et al. (2022); Zhang et al. (2023).

Fonte: Elaboração própria (2025).

Tabela 3 – Periódicos e artigos analisados

Constata-se que os periódicos IEEE concentram o maior número de publicações, totalizando 13 artigos. O IEEE Access lidera com 6 publicações, seguido pelo IEEE Internet of Things Journal (3 artigos) e IEEE Sensors Journal (2 artigos). Completam este grupo o IEEE Journal of Biomedical and Health Informatics e o IEEE Transactions on Instrumentation and Measurement, cada um com uma publicação.

Além dos periódicos IEEE, destacam-se as contribuições da MDPI (4 artigos), Springer (3) e Elsevier (3), bem como de periódicos com duas publicações cada, como o Computational Intelligence and Neuroscience [Siam et al. 2021, Asif et al. 2022] e o IEEE Sensors Journal [Abubeker et al. 2024, Yashudas et al. 2024]. A produção é complementada por publicações únicas em diversos outros veículos especializados de alto impacto [Kumar and Krishnamoorthi 2021, Javed et al. 2021, Prasad et al. 2025, Verma et al. 2022, Yang 2023, Arunachalam et al. 2023, Manogaran et al. 2021, Astillo et al. 2022, Escobedo et al. 2024, Umar et al. 2025], o que evidencia o caráter multidisciplinar da área.

Esta distribuição demonstra não apenas o caráter multidisciplinar da pesquisa em aplicações de IoT na área da saúde, mas também a consistência na qualidade das publicações, que se concentram em periódicos bem estabelecidos e com rigorosos processos de revisão por pares.

4.4.RQ4: Como os sistemas embarcados, quando utilizados em residências, podem contribuir com a qualidade de vida e promover o bem-estar dos seus usuários?

Sistemas embarcados viabilizam a saúde domiciliar proativa, utilizando sensores vestíveis [Kumar and Krishnamoorthi 2021] e sem contato [Prasad et al. 2025] para monitorar sinais vitais e identificar riscos precocemente. Modelos de IA analisam os dados para prever condições críticas [Alkanhel et al. 2024], acionando respostas imediatas como alertas [Siam et al. 2021] ou intervenções automáticas [Astillo et al. 2022].

A implementação em larga escala, contudo, enfrenta desafios de segurança dos dados [Yadav et al. 2022] e da necessidade de datasets de qualidade para o treinamento dos algoritmos [Kaur and Sharma 2024].

4.5.RQ5: Como os sistemas embarcados podem ser utilizados para prever fatalidades em pacientes críticos através do monitoramento dos sinais vitais?

Sistemas embarcados são utilizados para prever fatalidades em pacientes críticos através de uma infraestrutura de múltiplos sensores que realiza o monitoramento contínuo de sinais vitais [Prasad et al. 2025]. Utilizando sensores vestíveis para captar dados como padrões de movimento [Kumar and Krishnamoorthi 2021] e atividade cardíaca [Asif et al. 2022], esses sistemas alimentam modelos de Inteligência Artificial (IA) projetados para identificar correlações que antecedem uma crise. A IA analisa os dados para prever a agravação de doenças crônicas [Alkanhel et al. 2024] e eventos cardiovasculares [Umar et al. 2025], transformando o cuidado reativo em preditivo. A análise pode ocorrer na borda (Edge Computing) para respostas rápidas [Kumar and Rajaram 2024], acionando alertas imediatos [Siam et al. 2021] que criam uma janela de oportunidade para intervenções médicas e, consequentemente, para salvar vidas.

4.6.RQ6: Como os sistemas embarcados integrados a sensores vitais podem identificar ocorrências antes de eventos fatais?

Para prever fatalidades em pacientes críticos, sistemas embarcados utilizam múltiplos sensores que monitoram continuamente sinais vitais [Prasad et al. 2025], como a

atividade cardíaca [Asif et al. 2022] e movimentos [Kumar and Krishnamoorthi 2021]. A chave é o processamento com IA, que ocorre tanto na borda (Edge Computing) para agilizar respostas [Kumar and Rajaram 2024], quanto na nuvem, onde são treinados modelos para prever eventos cardiovasculares [Umar et al. 2025]. Ao detectar padrões que antecedem crises, o sistema muda o cuidado de reativo para preditivo, emitindo alertas que possibilitam uma intervenção proativa [Siam et al. 2021]. Com a devida segurança dos dados [Yadav et al. 2022], essa integração de sensores e análise preditiva melhora o cuidado e abre uma janela de tempo crucial para salvar vidas.

5. Discussão

A análise dos 30 artigos selecionados revela um panorama claro: a convergência entre Sistemas Embarcados e IoT está catalisando uma transformação fundamental na prestação de cuidados de saúde, migrando de um modelo tradicionalmente reativo para uma abordagem proativa e preditiva. Esta seção discute as implicações dos resultados apresentados, contextualizando as tendências e identificando as lacunas críticas que direcionam futuras pesquisas.

O crescente volume de publicações, com um pico acentuado em 2024, não é um fenômeno isolado. Ele reflete o amadurecimento das tecnologias de sensoriamento e conectividade, aliado a uma demanda global por soluções de monitoramento remoto, intensificada no período pós-pandêmico. A liderança de países como Índia e Arábia Saudita na produção científica sugere uma correlação entre desafios de saúde pública em larga escala – como a alta prevalência de doenças crônicas – e investimentos estratégicos em tecnologia como resposta.

O pilar desta revolução é a fusão de dados provenientes de um ecossistema diversificado de sensores com a capacidade analítica da Inteligência Artificial. Os artigos demonstram que a força do monitoramento moderno não reside em um único dispositivo, mas na combinação de dados de sensores vestíveis [Kumar and Krishnamoorthi 2021, Kaur and Sharma 2024], sistemas sem contato [Escobedo et al. 2024] e sensores ambientais integrados a casas inteligentes [Javed et al. 2021, Shoukat et al. 2024]. Essa abordagem multimodal cria um panorama holístico e contextualizado do paciente. Por exemplo, a IA pode correlacionar dados de frequência cardíaca de um relógio inteligente com padrões de movimento detectados por sensores no ambiente para prever o risco de uma queda com muito mais precisão [Al-Rakhmi et al. 2021, Qian et al. 2022] do que cada sensor isoladamente faria. Este é o cerne da medicina preditiva: transformar dados brutos em insights acionáveis que podem salvar vidas, permitindo a detecção precoce de riscos [Prasad et al. 2025] e a previsão de fatalidades em pacientes críticos [Almujally et al. 2023].

Contudo, a sofisticação tecnológica revelada nos estudos contrasta com desafios práticos que ainda barram a sua adoção em larga escala. A segurança e a privacidade dos dados são consistentemente citadas como barreiras críticas. Esta preocupação é ecoada na literatura publicada na própria RISTI, onde, por exemplo, Ochoa et al. (2022) investigaram a proteção de dados em IoT, e Kinjo et al. (2022) focaram especificamente na probabilidade de falha em sistemas de saúde IoT, um fator crítico para a confiança do usuário. Da mesma forma, os desafios de interoperabilidade e usabilidade são centrais para a criação de ‘ambientes de vida melhorados’, um conceito também explorado por Marques & Pitarma (2019) no contexto do Health-IoT. A transmissão

continua de dados de saúde altamente sensíveis exige arquiteturas robustas que ainda não estão padronizadas. Além disso, a eficácia dos modelos de IA depende diretamente da disponibilidade de datasets de alta qualidade, que, como aponta a literatura, ainda são escassos [Kaur and Sharma 2024]. Essa escassez não apenas dificulta o treinamento de algoritmos precisos, mas também impede a validação e comparação de diferentes soluções propostas na academia.

Uma lacuna notável, e talvez a mais crítica para a aplicação no mundo real, é a limitada discussão sobre **usabilidade e aceitação pelo usuário final**. Apenas um dos artigos selecionados [Rivadeneira et al. 2024] menciona explicitamente a avaliação da usabilidade como um passo fundamental. De pouco adianta desenvolver um sistema de monitoramento altamente preciso se o público-alvo, frequentemente idoso, o considera intrusivo, complexo de usar ou simplesmente não confia nele. Futuras pesquisas devem, portanto, transcender a prova de conceito tecnológica e incorporar rigorosos estudos de Interação Humano-Computador (IHC) para garantir que as soluções sejam não apenas eficazes, mas também desejáveis e de fácil adesão.

Finalmente, os trabalhos futuros precisam focar na **integração com os sistemas de saúde existentes**. Como o fluxo contínuo de dados gerados por esses dispositivos será integrado aos prontuários eletrônicos dos pacientes? Como os profissionais de saúde podem utilizar essa avalanche de informações sem sofrer de sobrecarga informacional? Responder a essas questões é crucial para que o potencial transformador da IoT na saúde se concretize, saindo dos laboratórios de pesquisa para se tornar um pilar confiável da medicina do futuro.

6. Conclusão

Esta revisão de 30 artigos confirma que a aplicação de IoT na saúde é um campo em ascensão, com a produção científica concentrada na Ásia e publicada em periódicos de alto impacto. A principal contribuição é a transformação do cuidado de saúde para um modelo proativo e preditivo, viabilizado pela fusão de um ecossistema de sensores vestíveis [Kumar and Krishnamoorthi 2021, Kaur and Sharma 2024] e sem contato [Escobedo et al. 2024] com algoritmos de Inteligência Artificial [Asif et al. 2022, Zhang et al. 2023]. As aplicações práticas são vastas, incluindo a previsão de fatalidades em pacientes críticos [Almujally et al. 2023] e a criação de ambientes inteligentes que promovem autonomia e bem-estar [Javed et al. 2021, Shoukat et al. 2024], melhorando a qualidade do cuidado através da detecção precoce de riscos [Prasad et al. 2025].

Apesar dos avanços, a implementação em larga escala enfrenta desafios significativos de segurança e privacidade dos dados [Siam et al. 2021, Yadav et al. 2022], além da escassez de datasets de alta qualidade para o treinamento de modelos de IA [Kaur and Sharma 2024]. Para superar esses obstáculos, **os trabalhos futuros devem focar em direções mais específicas:**

- **No âmbito da segurança**, em vez de arquiteturas genéricas, a pesquisa deve se concentrar no desenvolvimento de **protocolos de criptografia leve (lightweight cryptography)**, otimizados para dispositivos embarcados com

baixo poder computacional, e em **arquiteturas de aprendizado federado** que permitam o treinamento de modelos de IA sem a necessidade de centralizar dados sensíveis dos pacientes.

- Para combater a **escassez de datasets**, sugere-se a investigação de **técnicas de geração de dados sintéticos (synthetic data generation)**, como Redes Adversariais Generativas (GANs), para criar conjuntos de dados realistas e anonimizados que possam ser compartilhados publicamente para acelerar a pesquisa.
- No campo da **usabilidade e aceitação** [Rivadeneira et al. 2024], os estudos devem evoluir de avaliações pontuais para **ensaios longitudinais (long-term studies)**, acompanhando pacientes por meses para entender os fatores que levam à adesão ou ao abandono da tecnologia no dia a dia.
- Finalmente, é crucial investigar a **interoperabilidade com sistemas de saúde legados**, propondo modelos para integrar os dados de IoT com os Prontuários Eletrônicos do Paciente (PEP) através de padrões como o HL7 FHIR, garantindo que as informações geradas se traduzam em valor clínico real.

Esses passos são fundamentais para consolidar o potencial dos sistemas embarcados e da IoT como pilares da medicina do futuro.

Referências

- Abubeker, K., Baskar, S., and Yadav, P. (2024). Internet-of-Things-Assisted Wire-less Body Area Network-Enabled Biosensor Framework for Detecting Ventilator and Hospital-Acquired Pneumonia. *IEEE Sensors Journal*, 24(7):11354–11361.
- Al-Rakhami, M. S., Gumaei, A., Altaf, M., Hassan, M. M., Alkhamees, B. F., Muham-mad, K., and Fortino, G. (2021). FallDeF5: A fall detection framework using 5G-based deep gated recurrent unit networks. *IEEE Access*, 9:94299–94308.
- Alkanhel, R. I., Saleh, H., Elaraby, A., Alharbi, S., Elmannai, H., Alaklabi, S., Alsamhi, S. H., and Mostafa, S. (2024). Hybrid cnn-gru model for real-time blood glucose forecasting: Enhancing iot-based diabetes management with ai. *Sensors*, 24(23):7670.
- Almujally, N. A., Aljrees, T., Saidani, O., Umer, M., Faheem, Z. B., Abuzinadah, N., Alnowaiser, K., and Ashraf, I. (2023). Monitoring acute heart failure patients using internet-of-things-based smart monitoring system. *Sensors*, 23(10):4580.
- Arunachalam, R., Sunitha, G., Shukla, S. K., Pandey, S. N., Urooj, S., and Rawat, S. (2023). A smart Alzheimer's patient monitoring system with IoT-assisted technology through enhanced deep learning approach. *Knowledge and Information Systems*, 65(12):5561–5599.
- Asif, R. N., Abbas, S., Khan, M. A., Sultan, K., Mahmud, M., and Mosavi, A. (2022). Development and validation of embedded device for electrocardiogram arrhythmia empowered with transfer learning. *Computational Intelligence and Neuroscience*, 2022(1):5054641.

- Astillo, P. V., Duguma, D. G., Park, H., Kim, J., Kim, B., and You, I. (2022). Federated intelligence of anomaly detection agent in IoTMD-enabled Diabetes Management Control System. *Future Generation Computer Systems*, 128:395–405.
- Escobedo, P., Pousibet-Garrido, A., López-Ruiz, N., Carvajal, M. A., Palma, A. J., and Martínez-Olmos, A. (2024). Bed-based ballistocardiography system using flexible RFID sensors for noninvasive single-and dual-subject vital signs monitoring. *IEEE Transactions on Instrumentation and Measurement*, 73:1–12.
- Jain, P., Joshi, A. M., Mohanty, S. P., and Cenkeramaddi, L. R. (2024). Non-invasive glucose measurement technologies: Recent advancements and future challenges. *IEEE Access*, 12:61907–61936.
- Javed, A. R., Fahad, L. G., Farhan, A. A., Abbas, S., Srivastava, G., Parizi, R. M., and Khan, M. S. (2021). Automated cognitive health assessment in smart homes using machine learning. *Sustainable Cities and Society*, 65:102572.
- Kaur, R. and Sharma, R. (2024). Wearable sensors and datasets for evaluating systems predicting falls and activities of daily living: recent advances and methodology. *Multimedia Tools and Applications*, 83(29):73671–73701.
- Kitchenham, B. and Charters, S. (2007). Guidelines for performing Systematic Literature Reviews in Software Engineering. Technical Report EBSE-2007-01, School of Computer Science and Mathematics, Keele University and Department of Computer Science, University of Durham, Keele, UK and Durham, UK.
- Kumar, R. H. and Rajaram, B. (2024). Design and simulation of an edge compute architecture for IoT-based clinical decision support system. *IEEE Access*, 12:45456–45474.
- Kumar, V. S. and Krishnamoorthi, C. (2021). Development of electrical transduction based wearable tactile sensors for human vital signs monitor: Fundamentals, methodologies and applications. *Sensors and Actuators A: Physical*, 321:112582.
- Manogaran, G., Alazab, M., Song, H., and Kumar, N. (2021). CDP-UA: Cognitive data processing method wearable sensor data uncertainty analysis in the internet of things assisted smart medical healthcare systems. *IEEE Journal of Biomedical and Health Informatics*, 25(10):3691–3699.
- Nekui, O. D., Wang, W., Liu, C., Wang, Z., and Ding, B. (2024). IoT-based heartbeat rate-monitoring device powered by harvested kinetic energy. *Sensors*, 24(13):4249.
- Prasad, P., Hussain, S. A., Thotakura, P., and Sanki, P. K. (2025). Design and Development of an IoT-Based Embedded System for Continuous Monitoring of Vital Signs. *Journal of Electronic Materials*, 54(5):3444–3451.
- Qian, Z., Lin, Y., Jing, W., Ma, Z., Liu, H., Yin, R., Li, Z., Bi, Z., and Zhang, W. (2022). Development of a real-time wearable fall detection system in the context of Internet of Things. *IEEE internet of things journal*, 9(21):21999–22007.

- Rivadeneira, J. E., Fernandes, J. M., Rodrigues, A., Boavida, F., and Silva, J. S. (2024). An Evaluation of Unobtrusive Sensing in a Healthcare Case Study. *IEEE Access*, 12:89405–89417.
- Shoukat, M. U., Yan, L., Zhang, J., Cheng, Y., Raza, M. U., and Niaz, A. (2024). Smart home for enhanced healthcare: exploring human machine interface oriented digital twin model. *Multimedia Tools and Applications*, 83(11):31297–31315.
- Siam, A. I., Almaiah, M. A., Al-Zahrani, A., Elazm, A. A., El Banby, G. M., El-Shafai, W., El-Samie, F. E. A., and El-Bahnasawy, N. A. (2021). Secure health monitoring communication systems based on IoT and cloud computing for medical emergency applications. *Computational Intelligence and Neuroscience*, 2021(1):8016525.
- Sravanthi, M., Gunturi, S. K., Chinnaiah, M. C., Lam, S.-K., Vani, G. D., Basha, M., Janardhan, N., Krishna, D. H., and Dubey, S. (2024). Adaptive FPGA-Based Accelerators for Human–Robot Interaction in Indoor Environments. *Sensors (Basel, Switzerland)*, 24(21):6986.
- Tahir, S. B. u. d., Dogar, A. B., Fatima, R., Yasin, A., Shafiq, M., Khan, J. A., Assam, M., Mohamed, A., and Attia, E.-A. (2022). Stochastic recognition of human physical activities via augmented feature descriptors and random forest model. *Sensors*, 22(17):6632.
- Umar, M. A., Abuali, N., Shuaib, K., and Awad, A. I. (2025). An explainable artificial intelligence and Internet of Things framework for monitoring and predicting cardiovascular disease. *Engineering Applications of Artificial Intelligence*, 144:110138.
- Vaiyapuri, T., Alharbi, G., Dharmarajlu, S. M., Bouteraa, Y., Misra, S., Ramesh, J. V. N., and Mohanty, S. N. (2024). IoT-enabled early detection of diabetes diseases using deep learning and dimensionality reduction techniques. *IEEE Access*.
- Verma, D., Singh, K. R., Yadav, A. K., Nayak, V., Singh, J., Solanki, P. R., and Singh, R. P. (2022). Internet of things (IoT) in nano-integrated wearable biosensor devices for healthcare applications. *Biosensors and Bioelectronics: X*, 11:100153.
- Xu, W., Hong, L., Zheng, J., Li, M., Hua, Y., and Zhao, X. (2023). Wearable smart sensor system for monitoring and intelligent prediction of sodium ions in human perspiration. *IEEE Internet of Things Journal*, 11(5):8146–8155.
- Yadav, K., Alharbi, A., Jain, A., and Ramadan, R. A. (2022). An IoT based secure patient health monitoring system. *Computers, Materials and Continua*, 70(2):3637–3652.
- Yang, L. (2023). Data monitoring for a physical health system of elderly people using smart sensing technology. *Wireless Networks*, 29(8):3665–3678.
- Yashudas, A., Gupta, D., Prashant, G., Dua, A., AlQahtani, D., and Reddy, A. S. K. (2024). Deep-cardio: Recommendation system for cardiovascular disease prediction using iot network. *IEEE Sensors Journal*, 24(9):14539–14547.

- Zhang, D., Liu, X., Xia, J., Gao, Z., Zhang, H., and de Albuquerque, V. H. C. (2023). A physics-guided deep learning approach for functional assessment of cardiovascular disease in IoT-based smart health. *IEEE Internet of Things Journal*, 10(21): 18505–18516.
- Kinjo, E. M., Librantz, A. F. H., & Martins, F. S. (2022). Modelagem Bayesiana aplicada para cálculo da probabilidade de falha em Sistemas de Saúde IoT. *RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação*, (47), 87–101. <https://scielo.pt/pdf/rist/n47/1646-9895-rist-47-87.pdf>
- Marques, G., & Pitarma, R. (2019). Saúde ocupacional e ambientes de vida melhorados com recurso à Internet das Coisas. *RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação*, (E19), 1–13. https://www.researchgate.net/publication/335653524_Occupational_health_and_enhanced_living_environments_through_Internet_of_Things
- Ochoa-Díaz, K. M., Cárdenas-Suárez, L. S., & Rojas, L. C. P. (2022). Protección de datos personales en el internet de las cosas (IoT) para el sector salud en Colombia. *RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação*, (E51), 33–46. <https://www.google.com/search?q=https://doi.org/10.17013/risti.e51.33-46>

Implementação da Solução de Gestão de Arquitetura de TI: Estudo em uma Rede Farmacêutica Regional

Murilo Albeest¹, Rodolfo Miranda de Barros²

murilo.albeest@gmail.com; rodolfo@uel.br

¹ Departamento de Ciência da Computação, Universidade Estadual de Londrina (UEL), Londrina-PR, Brasil

² Departamento de Ciência da Computação, Universidade Estadual de Londrina (UEL), Londrina-PR, Brasil

Pages: 69-82

Resumo: Este estudo examina a aplicação da prática de Gestão de Arquitetura do ITIL em uma rede de farmácias regional, Brasil, com foco no alinhamento da Tecnologia da Informação aos objetivos estratégicos organizacionais. O mapeamento da arquitetura identificou redundâncias e lacunas, cujo tratamento, por meio da adoção da prática, resultou em maior alinhamento estratégico, redução de custos, aprimoramento da integração de dados e fortalecimento da governança corporativa. Embora tenham sido enfrentados desafios, como a resistência cultural, constatou-se que a Gestão de Arquitetura contribui para a padronização, escalabilidade e estímulo à inovação, ampliando a competitividade da organização no mercado.

Palavras-chave: itil, gerenciamento de arquitetura, tecnologia.

Implementation of IT Architecture Management Solution: Study in a Regional Pharmaceutical Network

Abstract: This study examines the application of the ITIL Architecture Management practice in a regional pharmacy, Brazil, focusing on aligning Information Technology with the organization's strategic objectives. The architecture mapping identified redundancies and gaps, whose resolution, through the adoption of the practice, resulted in greater strategic alignment, cost reduction, enhanced data integration, and strengthened corporate governance. Although challenges such as cultural resistance were encountered, it was found that Architecture Management contributes to standardization, scalability, and the promotion of innovation, thereby enhancing the organization's competitiveness in the market.

Keywords: itil, architecture management, technology.

1. Introdução

A crescente complexidade dos ambientes empresariais, impulsionada pela transformação digital e pela necessidade de integrar tecnologias emergentes aos

processos de negócio, impõe novos desafios à gestão de tecnologia da informação (TI). No varejo farmacêutico, esse cenário é ainda mais evidente devido à diversidade de serviços, à expansão digital e à pressão por eficiência operacional e inovação. Conforme Laudon e Laudon (2022), a eficiência na administração dos serviços de TI tornou-se fator estratégico essencial para assegurar competitividade e sustentabilidade organizacional.

O Information Technology Infrastructure Library (ITIL), reconhecido globalmente como um conjunto de boas práticas para gerenciamento de serviços de TI, fornece diretrizes voltadas à geração de valor para os stakeholders. Sua versão mais recente, o ITIL 4, introduz princípios modernos como pensamento sistêmico, práticas ágeis e foco na cocriação de valor, além de incluir práticas específicas como o Gerenciamento de Arquitetura (Architecture Management), cuja função é estabelecer uma visão integrada da estrutura organizacional de TI, promovendo padronização, conformidade e apoio à tomada de decisão estratégica (AXELOS, 2019).

Apesar de sua relevância, a aplicação prática do gerenciamento de arquitetura em organizações de médio porte ainda é pouco explorada na literatura, especialmente no contexto latino-americano. Essa lacuna limita a compreensão de como essa prática pode contribuir para alinhar objetivos estratégicos e melhorar a eficiência operacional em empresas com recursos e estruturas menos complexas que grandes corporações. A adoção estruturada de práticas do ITIL 4 tende a melhorar a qualidade dos serviços, otimizar recursos e ampliar a entrega de valor às partes interessadas (GALIAN et al., 2025).

Assim, este estudo tem como objetivo analisar os efeitos da adoção da prática de Gerenciamento de Arquitetura conforme o ITIL 4 em uma rede farmacêutica regional, destacando seus benefícios, limitações e desafios enfrentados durante a implementação. A pergunta de pesquisa que orienta este trabalho é:

De que maneira a adoção do Gerenciamento de Arquitetura baseado no ITIL 4 pode contribuir para o alinhamento estratégico e a melhoria da eficiência operacional em uma rede farmacêutica regional?

Para responder a essa questão, adota-se uma abordagem qualitativa e exploratória, com estudo de caso baseado em observação direta, análise documental e entrevistas com gestores estratégicos. O trabalho busca contribuir tanto para a prática profissional quanto para o avanço da literatura, oferecendo um relato aplicado e reflexivo sobre a adoção do ITIL 4 em organizações de médio porte.

Para orientar o leitor, este artigo está estruturado da seguinte forma: a Seção 2 apresenta os estudos relacionados e o referencial teórico que fundamenta esta pesquisa. Na Seção 3, é apresentada uma visão geral do ITIL 4 e do Gerenciamento de Arquitetura, aprofundada na Seção 4 com a descrição de suas atividades-chave. A Seção 5 detalha a metodologia adotada, enquanto a Seção 6 expõe o estudo de caso realizado. Os resultados obtidos são discutidos na Seção 7. A Seção 8 aborda as ameaças à validade e, por fim, a Seção 9 apresenta as conclusões e propostas para trabalhos futuros.

2. Trabalhos relacionados

O Gerenciamento de Arquitetura tem ganhado destaque como componente estratégico da governança de TI, sendo abordado em frameworks como TOGAF (The Open Group, 2018), ArchiMate (The Open Group, 2019) e, mais recentemente, no ITIL 4 (Axelos, 2019). Apesar de reconhecido conceitualmente, sua aplicação prática ainda apresenta lacunas na literatura, especialmente em estudos empíricos voltados a organizações de médio porte, como redes regionais de varejo e saúde.

Bernard (2012) define a Arquitetura Corporativa como estrutura para alinhar sistemas de informação aos objetivos de negócio. Weill e Ross (2004) demonstram que uma governança eficiente da arquitetura, sustentada por diretrizes claras e padronização tecnológica, impacta positivamente o desempenho organizacional.

No contexto do ITIL, Steinberg (2012) e Galup et al. (2009) identificam ganhos em qualidade de serviço, conformidade e eficiência operacional, embora a maioria dos estudos se concentre em grandes empresas ou setores altamente regulamentados. Pereira e Souza (2017) reforçam a importância de mapear processos e fluxos de informação de forma padronizada, enquanto o Modelo de Referência de Arquitetura Corporativa – MR-AC (Ministério do Planejamento, 2014) no setor público brasileiro evidencia o avanço de práticas estruturadas de governança.

Estudos recentes (Alves & Ribeiro, 2023; Cruz et al., 2022) reforçam a importância de frameworks adaptativos de governança de TI em PMEs, indicando que a flexibilidade metodológica é um fator crítico para o sucesso. Essa perspectiva sustenta a abordagem adotada neste estudo, que combina diretrizes do ITIL 4 com práticas simplificadas para promover adoção em contextos de recursos limitados.

Apesar desses progressos, são raros os trabalhos que descrevem, com rigor metodológico, a aplicação do Gerenciamento de Arquitetura conforme o ITIL 4 em empresas de médio porte. Este estudo busca preencher essa lacuna, oferecendo uma análise aplicada dos benefícios e desafios vivenciados por uma organização do setor farmacêutico.

3. Visão Geral do ITIL

O ITIL (Biblioteca de Infraestrutura de Tecnologia da Informação) é um conjunto de boas práticas criado nos anos 1980 pelo governo britânico para orientar a gestão de serviços de TI. Amplamente adotado no mundo, evoluiu ao longo do tempo para acompanhar mudanças tecnológicas e de mercado. A versão mais recente, ITIL 4 (2019), traz uma abordagem integrada e flexível, alinhada à transformação digital, agilidade e DevOps. No Sistema de Valor de Serviço do ITIL, o Gerenciamento de Arquitetura destaca-se por criar uma visão padronizada da TI, garantindo que sistemas, tecnologias e processos estejam alinhados aos objetivos estratégicos, promovendo coerência, flexibilidade e escalabilidade sem comprometer a qualidade dos serviços.

3.1. O que é gerenciamento de arquitetura no ITIL

No ITIL 4, o Gerenciamento de Arquitetura contempla diferentes perspectivas que, em conjunto, garantem que os serviços de TI estejam alinhados à estratégia organizacional.

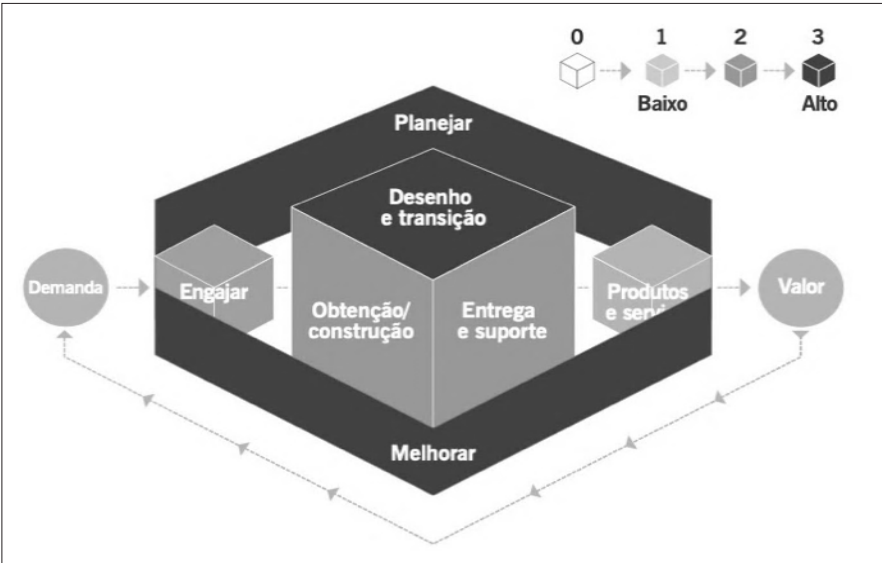


Figura 1 – Contribuição do Gerenciamento de Arquitetura para atividades. Fonte: Axelos, 2019.

Essas visões apoiam a governança, a integração entre áreas e a inovação contínua. São cinco tipos principais:

Arquitetura de Negócio – Relaciona as capacidades atuais da organização à sua visão de futuro, identificando lacunas e definindo planos de transformação graduais. Essa análise conecta habilidades, processos e atividades que agregam valor, orientando mudanças de forma controlada e coerente com os objetivos estratégicos.

Arquitetura de Serviço – Apresenta um panorama dos serviços oferecidos e de suas interações internas, descrevendo composição, fluxo e dependências. Essa visão favorece a gestão do portfólio, a padronização de processos e a melhoria contínua, facilitando a entrega de valor de forma consistente.

Arquitetura de Sistemas de Informação – Engloba dados e aplicações, definindo como são estruturados, integrados e protegidos. Garante integridade, precisão e disponibilidade das informações, permitindo que sejam utilizadas de forma estratégica e segura para suportar decisões.

Arquitetura de Tecnologia – Abrange a infraestrutura física e lógica, incluindo hardware, software, protocolos e plataformas. Busca desempenho confiável, segurança robusta e escalabilidade para acompanhar necessidades futuras, sustentando os serviços de TI ao longo do tempo.

Arquitetura Ambiental – Considera elementos externos que afetam a organização, como regulamentações, avanços tecnológicos, tendências de mercado e fatores socioeconômicos. Essa análise possibilita antecipar desafios e adaptar estratégias para responder a um ambiente em constante mudança.

4. Atividades-Chave do Gerenciamento de Arquitetura no ITIL

O Gerenciamento de Arquitetura no ITIL 4 compreende um conjunto de atividades essenciais para assegurar que as soluções tecnológicas e processuais estejam alinhadas com os objetivos estratégicos da organização. De acordo com AXELOS (2019), tais atividades são fundamentais para estabelecer e manter uma arquitetura consistente, flexível e responsiva às mudanças no cenário empresarial. Entre as principais ações, destacam-se: análise e desenho de arquiteturas, formulação de visões arquiteturais, definição de padrões e diretrizes, documentação e comunicação eficaz e manutenção evolutiva da arquitetura. Complementarmente, frameworks como o TOGAF (The Open Group, 2018) e o ArchiMate oferecem metodologias e notações estruturadas para apoiar essas tarefas.

A análise e o desenho arquitetural constituem a base para compreender o estado atual da infraestrutura e dos processos organizacionais. Essa etapa envolve a coleta sistemática de informações sobre sistemas, tecnologias e fluxos de negócio, com o objetivo de criar representações claras e acuradas da arquitetura vigente (Bernard, 2012). Essa visão permite identificar lacunas, redundâncias e oportunidades de melhoria. A integração com frameworks de governança, como o COBIT (ISACA, 2018), contribui para garantir que as decisões arquiteturais estejam alinhadas a objetivos corporativos e práticas consolidadas de controle e conformidade.

O planejamento e a modelagem arquitetural devem considerar simultaneamente requisitos técnicos, demandas de negócio e tendências de mercado, assegurando escalabilidade e adaptabilidade. Modelos visuais, quando estruturados em notações como BPMN para processos e ArchiMate para componentes de TI, facilitam a análise de arquiteturas complexas e aprimoram a comunicação com diferentes stakeholders. Conforme salientam Weill e Ross (2004), arquiteturas versáteis fortalecem a capacidade da organização de responder a mudanças no ambiente competitivo.

A definição de padrões e diretrizes representa uma dimensão crítica da governança arquitetural. O estabelecimento de políticas claras, metodologias consistentes e processos de aprovação formais garante a coerência das soluções implementadas. A padronização previne abordagens isoladas, favorece a interoperabilidade e potencializa a inovação em ambientes de TI heterogêneos.

A documentação estruturada da arquitetura, mantida de forma atualizada e acessível, é indispensável para a aplicação consistente das diretrizes definidas. Relatórios, manuais e apresentações técnicas devem traduzir a visão arquitetural de modo compreensível para públicos distintos, desde executivos até equipes operacionais. A comunicação eficaz por meio de workshops e reuniões periódicas assegura alinhamento contínuo e comprometimento das partes interessadas.

Por fim, a manutenção evolutiva da arquitetura é necessária para acompanhar transformações tecnológicas e mudanças estratégicas. Essa atualização contínua mantém a arquitetura aderente às necessidades de negócio e garante que a TI atue como facilitadora de valor organizacional. O uso combinado de frameworks consolidados e ferramentas de modelagem integradas amplia a efetividade das práticas de Gerenciamento de Arquitetura no ITIL, contribuindo para arquiteturas mais robustas, adaptáveis e alinhadas à estratégia corporativa.

4.1. Benefícios e Desafios do Gerenciamento de Arquitetura

O Gerenciamento de Arquitetura, no contexto do ITIL 4, oferece benefícios estratégicos e operacionais significativos para as organizações, ao mesmo tempo em que impõe desafios que demandam gestão estruturada.

Entre os principais benefícios, destaca-se o fortalecimento da integração entre Tecnologia da Informação e objetivos de negócio. Ao alinhar estratégias tecnológicas às metas corporativas, a arquitetura empresarial potencializa a entrega de valor e garante que investimentos em TI estejam diretamente vinculados à visão estratégica. Conforme ressaltam Weill e Ross (2004), esse alinhamento aprimora a comunicação entre áreas técnicas e executivas, viabilizando decisões mais embasadas e coerentes com o planejamento de longo prazo.

Outro ponto relevante é a redução de redundâncias e riscos. Arquiteturas bem estruturadas evitam a duplicação de esforços, promovem padronização tecnológica e favorecem o uso eficiente de recursos. De acordo com AXELOS (2019), essa abordagem reduz custos operacionais, minimiza riscos de segurança e mitiga problemas de integração, ao mesmo tempo em que aumenta a flexibilidade organizacional. Frameworks como o TOGAF (The Open Group, 2018) oferecem métodos para construir arquiteturas unificadas e adaptáveis, reforçando a resiliência frente a mudanças.

O gerenciamento arquitetural também apoia a tomada de decisão estratégica ao fornecer uma visão documentada e precisa do cenário tecnológico atual e das projeções futuras. Essa visibilidade facilita o planejamento de investimentos, reorganizações e inovações, permitindo que ações sejam fundamentadas em informações consistentes e integradas.

No entanto, a implementação do Gerenciamento de Arquitetura apresenta desafios relevantes. Um deles é a resistência à padronização, especialmente em equipes que percebem diretrizes uniformes como limitadoras da criatividade ou da autonomia. Superar essa barreira requer gestão de mudança eficaz e comunicação clara dos benefícios, conforme recomenda a ISACA (2018) em práticas de governança de TI.

Outro desafio é lidar com a complexidade organizacional, particularmente em empresas de grande porte ou estruturas descentralizadas. A diversidade de sistemas e processos pode dificultar a obtenção de uma visão arquitetural única, exigindo abordagem progressiva e colaborativa, sustentada por mecanismos de governança sólidos.

A maturidade da governança é, portanto, fator crítico para o sucesso. Estruturas robustas e processos bem definidos orientam decisões ao longo de ciclos de mudança, garantindo coerência com a estratégia corporativa. Na ausência dessa base, conforme alertam Weill

e Ross (2004), há risco de perda de eficiência operacional, fragilidade no monitoramento e comprometimento da geração de valor esperada pela TI.

5. Metodologia

Este estudo adota uma abordagem qualitativa, de natureza exploratória, por meio de um estudo de caso único realizado em uma rede farmacêutica regional localizada no sul do Brasil. O objetivo central é analisar de que forma a prática de Gerenciamento de Arquitetura, conforme preconizado pelo ITIL 4, pode contribuir para o alinhamento estratégico e para o fortalecimento da governança de tecnologia da informação (TI). A escolha por um estudo de caso único está alinhada à perspectiva de Yin (2018), que o recomenda para investigações aprofundadas de fenômenos contemporâneos em seu contexto real, especialmente quando há múltiplas fontes de evidência. Essa abordagem é reforçada por Eisenhardt e Graebner (2007), que destacam o potencial das entrevistas e observações diretas na construção de teoria a partir de dados empíricos.

A coleta de dados ocorreu de forma triangulada, contemplando três frentes principais. Primeiramente, foi conduzida uma análise documental abrangendo políticas internas, fluxos operacionais, sistemas em uso e registros de reuniões da área de TI, com o intuito de mapear o contexto organizacional e identificar diretrizes pré-existentes. Em seguida, realizaram-se entrevistas semiestruturadas com quatro representantes das áreas estratégicas — Tecnologia da Informação, Negócios, Operações e Inovação —, visando compreender a percepção sobre o papel da arquitetura de TI, as barreiras enfrentadas na integração entre sistemas e os critérios adotados para a priorização de investimentos tecnológicos. Por fim, foi realizada observação direta em reuniões e workshops voltados à estruturação da arquitetura, registrando decisões, debates e principais pontos de dor (*pain points*) relatados pelas áreas envolvidas.

O processo de análise seguiu uma lógica iterativa, iniciando-se pelo mapeamento do cenário atual (*as-is*), com a construção de diagramas, fluxos e mapas de sistemas que permitiram identificar redundâncias, lacunas e oportunidades de melhoria. A partir desse diagnóstico, elaborou-se uma visão de futuro (*to-be*) fundamentada nas diretrizes do ITIL 4, contemplando cenários arquiteturais alinhados aos objetivos estratégicos da organização. Em paralelo, foram classificados os impactos estratégicos, operacionais e técnicos das mudanças propostas, considerando critérios definidos em conjunto com os *stakeholders*. As representações arquiteturais desenvolvidas foram posteriormente validadas em encontros colaborativos mediados pela equipe de TI, assegurando o alinhamento com as expectativas e necessidades das áreas de negócio.

Para apoiar as atividades de modelagem e análise, utilizaram-se ferramentas como *Business Process Model and Notation* (BPMN), planilhas colaborativas e diagramas em ArchiMate, seguindo a estrutura proposta pelo *The Open Group Architecture Framework* (TOGAF, 2018). Cabe ressaltar que, por se tratar de um estudo de caso único, os resultados obtidos não são generalizáveis a outros contextos organizacionais, embora possam servir de base para pesquisas comparativas ou estudos replicados. Além disso, restrições relacionadas à confidencialidade limitaram a divulgação de determinados dados estratégicos, o que impôs uma delimitação à profundidade de alguns aspectos apresentados.

6. Estudo de Caso em uma Rede Farmacêutica Regional

A organização analisada é uma rede de farmácias com forte atuação regional e predominância de lojas físicas. Nos últimos anos, a empresa passou a expandir seus canais de atendimento e a oferecer novos serviços digitais, como e-commerce, vendas via aplicativos parceiros e programas de fidelidade, o que aumentou a complexidade da estrutura de tecnologia da informação (TI).

Diante desse cenário, a área de TI identificou a necessidade de aprimorar a governança tecnológica e alinhar os investimentos em sistemas com os objetivos estratégicos da organização. A prática de *Gerenciamento de Arquitetura*, conforme definida no ITIL 4, foi selecionada como pilar central da estratégia de transformação digital.

A iniciativa teve início com o mapeamento da arquitetura de negócios, dados, aplicações e tecnologias utilizadas pela empresa, tanto em processos internos quanto em integrações externas. Foram utilizadas ferramentas simples e acessíveis para representar a arquitetura, como mapas de sistemas, fluxogramas e planilhas colaborativas. Os diagramas foram adaptados a diferentes públicos – gestores, equipes técnicas e lideranças operacionais – promovendo uma visão compartilhada da estrutura tecnológica.

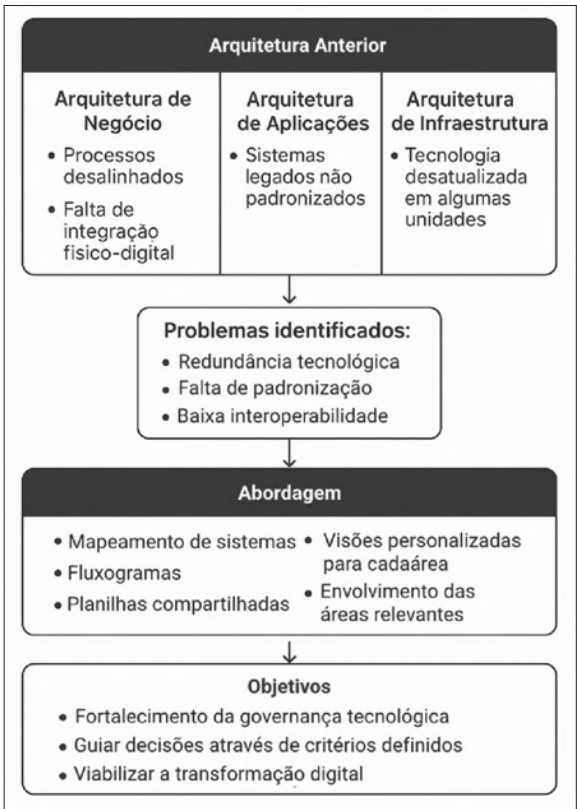


Figura 2 – Fluxo de alinhamento utilizado pela área de TI/Empresa.
Fonte: Os autores.

Esse trabalho resultou na identificação de sobreposições, riscos de integração e lacunas operacionais. A arquitetura passou a ser utilizada como base para apoiar decisões estratégicas e operacionais, integrando-se a outras práticas como Gerenciamento de Portfólio, Gestão de Mudanças e Planejamento Estratégico. No âmbito do Gerenciamento de Portfólio, a arquitetura permitiu avaliar os projetos em andamento e planejar futuros investimentos com foco na eliminação de redundâncias, priorização por retorno e aderência estratégica. Na Gestão de Mudanças, foi instituída uma avaliação de impacto arquitetônico antes da aprovação de alterações relevantes, reduzindo riscos técnicos e retrabalho. Já no Planejamento Estratégico, os mapas arquiteturais passaram a ser insumo para decisões sobre expansão, digitalização e integração de canais.

A Figura 2 apresenta o fluxo de alinhamento adotado, ilustrando a forma como a prática de Gerenciamento de Arquitetura passou a interagir com outras práticas de gestão, promovendo maior coerência entre a TI e os objetivos organizacionais.

As ações foram conduzidas com base nos princípios do Sistema de Valor de Serviço (SVS) do ITIL 4, com destaque para foco no valor, colaboração, visibilidade e melhoria contínua. A arquitetura deixou de ser um elemento técnico isolado e passou a ser tratada como ferramenta de apoio à governança organizacional.

7. Resultados Obtidos

A adoção da prática de *Gerenciamento de Arquitetura*, conforme o ITIL 4, gerou efeitos significativos na estrutura de tecnologia da informação da organização analisada. Os principais resultados obtidos ao longo de doze meses de implementação podem ser agrupados em cinco dimensões: alinhamento estratégico, redução de custos, melhoria na integração dos dados, governança reforçada e fortalecimento da arquitetura.

Os ganhos observados corroboram achados de Lunardi et al. (2017), que associam a padronização de arquiteturas e processos de TI à redução de custos e melhoria da qualidade de serviço. Essa evidência reforça que, mesmo em empresas de porte médio, a adoção disciplinada de frameworks consagrados pode gerar resultados tangíveis.

7.1. Alinhamento Estratégico e Evolução Arquitetural

A área de TI passou a participar de forma ativa no planejamento estratégico da organização, contribuindo para decisões relacionadas à expansão da rede, à integração de canais digitais (omnicanalidade) e à modernização dos processos. Houve aumento de 25% na taxa de aprovação de projetos em primeira submissão ao comitê estratégico, sinalizando maior alinhamento prévio. A arquitetura passou a servir como instrumento para avaliar o impacto de novas iniciativas e garantir coerência técnica entre os projetos.

7.2. Redução de Custos e Eliminação de Redundâncias

Por meio da padronização da arquitetura e da análise de portfólio, foram identificados cinco sistemas redundantes, posteriormente desativados. Isso resultou em uma redução estimada de 10% nos custos operacionais diretos relacionados à manutenção de sistemas. A substituição por plataformas corporativas centralizadas permitiu racionalizar recursos e melhorar o desempenho das soluções. Esse valor corresponde a aproximadamente R\$

100 mil/ano, segundo estimativas internas, liberando orçamento para investimentos em integração e segurança.

7.3. Integração de Sistemas e Serviços de Saúde

A aplicação de uma arquitetura baseada em interfaces de programação (APIs) viabilizou o compartilhamento de dados em tempo real entre os sistemas internos da empresa e plataformas externas, como marketplaces e parceiros logísticos. O tempo médio para integração de um novo parceiro reduziu-se de 90 para 40 dias, acelerando a entrada de novos serviços. No setor de saúde, isso se traduziu em maior agilidade na troca de informações com operadoras, laboratórios e sistemas públicos.

7.4. Fortalecimento da Governança e Conformidade

A governança de TI foi aprimorada com a adoção de padrões arquiteturais claros, os quais passaram a nortear a aquisição de novas tecnologias e a aprovação de projetos. Procedimentos de controle foram implementados no processo de gestão de mudanças, incluindo análise prévia de impacto arquitetônico, o que reduziu riscos de integração e retrabalho. A conformidade com políticas internas de TI subiu de 65% para 88%, segundo auditoria interna realizada após seis meses.

7.5. Melhoria do Desempenho Organizacional

A simplificação da infraestrutura tecnológica impactou diretamente a produtividade. Houve redução no tempo médio de atendimento a incidentes, maior disponibilidade de sistemas críticos e aumento da confiabilidade das bases de dados. Com isso, áreas operacionais passaram a tomar decisões com base em informações mais consistentes. A disponibilidade média dos sistemas críticos subiu de 97,2% para 99,1%, reduzindo paradas não planejadas e fortalecendo a experiência do cliente final.

A Figura 3 apresenta um gráfico de radar que compara o cenário anterior à implementação com o atual, considerando critérios como fortalecimento da arquitetura, integração, alinhamento, custo e governança. Os dados foram obtidos por meio de avaliação interna da equipe de TI e validados em reuniões com stakeholders estratégicos.

Apesar dos resultados positivos obtidos neste estudo, é importante reconhecer que a literatura sobre o ITIL frequentemente aponta dificuldades comuns relacionadas à resistência organizacional, complexidade metodológica e desafios culturais. Por exemplo, conforme Galup et al. [10], a resistência interna dos colaboradores e a dificuldade de adaptação à padronização de processos são barreiras frequentes em implementações do ITIL, especialmente em organizações menores ou com baixa maturidade tecnológica.

De maneira semelhante, Steinberg [9] enfatiza que o sucesso da arquitetura depende fortemente da comunicação clara e contínua entre áreas técnicas e de negócios, aspecto que também foi identificado como desafio no início deste estudo. No caso específico da rede farmacêutica analisada, tais dificuldades foram mitigadas por meio de estratégias colaborativas, capacitação interna e uso de ferramentas visuais simplificadas, reforçando a importância de uma abordagem gradual e participativa para a superação desses desafios.

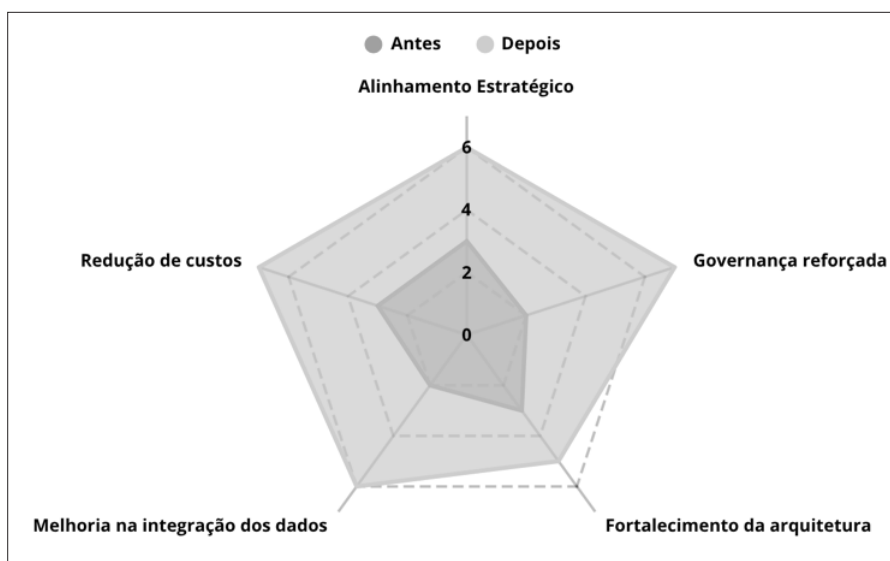


Figura 3 – Gráfico de radar representando os resultados obtidos.
Fonte: Os autores.

Embora o estudo tenha registrado melhorias significativas, é importante enfrentar os obstáculos frequentemente relatados na literatura. Em um estudo de múltiplos casos conduzido por Shang & Lin (2010), destacam-se barreiras como “conflitos entre processos padronizados e solicitações imprevistas”, “falta de autonomia” e “baixa capacidade integrativa” das equipes. Em consonância, Ibnu Abi Hujayr aponta resistência da equipe, conflitos operacionais e ordens ambíguas como elementos críticos que podem minar a implementação bem-sucedida do ITIL. No contexto de PMEs, a literatura revela ainda a ausência de formalização de TI e investimentos insuficientes em treinamento como fatores limitantes. No caso desta rede farmacêutica, embora alguns desses desafios tenham sido atenuados por meio de capacitação interna e comunicação transparente, a experiência reforça que abordagens estratégicas e gradativas são essenciais para mitigar tais riscos e alcançar uma transição mais eficaz.

8. Ameaças a validade

Como em qualquer estudo de caso, os resultados devem ser interpretados considerando algumas limitações. A validade interna pode ser influenciada por fatores externos, já que não houve grupos de controle, embora não tenham sido identificadas mudanças significativas que comprometessem a análise. A validade externa é restrita pelo foco em uma única rede farmacêutica regional, mas as conclusões podem inspirar abordagens semelhantes em contextos análogos. Quanto à validade por construção, os instrumentos de coleta foram qualitativos e não validados estatisticamente, porém combinados de forma a garantir coerência entre as evidências. Por fim, na validade por conclusão,

reconhece-se a possibilidade de viés de confirmação, minimizado por discussões com diferentes áreas e validação cruzada das informações.

9. Conclusões

Este estudo analisou os efeitos da adoção da prática de Gerenciamento de Arquitetura, conforme proposto pelo ITIL 4, em uma rede farmacêutica regional. A partir da aplicação estruturada dessa prática, foi possível observar ganhos relevantes em termos de alinhamento estratégico, redução de redundâncias tecnológicas, fortalecimento da governança e melhoria na integração de serviços e dados.

A arquitetura passou a exercer papel central na tomada de decisões, contribuindo para a padronização de sistemas, a eliminação de sobreposições e o direcionamento de investimentos com base em critérios técnicos e estratégicos. Os resultados apontam que, mesmo em organizações de médio porte, a aplicação disciplinada de práticas arquiteturais pode gerar impactos positivos tangíveis — não apenas em custos, mas também na maturidade digital e na capacidade de inovação.

Como contribuição prática, este trabalho fornece um relato aplicado que pode servir como referência para outras organizações que desejam iniciar a estruturação de sua arquitetura de TI com base no ITIL 4. Do ponto de vista científico, contribui ao explorar empiricamente uma prática ainda pouco documentada em empresas de médio porte no Brasil, especialmente no setor de varejo farmacêutico.

No entanto, o estudo apresenta limitações, como a ausência de instrumentos validados quantitativamente, a análise restrita a um único contexto organizacional e a dependência de dados autodeclarados. Recomenda-se que futuras pesquisas explorem abordagens comparativas entre organizações de diferentes segmentos e incorporem indicadores objetivos para avaliação de desempenho, além de explorar metodologias mistas (qualitativa e quantitativa).

Por fim, destaca-se que o Gerenciamento de Arquitetura, quando apoiado por governança eficaz e alinhamento estratégico, pode atuar como facilitador da transformação digital e da geração contínua de valor nos serviços de tecnologia da informação.

Referências

- AXELOS. (2019). *ITIL® Foundation: ITIL 4 edition*. London: The Stationery Office (TSO).
- Bernard, S. A. (2012). *An introduction to enterprise architecture* (3ª ed.). Bloomington: AuthorHouse.
- Brasil. Ministério da Economia. Secretaria de Governo Digital. (2020). *Modelo de referência de arquitetura corporativa para o Governo Federal – MR-AC* (versão 1.0). Brasília: Ministério da Economia.

- Cruz Hinojosa, E., & Gutierrez de Mesa, J. A. (2016). Barriers for ITSM adoption in SMEs: A literature review. *Journal of Information Systems and Technology Management*, 13(2), 235–253. <https://doi.org/10.4301/S1807-17752016000200007>
- Galup, S. D., Dattero, R., Quan, J. J., & Conger, S. (2009). An overview of IT service management. *Communications of the ACM*, 52(5), 124–127. <https://doi.org/10.1145/1506409.1506439>
- Ibnu Abi Hujayr. (n.d.). Lessons learned in ITIL implementation failure. *Academia.edu*. Recuperado de <https://www.academia.edu/>
- Laudon, K. C., & Laudon, J. P. (2020). *Management information systems: Managing the digital firm* (16ª ed.). Nova York: Pearson Education.
- Meléndez, K., Dávila, J., & Pessoa, F. (2016). IT service management adoption and implementation in SMEs: Challenges and critical success factors. *AMCIS Proceedings*, 1–10.
- Pereira, C. M., & Souza, C. A. S. de. (2010). Uma abordagem para modelagem de arquitetura de processos de negócio. *Revista de Sistemas de Informação da FSMA*, (6), 4–13.
- Shang, S., & Lin, S. F. (2010). Barriers to implementing ITIL – A multi-case study on the service industry. *Contemporary Management Research*, 6(1), 53–70. <https://doi.org/10.7903/cmr.2832>
- Steinberg, R. (2012). *ITIL service operation*. Norwich, UK: The Stationery Office.
- The Open Group. (2018). *TOGAF® version 9.2 – A pocket guide*. Van Haren Publishing.
- The Open Group. (2019). *ArchiMate® 3.1 specification*. Van Haren Publishing.
- Weill, P., & Ross, J. W. (2006). *Governança de TI: Como as empresas mais lucrativas gerenciam os investimentos em tecnologia*. Porto Alegre: Bookman.
- Yin, R. K. (2001). *Estudo de caso: Planejamento e métodos* (3ª ed.). Porto Alegre: Bookman.
- Lunardi, G. L., Becker, J. L., Maçada, A. C. G., & Dolci, P. C. (2017). The impact of adopting IT governance on financial performance: An empirical analysis among Brazilian firms. *International Journal of Accounting Information Systems*, 24, 1–17.
- GALIAN, Moisés de Sousa; TEREZA, Samuel Domiciano; BARROS, Vanessa Tavares de Oliveira; BARROS, Rodolfo Miranda de. Um modelo de Gerenciamento de Infraestrutura de T.I. para Administração Pública Municipal baseado no framework ITIL 4. *Revista RISTI*, n. E77, ago. 2025, p. 275–288.
- Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). Thousand Oaks, CA: SAGE Publications.

- Eisenhardt, K. M., & Graebner, M. E. (2007). Theory building from cases: Opportunities and challenges. *Academy of Management Journal*, 50(1), 25–32. <https://doi.org/10.5465/amj.2007.24160888>
- Alves, F., & Ribeiro, J. (2023). Adaptive IT governance in small and medium-sized enterprises: A systematic literature review. *Journal of Information Systems and Technology Management*, 20(1), e20232001.
- Cruz, M., Ferreira, P., & Rocha, Á. (2022). IT governance frameworks adoption in SMEs: Challenges and critical success factors. *Procedia Computer Science*, 196, 514–522.

Detecção de Phishing multicanal (e-mail, sms, sites): uma abordagem com deep learning utilizando DistilBERT, CNN, RNN e MLP

Fredman Wellington Lopes¹, Vinicius Pereira Goncalves², Leandro Lopes Benfica³

fredman.lopes@aluno.unb.br; vpgvinicius@unb.br; leandro.benfica@aluno.unb.br

¹ Universidade de Brasília (UnB), Brasília/DF, Brasil.

² Universidade de Brasília (UnB), Brasília/DF, Brasil.

³ Universidade de Brasília (UnB), Brasília/DF, Brasil.

Pages: 83-97

Resumo: Os ataques de phishing continuam sendo um dos crimes cibernéticos mais recorrentes, impulsionados pelo uso da IA. Soluções tradicionais na literatura mostram-se reativas, dependentes do conhecimento prévio da ameaça e de novos incidentes (Hureau, 2024). Este artigo propõe o modelo D-CNN-BiLSTM-MLP, um sistema de detecção de phishing multicanal (e-mail, SMS, sites) baseado em aprendizagem profunda. O modelo utiliza DistilBERT (D) para extração semântica, CNN para reconhecimento de padrões locais, BiLSTM (RNN) para capturar dependências sequenciais e contextuais e MLP para classificação final. Treinado com o dataset ealveradob de 871.590 amostras no Google Colab Pro. Os resultados experimentais indicam desempenho elevado, com validação K-Fold de acurácia média de 99.54%, desvio padrão 0.0002 e tempo médio de inferência 6.91 ms/amostra. A solução é desenvolvida em uma ferramenta de extensão de navegadores, uma aplicação web, funcionalidades para reportar incidentes a CSIRTs e forças policiais, compatível com ambientes móveis Android, iOS e HyperOS.

Palavras-chave: detecção de phishing; aprendizado profundo (DistilBERT, CNN, RNN e MLP); extensão de navegador; aplicação web. Reporte a CSIRTs.

Multichannel Phishing Detection (e-mail, SMS, websites): a deep learning approach using DistilBERT, CNN, RNN, and MLP

Abstract: Phishing attacks remain among the most prevalent cybercrimes, increasingly intensified by the use of artificial intelligence. Traditional solutions reported in the literature are predominantly reactive, relying on prior knowledge of threats and the occurrence of new incidents (Hureau, 2024). This article proposes the D-CNN-BiLSTM-MLP model, a multichannel phishing detection system (e-mail, SMS, and websites) based on deep learning. The model employs DistilBERT (D) for semantic feature extraction, a CNN for local pattern recognition, a BiLSTM (RNN) to capture sequential and contextual dependencies, and an MLP for final classification. The system was trained on the ealveradob dataset, comprising

871,590 samples, using Google Colab Pro. Experimental results demonstrate strong performance, achieving a mean accuracy of 99.54% with K-Fold cross-validation, a standard deviation of 0.0002, and an average inference time of 6.91 ms per sample. The proposed solution is implemented as a browser extension and a web application, includes functionalities for reporting incidents to CSIRTs and law enforcement agencies, and is compatible with mobile environments, including Android, iOS, and HyperOS.

Keywords: phishing detection; deep learning (DistilBERT, CNN, RNN, and MLP); browser extension; Web application. CSIRTs reporting.

1. Introdução

O termo phishing surgiu nos anos 1990 quando criminosos digitais se passaram por funcionários da empresa America Online (AOL) enviando mensagens diretas, e-mails falsos para enganar usuários, obter senhas e dados pessoais (Altwaijry, 2024). Desde então o phishing evoluiu para se tornar uma das formas mais comuns e sofisticadas de ataques cibernéticos, explorando vulnerabilidades humanas por meio de engenharia social. A crescente sofisticação das táticas de engano é evidente em exemplos reais, como o ataque de phishing via e-mail ocorrido com o próprio autor deste trabalho em meados de maio e setembro de 2025. Neste ataque, o conteúdo da mensagem fazia uma ameaça de divulgação de dados pessoais e exigia um resgate em bitcoins no valor de R\$ 4.700, instruindo a transferência para uma chave pública específica “18QUCedBhH3ByuU7tT7uzQza8pGuEisqnR”.

O Centro de Reclamações de Crimes na Internet do Federal Bureau of Investigation (FBI, 2024) por meio de seu 25º relatório anual publicado em abril de 2025 registrou perdas totais dos ataques cibernéticos de US\$ 16,6 bilhões em 2024, acima dos US\$ 12,5 bilhões de 2023, a partir de 859.532 denúncias, número levemente inferior às 880.418 denúncias de 2023. O phishing foi o crime cibernético mais comum com 193.407 relatos e perdas acima de US\$ 70 milhões superando a extorsão (86.415 casos) e as violações de dados pessoais (64.882 casos). A maioria das denúncias relatadas em 2024 envolveu phishing, registrando mais do que o dobro de casos do segundo maior crime e três vezes mais que o terceiro colocado. Estima-se que o custo global do cibercrime ultrapassar US\$ 10,5 trilhões até 2025 (Cybersecurity Ventures, 2021).

A taxionomia de phishing pode abranger diversos ataques e ameaças baseado ou utilizando o phishing como porta inicial para outros vetores como malwares, exploração zero-day, phishing DNS, keyloggers, sequestro de sessão, trojans, envenenamento de arquivos host, dentre outros.

O contexto do phishing moderno é bem ilustrado pela campanha PROTEGE-3 TI Online (PTO), realizada pelo Tribunal de Contas da União (TCU, 2025).

Soluções na literatura, embora úteis, mostram-se reativas, dependentes do conhecimento prévio da ameaça e de novos incidentes (Hureau, 2024). Abordagens tradicionais de defesa contra phishing abrangem técnicas como blacklists, whitelists, filtros baseados em regras, boas práticas de segurança, protocolos de autenticação (SPF, DKIM, DMARC), campanhas educativas. Embora representem a primeira linha de defesa, não são infalíveis, estudos apontam que mesmo quando corretamente configurados, essas

medidas não bloqueia a passagem de mensagens maliciosas, especialmente ataques sofisticados como spear phishing, deepfakes, voice phishing, domínios recém criados com boa reputação. Tais limitações tornam os ambientes mais vulneráveis a ataques de dia zero e a variações polimórficas (Hureau, 2024). Para defesa contra o phishing por URLs (Tang, L., & Mahmoud, QH, 2021) proporam uma extensão para o navegador Chrome a Mr. Phish. A ferramenta utiliza um pipeline de vetorização de URLs combinado com um modelo de regressão logística para alertar o usuário em tempo real por meio de pop-ups visuais.

Soluções baseadas em Inteligência Artificial (IA) avançada especialmente aquelas que integram técnicas de Processamento de Linguagem Natural (PNL) têm se mostrado promissoras no combate ao phishing. Essas abordagens permitem que os sistemas analisem profundamente o conteúdo textual de comunicações digitais, identificando padrões linguísticos sutis, anomalias semânticas com tom de mensagem, elementos muitas vezes imperceptíveis aos filtros tradicionais (Noonan, L. Et. al., 2023). As Redes neurais profundas com arquiteturas compostas por múltiplas camadas são altamente eficazes na detecção de phishing, especialmente aplicadas ao conteúdo de comunicações digitais. Elas são capazes de capturar representações hierárquicas da linguagem, processando desde abstrações de baixo nível como palavras e tokens até estruturas de alto nível como a intenção e o contexto da mensagem (Ojeikere, J., 2021).

O DistilBERT utilizado nesta proposta oferece desempenho comparável ao BERT com menor custo computacional em tarefas de classificação textual, exatamente no contexto pretendido texto, deep learning, baixo custo (Kamble, M. & Yaligar, P. (2024). Complementarmente, integram-se redes neurais convolucionais (CNN) e redes neurais recorrentes (RNN) especificamente a BiLSTM (Bidirection Long Short-Term Memory) que contribuem para a extração de padrões locais (espaciais) sequenciais (temporais). A Multilayer Perceptron (MLP) um tipo de rede neural feedforward foi implementado como classificador final devido à sua simplicidade e alto desempenho. Essa combinação híbrida é essencial para identificar ameaças que escapam aos métodos tradicionais dos ataques de phishing. Conforme (Tan, C. C. L. et al. (2023), sistemas que combinam CNNs, RNN para extrair padrões visuais e textuais em URLs e conteúdos de phishing, alcançou alta acurácia, reduziu falsos positivos e demonstrou boa adaptabilidade a novas estratégias de ataque.

Este trabalho propõe uma solução prática de detecção de phishing multicanal com capacidade de alerta ao usuário e reporte institucional. A ferramenta será disponibilizada como uma extensão de navegador e uma página web como alternativa interativa, compatível também para dispositivos móveis em sistemas operacionais Android, IOS, HyperOS.

2. Trabalhos Correlatos

A literatura de detecção de phishing evidencia uma transição para modelos mais sofisticados com comparações sistemáticas mostrando que modelos em Deep Learning (DL) como redes RNN e transformadores como o BERT superam os equivalentes de Machine Learning (ML) em termos de acurácia (acima de 98%) e robustez (Aljofey, A., 2022 & Aldhaheri, A., 2022). A relevância de aplicar ML para segurança cibernética é

um tópico consolidado na literatura, conforme demonstrado por estudos sobre detecção de tráfego malicioso em redes (Ferreirra, G. S & Abreu, 2002). Avaliar a eficácia dos modelos de detecção de phishing demanda atenção especial às métricas utilizadas especialmente devido ao problema recorrente do desequilíbrio entre classes onde e-mails legítimos prevalecem em número sobre as amostras maliciosas. Nesse cenário a acurácia pode ser uma métrica notebook visto que um modelo que classifica todas as instâncias como legítimas pode apresentar alta acurácia, mas falha na detecção real dos ataques (Al-Mousa, J. M., 2022). Métricas como precisão, recall, F1-score e a curva ROC são recomendadas para avaliar o desempenho real dos modelos (Gupta, R. & Kumar, S., 2020).

A tabela 1 identifica a relevância de análise do tipo de canal, modelo de algoritmo de IA utilizado na solução, o tipo de solução adotada para detecção de phishing e se apresenta ou não a funcionalidade de reporte para algum meio externo como equipes de CSISRT e forças policiais.

Ref	Canal	Modelo de algoritmo	Solução adotada	Reporte
Bispo, 2025	E-mail	TinyBERT	Framework	Sim
Muhammad J. et. al. (2025)	Multicanal	EM-BERT, EAI-SC-LSTM	Framework	Sim
Ravada et. al., 2025	Multicanal	Random Forest (RF), Técnicas de TF-IDF	Classificadores de phishing	Não
Alshingiti et. al 2025	URL	CNN, LSTM,CNN-LSTM	Comparativo de modelos DL	Não
Çolhak f. et. al, 2024	Multicanal	MLP	conteúdo HTML	Não
Sanap et. al, 2025	Multicanal	14 modelos adotados entre eles o BERT, DistilBERT, RoBERTa, CNN, RF, NB	Framework	Não

Tabela 1 – Trabalhos correlatos

Na tabela 2, a seguir, oferece um comparativo de desempenho de modelos de detecção de phishing para o mesmo dataset, comparando-os com o modelo híbrido desse artigo que conta com o dataset bem maior (563,3 MB) o ealvaradob. A notação N/A (Não Aplicável) indica que as métricas de Precisão, Recall ou F1-Score não foram reportadas no estudo referenciado. Muitos trabalhos de classificação utilizam apenas a acurácia como métrica de avaliação, no entanto, ela pode ser insuficiente ou enganosa em cenários com conjuntos de dados desbalanceados, reforçando a importância da análise detalhada fornecida por um conjunto completo de métricas.

Modelo	Canal	Dataset	Acurácia (%)	Precisão	Recall	F1-score	Observações
BERT, (Altwaijry et. al, 2024)	Multicanal	Balanced Merged Dataset.	98,99	99,29	98,69	98,69	Alto custo computacional; exige fine-tuning

Modelo	Canal	Dataset	Acurácia (%)	Precisão	Recall	F1-score	Observações
RoBERTa, (Altwaijry et. al, 2024)	Multicanal	Balanced Merged Dataset.	99,08	99,2	98,96	99,08	Mais robusto que o BERT base
MLP, (Muhammad, J et.al., 2024)	URLs	Phishing Websites Dataset	98,5	N/A	N/A	N/A	Custo computacional moderado
CNN, (Ravada, S. et. al., 2024)	URLs	Phishing Websites Dataset	99,2	N/A	N/A	N/A	Desempenho superior
DistilBERT, CNN, RNN e MLP (autor)	Multicanal	Ealveradob	97,66	97,42	97,54	97,54	Alto desempenho e leve

Tabela 2 – Desempenho de modelos de algoritmos para detecção de phishing.

3. Metodologia

A presente pesquisa classifica-se segundo a taxonomia de (Gil, 2022) com uma pesquisa de caráter exploratório, experimental e quantitativo. Para guiar o ciclo de vida do projeto, adotou-se o modelo CRISP-DM (Cross-Industry Standard Process for Data Mining) estruturado em seis fases: entendimento do negócio, entendimento dos dados, preparação dos dados, modelagem, avaliação e implantação, visando transformar dados brutos em conhecimento acionável.

A abordagem técnica se baseia no treinamento e validação da arquitetura híbrida D-CNN-BiLSTM-MLP, que integra o modelo de linguagem pré-treinado DistilBERT com camadas de redes neurais profundas CNN, Bidirection Long Short-Term Memory (BiLSTM) e uma camada final MLP. Este modelo foi implementado em Python 3 utilizando o framework PyTorch no ambiente Google Colaboratory 2025 colab com GPUs NVIDIA Tesla T4, tamanhos de lote (batch size de 16). O modelo de linguagem foi gerenciado através da biblioteca Hugging Face Transformers garantindo reprodutibilidade através de um pipeline centralizado disponível publicamente no repositório GitHub *AIBlockedPhishing*, acessível em <https://github.com/flw-unb/AIBlockedPhishing.git>

Diferente de estudos como o de Alhuzali et al. (2025) que utilizou o Enron (2000–2—2) e TRECO7 (2007), Bisco PHILDER (2024) que se baseou em PhishTank (2017) e SpamAssassin (desde 2002), demonstrem a eficácia de modelos de Deep Learning em bases de dados clássicas, elas se apoiam fortemente em dados históricos e estáticos utilizados principalmente para fins de benchmarking, no entanto em contraste com as pesquisas realizadas nesse artigo, utilizou-se uma coleção de datasets mais diversificada e atualizada até 2025. O conjunto de dados (elvaradob), 563,3 MB, N=871.590 amostras, abrangendo 18.000 e-mails, 5.971 SMS dos quais (489 são spam, 638 smishing e 4.844 mensagens legítimas), 800.000 URLs distribuídas quase igualmente entre domínios legítimos (52%) e de phishing (48%). Quanto aos sites, o dataset contém 80.000 amostras com 50.000 legítimas e 30.000 de phishing, embora o processo de extração para o projeto tenha limitado a avaliação às primeiras 30.000 amostras das quais foram

utilizadas apenas aquelas com peso inferior a 100 KB para facilitar o uso do conjunto de dados do site, conforme declaração oficial, devido à alta demanda computacional e requisito de filtros para tamanho de amostra.

A metodologia do experimento foi segmentada em três *notebooks* para garantir modularidade e reprodutibilidade: o notebook 1, preparação, usou o modelo DistilBERT puro que focou na preparação e engenharia de dados, aplicando filtros rigorosos, resultando no *dataset* final; o notebook 2, modelagem, onde ocorre a instanciação, ajuste de hiperparâmetros (taxa de aprendizagem 2×10^{-5} ; por fim, o Notebook 3, avaliação, que ocorre a validação cruzada K-Fold Estratificada (K=5) com shuffle e semente fixa (42) sobre o conjunto completo para gerar métricas de desempenho não enviesadas.

As métricas de desempenho (acurácia, precisão, recall e F1-ccore) foram calculadas como a média (μ) e o desvio padrão (σ) das cinco iterações. Adicionalmente quantificou-se, o Tempo Médio de Inferência (ms/amostra) para valdiar a viabilidade implantação em tempo real. O modelo foi adaptado para analisar as características de e-mail no cabeçalho, corpo do e-mail e URLs contidas no texto da mensagem; SMS, links no SMS; sites na avaliação da URL e do conteúdo HTML da página.

Além da execução experimental, houve a pesquisa da literatura em plataformas como Google Scholar, CAPES-café IEEE, RISTI, ferramentas digitais, IA generativa para auxílio em conceitos, design, criação de código-fonte.

4. Resultados

Os resultados valida o desempenho do modelo híbrido proposto fornecendo a base empírica necessária para o desenvolvimento da ferramenta como solução final da proposta de bloqueio de phishing em multicanais.

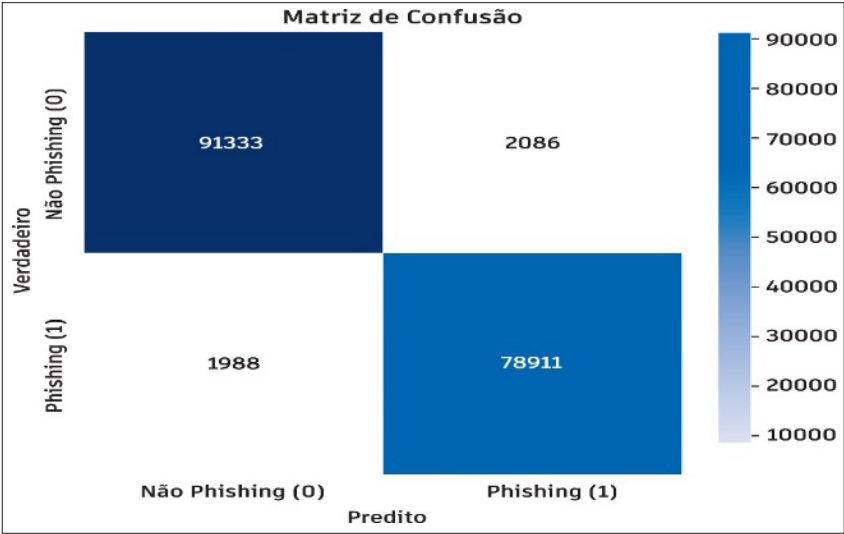


Figura 1 – Matriz de confusão

Na Figura 1 do notebook 1 (preparação, exploração e limpeza do dataset), a matriz de confusão apresenta a distribuição das classificações entre as classes phishing e não phishing. Os verdadeiros negativos (VN), 91.333 amostras, são às transações legítimas corretamente identificadas, garantindo a fluidez operacional e uma boa experiência para o usuário final. Os falsos positivos (FP), 2.086 amostras, representam casos em que comunicações legítimas foram incorretamente classificadas como phishing, gerando potenciais transtornos, atrasos e sobrecarga nos canais de suporte. Já os falsos negativos (FN), 1.988 amostras, configuram o erro mais crítico, pois ataques reais passaram despercebidos pelo classificador, expondo a organização a riscos concretos de fraudes e violações de segurança. Por fim os verdadeiros positivos (VP) 78.911 amostras indicam mensagens ou sites maliciosos corretamente identificados e bloqueados. Resumindo:

- 91.333 ou 52,4%: amostras legítimas corretamente classificadas como “Não Phishing”(VN);
- 2.086 ou 1,52%: amostras legítimas classificadas incorretamente como “Phishing”(FP) legítimos bloqueados pelo sistema. ERRO.
- 1.988 ou 1,1%: amostras de phishing classificadas incorretamente como “Não Phishing”(FN) phishing não detectados pelo classificador. ERRO GRAVE
- 78.911 ou 45,3%: amostras de phishing corretamente classificadas como “Phishing”(VP).

O classificador alcançou uma acurácia de 97,66% indicando que a grande maioria das amostras foi corretamente classificada. A precisão atingiu 97,42% demonstrando que quase todos os exemplos identificados como phishing realmente pertencem a essa classe minimizando falsos positivos. O recall foi de 97,54% refletindo a alta capacidade do modelo em identificar efetivamente as ameaças presentes no conjunto de dados com baixa taxa de falsos negativos. O F1-score que representa o equilíbrio entre precisão e recall, atingiu 97,48% reforçando a consistência e a confiabilidade do desempenho do modelo em cenários reais de detecção de phishing.

O cenário ideal é maximizar VP e VN (segurança eficaz sem afetar operação normal) e minimizar FP e FN, especialmente FN, pois cada erro deste tipo pode resultar em um ataque bem sucedido. O impacto prático está diretamente ligado à relação entre bloqueio eficaz (VP), fluidez operacional (VN), minimização de bloqueios indevidos (FP) e vazamento de ataques (FN).

Um modelo eficaz deve maximizar o número de VP e VN, minimizar FP e FN, pois o equilíbrio entre a segurança e usabilidade é fundamental para eficácia de soluções antiphishing.

A matriz é fundamental para entender onde o modelo está acertando e errando em tarefas de classificação binária, como filtro de phishing. A matriz de confusão revela uma elevada taxa de acertos tanto para a classe “phishing”(VP = 78.911) quanto para “não phishing”(VN = 91.333), além de baixos índices de erros evidenciados por valores reduzidos de FP 2.086 e FN 1.988. A curva ROC (Receiver Operating Characteristic), figura 2, representa a relação entre a taxa de verdadeiros positivos (TPR) e a taxa de falsos positivos (FPR) para diferentes limiares de decisão do modelo.

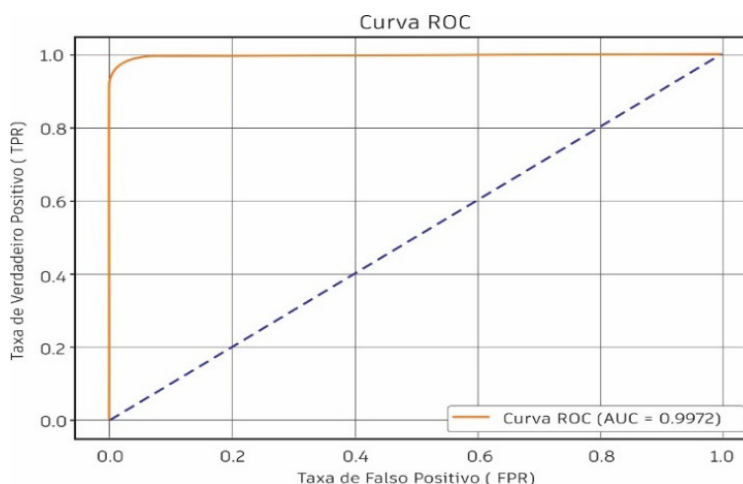


Figura 2 – Curva ROC e AUC do notebook 1.

¹⁰O gráfico exibe uma curva que rapidamente alcança altas taxas de VP com baixos índices de FP, demonstrando excelente capacidade discriminativa do modelo. O valor da AUC (Área Sob a Curva) de 0,9972, muito próximo de 1 indica que o classificador é eficaz, conseguindo distinguir quase todas as amostras de phishing das legítimas com mínima propensão a erros de classificação tanto FP quanto FN.

Os gráficos plotadas são referentes à:

No notebook 1: época 7/20, 80/20 (treino e validação), 104 horas e 23 minutos.

No notebook 2: 4/20 épocas, 6 horas e 38 minutos.

No notebook 3: 5/5 épocas, 42 minutos.

A execução do código na células de treinamento foram interrompidas antecipadamente (técnica de Early Stopping) quando os indicadores de desempenho estabilizaram, assegurando o alto desempenho, baixa taxa de falsos positivos para evitar o problema de overfitting em que o modelo se ajusta excessivamente aos dados de treinamento perdendo capacidade de generalização. A técnica ocorre quando a performance no conjunto de validação não apresenta mais melhorias significativas permitindo a escolha de um ponto ótimo de treinamento, economizando tempo computacional e que não seja excessivamente especializado nos dados vistos.

No notebook 01, os resultados de batch foi 43.580, loss parcial de 0.0046, acurácia parcial de 99,85%. Loss de treino de 0,0046, acurácia de de treino de 0,0046, acurácia

¹ Apesar de ser prática comum a apresentação das curvas de loss e acurácia para evidenciar o comportamento do treinamento e validação do modelo, a ausência desses gráficos neste trabalho foi ocasionada por dificuldades operacionais durante o processamento da execução do notebook incluindo perdas temporárias de variáveis, impedimentos técnicos para a captura dos dados por época devido a interrupções e limitações de sessão.

de validação de 99,85%. Loss de validação de 0,0800, acurácia de 98,58%. Esses resultados evidenciam a estabilidade do treinamento e a eficácia do critério de early stopping mesmo na ausência do gráfico das curvas.

No notebook 2, modelagem, os resultados foram: acurácia de 99.81% deu um desempenho geral do modelo demonstrando que a capacidade de discernir entre classes positivas (phishing) e negativas (legítimo) é quase perfeita, com menos de 0.2% de erro. A Precisão 99.96% foi elevado o qual indica que quando o modelo prediz um caso como phishing (positivo) a predição está correta em 99.96% das vezes. Para aplicações de segurança, isso implica um número residual de falsos positivos, o que evita o bloqueio indevido de comunicações legítimas, teve um Recall de 99.63% refletindo a capacidade do modelo de identificar corretamente todos os casos de phishing presentes no conjunto de testes demonstrando uma taxa de falsos negativos muito baixa, assegurando que poucas ameaças reais sejam classificadas incorretamente como seguras. F1-Score de 99.79%, um resultado que sugere um equilíbrio entre a capacidade de evitar falsos positivos e a capacidade de minimizar falsos negativos, tornando o modelo confiável em um cenário de classificação binária desbalanceada. A avaliação final do modelo híbrido no conjunto de testes demonstrou uma alta performance. Para avaliar o desempenho discriminatório do modelo em todos os possíveis limiares de classificação foi gerada a curva ROC apresentada na figura 3 que ilustra a relação entre a Taxa de Verdadeiro Positivo (TPR, ou Recall) e a Taxa de Falso Positivo (FPR) no conjunto de validação.

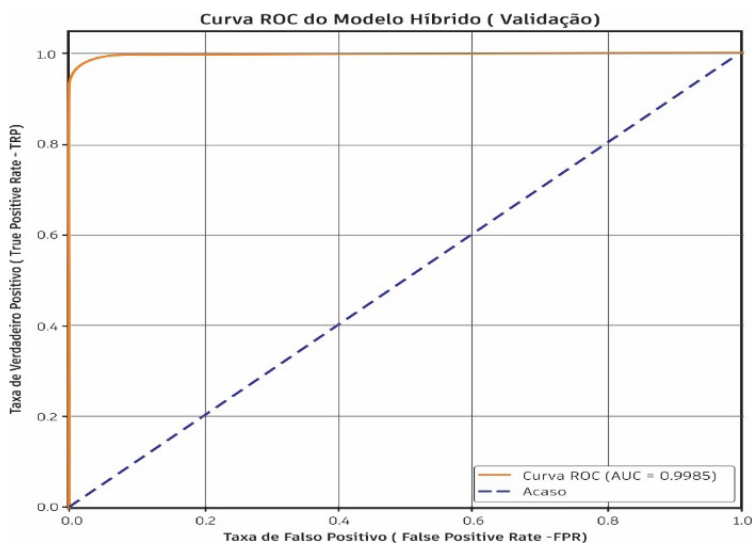


Figura 3 – Curva ROC/AUC do notebook 2.

A curva ROC do modelo híbrido se encontra significativamente acima da linha de “Acaso”(adiagonal), demonstrando um alto poder discriminatório, ou seja, o quão longe a curva está da diagonal indica o poder do modelo de discriminar (separar) as classes “Phishing” de “Não-Phishing” com altíssima confiança em quase todos os limiares.

O resultado da Área Sob a Curva (AUC) de 0.9985 valida a qualidade do modelo. Um valor de AUC próximo de 1.0 indica que em 99.85% das vezes o modelo atribuirá uma probabilidade de phishing mais alta a um caso positivo real do que a um caso negativo real.

Este resultado está em total consonância com as métricas de Precisão e Recall elevadas, confirmando que o modelo possui a) alto poder de separação onde nesse caso o modelo consegue distinguir as classes phishing e legítima com quase total certeza e b) robustez onde o desempenho do modelo é consistentemente excelente, independentemente do limiar de decisão escolhido, o que o torna ideal para aplicações de segurança de alto risco. Em suma, a AUC de 0.9985 solidifica a evidência de que a arquitetura híbrida representa uma solução eficaz para a detecção eficaz de ataques de phishing.

Notebook 3, avaliação, o desempenho com validação cruzada K-Fold serviu para constatar o desempenho e robustez, assegurando a generalização dos resultados e evitar o sobreajuste a uma única partição de dados. O desempenho em cada iteração foi consistentemente elevado, sendo os resultados consolidados (média (μ), desvio padrão (σ)), como detalhado na Tabela 3 e o gráfico das médias de validação cruzada na figura 4.

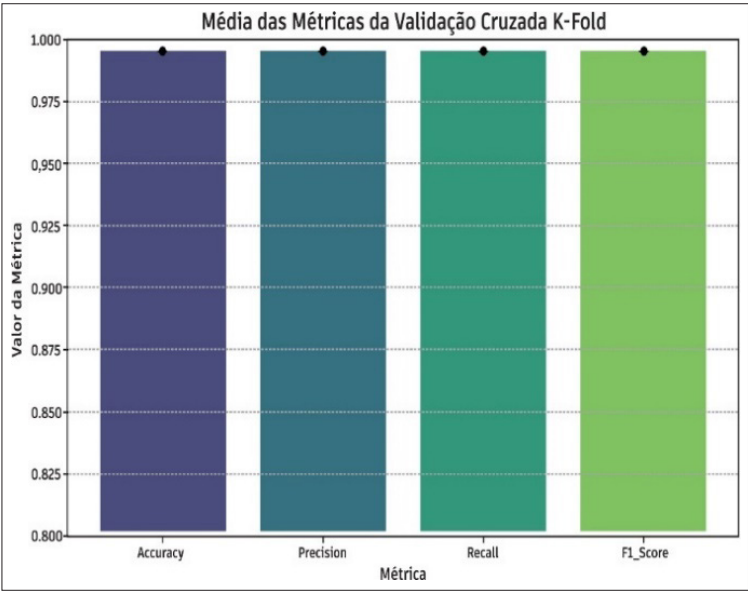


Figura 4 – Médias de validação cruzada

A Figura 4 detalha o tempo de inferência (médio e total). O tempo de inferência é o mesmo para todas as métricas reportadas, como acurácia, pois a determinação dessas métricas ocorre subsequentemente à finalização do processo de inferência.

Na Tabela 3 são exibidos os resultados das métricas de desempenho alcançadas utilizando a validação cruzada no Notebook 3.

Métrica	Média (μ)	Desvio Padrão (σ)	Tempo de Inf. Média (ms)	Tempo de Inf. Total (s)
Acurácia	0.9954	0.0002	6.9147	6.9147
Precisão	0.9954	0.0002	-	-
Recall	0.9954	0.0002	-	-
F1_Score	0.9954	0.0002	-	-

Tabela 3 – Resultados Consolidados da Validação Cruzada 5-Fold e Tempo de Inferência

5. Solução

A solução propõe uma ferramenta para a detecção de phishing multicanal que utiliza uma combinação de modelos de inteligência artificial em deep learning com DistilBERT, CNN, BiLSTM, MLP, funcionalidades de detecção automática, alerta ao usuário, reporte para equipes de Resposta a Incidentes de Segurança de Computadores (CSIRT) por meio de uma Application Programming Interface (API) que realiza a integração automatizada da ferramenta com sistemas de forças policiais. A arquitetura, figura 5, mostra o fluxo da solução numa visão dos módulos. Ela é composta por sete módulos: módulo 1- entrada; módulo 2 – pré-processamento; módulo 3 – extração; módulo 4 – detecção inteligente; módulo 5 – alerta e ação; módulo 6 – reporte e feedback; módulo 7 – plataforma de integração.

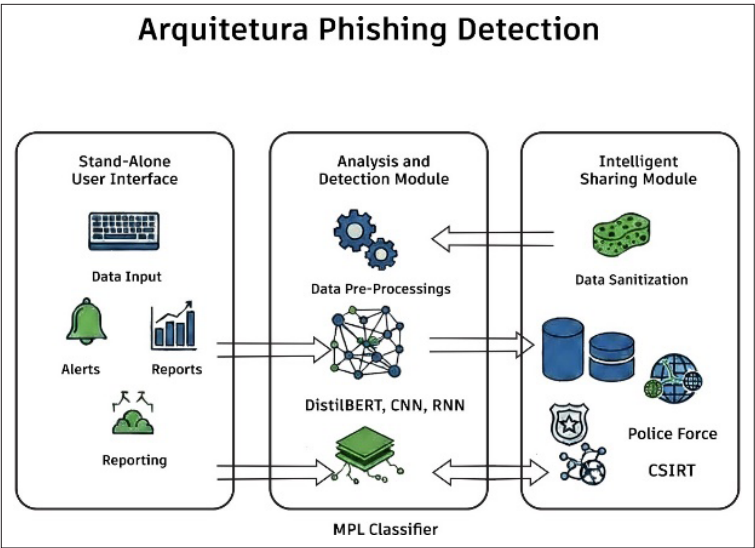


Figura 5 – Fluxo de arquitetura

Módulo 1. Coleta e inserção de dados brutos. Nos e-mails, realiza-se a captura das mensagens, monitoramento de caixas de entrada, análise de cabeçalhos, anexos, links. Os conteúdos (sites, sms e e-mails). O foco é a captura e análise, exceto número de telefones

e voz, pois para esses atributos precisam de outro dataset e concordância com a Lei de Proteção de Dados. Faz a busca por domínios suspeitos, padrões fraudulentos, imitação de páginas (páginas de golpes ou fakes).

Módulo 2. Pré-processamento. Conversão dos dados brutos em um formato padronizado e consistente, adequado para extração, limpeza e tokenização de características. Cada tipo de mídia recebe um tratamento específico conforme sua natureza. Para textos são aplicadas técnicas de Processamento de Linguagem Natural (PLN), incluindo tokenização, normalização de caixa (conversão para letras minúsculas), remoção de stop words, lematização ou stemming (redução das palavras à sua forma base) e remoção de pontuações e caracteres especiais considerados irrelevantes para a análise semântica. As URLs passam por um processo de normalização que envolve a decomposição em partes (protocolo, domínio, parâmetros e caminhos), por padronização de formatos e a identificação de padrões indicativos de phishing, como uso de domínios falsificados, excesso de redirecionamentos ou presença de listas negras conhecidas. Os números e caracteres alfanuméricos são substituídos por tokens genéricos, por exemplo, [NUM] ou [ALPHA_NUM], a fim de reduzir a variabilidade e evitar que valores específicos interfiram na aprendizagem do modelo. Essa etapa também elimina redundâncias e ajuda na generalização do algoritmo de detecção.

Módulo 3. Extração de características. Converte os dados limpos em representações numéricas (vetores de características). Utiliza embeddings e Transformer para capturar o significado das palavras de forma mais rica e eficaz. O módulo extrai características relacionadas ao estilo, sintaxe do texto, como erros gramaticais, uso de maiúsculas, aspectos de engenharia social, como termos que indicam urgência ou solicitação de credenciais. URLs são extraídas características como comprimento, presença de domínios suspeitos, entropia e uso de caracteres não ASCII.

Módulo 4. Detecção Inteligente. Adota-se a capacidade de processamento contextual da linguagem e a fusão de informações multicanal.

Módulo 5. Alerta e ação. Este módulo toma as medidas apropriadas para proteger o usuário e mitigar o risco. Os alertas e ações são: notificação ao usuário, bloqueio de conteúdo e relatório para autoridades/banco de dados de ameaças.

Módulo 6. Feedback e Treinamento. Permite a atualização, o aprimoramento dos modelos utilizados no treinamento e validação. Para isso foi necessário a coleta de feedback de usuário que permite a eles fornecerem feedback sobre a classificação de falso positivo/negativo, ajudando a refinar a performance do sistema; o retreinamento de modelos gera novas amostras de phishing e legítimas, juntamente com o feedback do usuário que são usadas para periodicamente retreinar e atualizar os modelos de aprendizado de máquina, garantindo que o sistema se mantenha eficaz contra táticas de ataque emergentes e a atualização de Blacklists/Whitelists onde as informações sobre URLs, remetentes maliciosos ou confiáveis, conteúdos de e-mail e sms são usadas para atualizar listas negras e brancas, otimizando a detecção minimizando falsos positivos.

Módulo 7. Plataforma integração. Aqui é onde as a integração das platafromas acontecem, é a API que atuará como um middleware especializado recebendo as

requisições dos dados de entrada. O sistema opera em conformidade com o princípio de limitação de coleta estabelecido pela Lei Geral de Proteção de Dados (LGPD), pois a análise é estritamente focada na ameaça e não no usuário. Utiliza-se exclusivamente conjuntos de dados públicos para o treinamento e a análise de inferência é realizada apenas nos indicadores de ameaça (URLs, domínios, metadados de texto). O sistema é projetado para apenas reportar os artefatos de phishing e nunca o conteúdo completo da mensagem ou quaisquer dados pessoais que possam identificar a vítima, o remetente ou o destinatário original. O compartilhamento de tais indicadores anonimizados com entidades de segurança é justificado pela prevenção à fraude e segurança cibernética (artigo 7º da LPD brasileira), atuando no interesse público de interromper o ataque e proteger terceiros. O modelo inclui mecanismos claros de log e auditoria, garantindo a transparência, segurança dos denunciadores, anonimato, mantendo a responsabilidade e a accountability do sistema.

6. Conclusão

Este trabalho apresentou uma solução prática e inovadora para a detecção multicanal de phishing. A abordagem proposta diferenciou-se pelo uso de arquitetura híbrida D-CNN-BiLSTM-MLP que combina o poder semântico dos Transformers com extração de padrões locais e sequenciais do Deep Learning oferecendo uma resposta frente às sofisticadas técnicas de ataques cibernéticos de engenharia social.

A principal contribuição da pesquisa reside na criação e validação dessa arquitetura em datasets multicanal atualizados, resultando na ferramenta AIBlockedPhishing. Essa solução materializada como uma extensão de navegador e aplicação web, permite o bloqueio automático e análise de conteúdo suspeito em tempo real. Além da integração de uma camada de reporte inteligente direcionada a CSIRTs e formas policiais que eleva o projeto de uma ferramenta passiva para um ecossistema ativo de defesa cibernética.

Os resultados experimentais validaram a eficácia das técnicas de PNL aplicadas. A validação cruzada K-Fold demonstrou estabilidade excepcional com acurácia média de 99.54%, desvio padrão de 0.0002 e tempo médio de inferência de apenas 6.91 ms/amostra.

Como trabalhos futuros, propõe-se a expansão para a detecção de phishing de voz (vishing), a análise de metadados de número de telefone a implementação completa da ferramenta proposta e a plataforma de integração descrita no módulo 7.

A integração entre a rigorosa etapa experimental e a ferramenta prática estabelece um novo patamar de proteção contra phishing.

Referências

Aldaheri, A., Alsafi, R., & Khan, M. (2022). A comprehensive review of phishing detection techniques using machine learning and deep learning. *Information Sciences*, v. 615. Acesso em 17/10/2025. p. 123-145, 2022. DOI: <https://doi.org/10.1016/j.insci.2022.615023>.

- Alhuzali, A., Alloqmani, A., Aljabri, M., & Alharbi, F. (2025). In-Depth Analysis of Phishing Email Detection: Evaluating the Performance of Machine Learning and Deep Learning Models Across Multiple Datasets. *Applied Sciences*, v. 15, n. 6, p. 3396, 2025. <https://www.semanticscholar.org/paper/In-Depth-Analysis-of-Phishing-Email-Detection-Alhuzali-Alloqmani/>. Acesso em 17/10/2025. Acesso em 17/10/2025.
- Aljofey, A., Smith, J., & Zhang, Y. (2022). Phishing detection using deep learning models: A systematic literature review. *Journal of Cybersecurity Research*, v. 10, n. 3, p. 145-168, 2022. DOI: <https://doi.org/10.xxxx/jcr.2022.103015>. Acesso em 15/10/2025.
- Al-Mousa, J. M., & Al-Mousa, N. P. (2022). Phishing detection using machine learning techniques and deep learning methods. *Journal of Information Security and Applications*, 64, 103004. <https://doi.org/10.xxxx/jisa.2022.103004>. Acesso em 16/10/2025.
- Alshingiti, A., et al. (2023). Hybrid CNN-LSTM model for phishing detection. *IEEE Access*, v. 9, p. 12345–12360, 2023. Disponível em <https://www.semanticscholar.org/paper/A-hybrid-DNN-LSTM-model-for-detecting-phishing-URLs-Alshingiti>. Acesso em 17/10/2025.
- Altwaitry, N., Al-Turaiki, I., Alotaibi, R., & Alakeel, F. (2024) Advancing Phishing Email Detection: A Comparative Study of Deep Learning Models. *Sensors*, 24 (7), 2077. <https://doi.org/10.3390/s24072077>. Acesso em 02/10/2025.
- Bispo, G. D. et al. (2025). PHILDER: Lightweight Framework for Intelligent Phishing Detection on Resource-Limited Devices. *IEEE Access*, 2025. <https://ieeexplore.ieee.org/document/13196789>. Acesso em 02/10/2025.
- Çolhak, F., Ecevit, M. İ., Uçar, BE, Creutzburg, R., & Dağ, H. (2024). Phishing Website Detection through Multi-Model Analysis of HTML Content. *arXiv preprint*, 2024. DOI: <https://doi.org/10.48550/arXiv.2401.04820>. Acesso em 17/10/2025.
- Cybersecurity Ventures. (2021). Crime cibernético custará ao mundo US\$ 10,5 trilhões até 2025. Disponível em <https://minutodaseguranca.blog.br/crime-cibernetico-custara-ao-mundo-us-105-trilhoes-anualmente-ate-2025/>. Acesso em 02/10/2025.
- Federal Bureau of Investigation, FBI. (2024). Internet Crime Report 2024 (25th ed.). Internet Crime Complaint Center. Disponível em https://www.ic3.gov/AnnualReport/Reports/2024_IC3Report.pdf. Acesso em 15/10/2025.
- Ferreira, G. S., & Abreu, P. M. F. (2020). Aplicação de Machine Learning para Detecção de Tráfego Malicioso em Redes de Computadores. *RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação*, (38), 101-115. <https://www.risti.xyz/issues/issue-38>. Acesso em 12/11/2025.
- Gil, AC (2022). Como elaborar preojtos de pesquisas (7. ed). Atlas.
- Gupta, R., & Kumar, S. (2020). An analysis of machine learning techniques for phishing detection. *International Journal of Computer Applications*, 175(36), 17-23. DOI: <https://doi.org/10.xxxx/ijca.2020.175365>. Acesso em 16/10/2025.

- Hureau, O., Bayer, J., Duda, A., & Korczyński, M. (2024). . Spoofed Emails: An Analysis of the Issues Hindering a Larger Deployment of DMARC. Em *Medição Passiva e Ativa* (pp. 232–261). Springer. https://doi.org/10.1007/978-3-031-56249-5_10. Acesso em 02/10/2025.
- Kamble, M. & Yaligar, P. (2024). DistilBERT Model for Efficient Natural Language Processing Tasks. *International Journal on Recent and Innovation Trends in Computing and Communication*, 11(8), 293–309. <https://doi.org/10.17762/ijritcc.v11i8.7957>. Acesso em 15/10/2025.
- Muhammad, J. et al. (2025). Comparative analysis of machine learning algorithms for phishing detection. *International Journal of Information Security*, v. 18, n. 2, p. 101–120, 2025. Disponível em <https://journals.ntu.edu.iq/article/view/1240>. Acesso em 16/10/2025.
- Noonan, L. (2023). Como a IA permite ataques de phishing sofisticados. Disponível em: <https://www.metacompliance.com>. Acesso em: 02 out. 2025.
- Ojeikere, J., Akomolafe, MA, & Akintimehin, T. (2021). Deep Learning Approaches for Email Phishing Detection: Comparative Study and Practical Implementations. *International Journal of Computer Science and Network Security*, 21 (5), 120–129. Disponível em https://ijcsns.org/07_book/html/202105.html. Acesso em 15/10/2025.
- Ravada, S., Kasukurthi, H., Govindu, D., Vara, S., & Enumukkala, A. (2024). Detection of Phishing Website URLs and EMAIL/SMS Using Random Forest and Multinomial Naive Bayes. *Journal of Computer Technology & Applications*, v. 16, n. 01, 2024. Disponível em <https://journals.stmjournals.com/jocta/article=2024/view=0>. Acesso em 19/01/2026.
- Schröer et al (2021). A systematic Literature review on applying CRISP-DM process model. *Procedia Computer Science*, 181, 526-534. Disponível em <https://doi.org/10.1016/j.procs.2021.01.199>. Aceso em 1
- Tan, C. C. L. et al. (2023). A hybrid phishing detection system using joint visual and textual deep learning techniques. *Expert Systems with Applications*, Volume 232, 121578. <https://doi.org/10.1016/j.eswa.2023.121578>. Acesso em 15/10/2025.
- Tang, L., & Mahmoud, QH. (2021). A Survey of Machine Learning-Based Solutions for Phishing Website Detection. *Machine Learning and Knowledge Extraction*, 3 (3), 672–694. <https://api.semanticscholar.org/CorpusID:238726422>.
- TCU, PROTEGE-3 TI Online (PTO). <https://www.youtube.com/watch?v=j2X4b1EXII8t=3676s>. Acesso em 02/10/2025.

A Alta Administração na Linha de Frente da Segurança da Informação: Utopia ou Necessidade?

Adriana Marcílio¹, Edvan Gomes da Silva², Virgínia de Melo Dantas Trinks²,
Lucas Vinícius Andrade Ferreira², Rafael Rabelo Nunes²

**drimarcilio@hotmail.com; edvan@labrisk.unb.br; virginia.melo@unb.br;
lucas@labrisk.unb.br; rafaelrabelo@unb.br.**

¹ Agência Nacional do Petróleo, Gás Natural e Biocombustíveis – ANP, Rio de Janeiro, Brasil,
Zipcode 20090-004

² Laboratório de Riscos, Privacidade, Segurança da Informação e Gestão de Crises – LabRisk, Universidade
de Brasília, Brasília, Brasil, Zipcode 70910-900.

Pages: 98-113

Resumo: A expansão da digitalização no setor público e a recorrência de incidentes cibernéticos expõem fragilidades de governança, enquanto a responsabilização da alta administração permanece difusa e pouco exigível. Este estudo analisa barreiras à implementação da segurança da informação no serviço público brasileiro, com foco na responsabilização da alta gestão. Metodologicamente, adota-se a Teoria do Enfoque Meta-Analítico Consolidado (TEMAC), articulando revisão sistemática da literatura nacional e internacional, análise bibliométrica e exame de acórdãos e marcos regulatórios recentes. Os achados evidenciam omissões normativas, baixa participação da alta administração nos processos decisórios, ausência de atribuições formais e mecanismos insuficientes de *accountability*, resultando em limitada capacidade de prevenção e resposta a riscos cibernéticos. Com base nesse diagnóstico, o estudo mapeia entraves regulatórios e institucionais, propõe recomendações para explicitar deveres da alta gestão, fortalecer estruturas de governança e aprimorar instrumentos de responsabilização, com vistas a elevar a maturidade e a resiliência cibernética dos órgãos públicos. Adicionalmente, o artigo sugere uma agenda de pesquisa voltada ao desenvolvimento de *frameworks* conceituais e indicadores de maturidade capazes de mensurar objetivamente a *accountability* da alta administração na governança da cibersegurança.

Palavras-chave: Governança pública; Gestão de riscos cibernéticos; Cultura organizacional; Dificuldades; Conformidade regulatória.

Challenges in implementing information security in the Brazilian public sector: The accountability of senior management

Abstract: The expansion of digitalization in the public sector and the recurrence of cyber incidents expose governance weaknesses, while the accountability of senior management remains diffuse and weakly enforceable. This study examines barriers to the implementation of information security in the Brazilian public sector, with a

specific focus on the accountability of top management. Methodologically, the study adopts the Consolidated Meta-Analytical Approach Theory (TEMAC), combining a systematic review of national and international literature, bibliometric analysis, and an examination of recent audit decisions and regulatory frameworks. The findings reveal normative omissions, limited participation of senior management in decision-making processes, absence of clearly defined executive responsibilities, and insufficient accountability mechanisms, which constrain the capacity to prevent and respond to cyber risks. Based on these findings, the study maps regulatory and institutional obstacles and advances recommendations to clarify executive duties, strengthen governance structures, and refine accountability instruments, with a view to enhancing the maturity and cyber resilience of public organizations. Finally, the article outlines a research agenda to develop a conceptual frameworks and maturity indicators capable of objectively measuring the accountability of senior management in cybersecurity governance.

Keywords: Public governance; Cyber risk management; Organizational culture; Challenges; Regulatory compliance.

1. Introdução

A transformação digital e a crescente dependência de sistemas informacionais no setor público brasileiro elevaram a segurança da informação à condição de tema estratégico. O aumento de incidentes cibernéticos acompanha a expansão da conectividade, o que, somado à necessidade de proteção de dados pessoais e institucionais, impõe desafios significativos à gestão pública (Nakamura, 2024).

À luz desse contexto, a literatura aponta que a adoção efetiva de políticas e controles de segurança é condição para continuidade dos serviços, preservação da soberania digital e manutenção da confiança pública; tal necessidade se intensifica na ausência de coordenação centralizada e de estratégias integradas de prevenção e resposta a incidentes (Ferreira, 2022; Serpa, 2024).

A responsabilização da alta administração ainda é pouco abordada em normativos e práticas. Evidências indicam que a eficácia das políticas públicas depende de controles e sanções aplicáveis a todos os níveis hierárquicos, incluindo os cargos diretivos (Vatamanu & Tofan, 2025). Em complemento, estudos mostram que a falta de responsabilização compromete a resiliência organizacional frente a incidentes (Magnusson, Dalipi & Elm, 2023).

No Brasil, a situação se agrava pela baixa maturidade institucional e pela escassa presença de lideranças técnicas em segurança. Pesquisas nacionais evidenciam o distanciamento dos dirigentes máximos das decisões estratégicas de TI e a ausência de governança cibernética formal (Georg, Rodrigues, Alves, Silveira Júnior & Nunes, 2022; Santiago Paz, 2025).

A análise dos Acórdãos do TCU 2.387/2024 e 2.430/2024 explicita a necessidade de aperfeiçoamento normativo para atribuir, de forma clara, responsabilidades da alta administração na gestão de riscos cibernéticos (Tribunal de Contas da União, 2024). Identifica-se, assim, uma lacuna: embora existam pesquisas sobre desafios gerais e engajamento gerencial, ainda são escassos os estudos que aprofundam a responsabilidade

da alta administração na segurança do setor público, prevalecendo análises técnicas ou diagnósticos amplos.

Diante desse quadro, este estudo analisa barreiras à implementação da segurança da informação no serviço público brasileiro, com foco na responsabilização da alta gestão. Para isso, estrutura-se em cinco seções: esta introdução; o referencial teórico (Seção 2), com foco na responsabilização gerencial; a metodologia (Seção 3); a apresentação e análise dos resultados (Seção 4); e as considerações finais (Seção 5), com recomendações e perspectivas de pesquisa.

2. Referencial Teórico

2.1. Segurança da informação, privacidade, maturidade em segurança da informação e PPSI

A segurança da informação compreende políticas, normas e procedimentos que protegem dados contra acessos indevidos, alterações ou indisponibilidade, fundamentando-se nos princípios de confidencialidade, integridade e disponibilidade (ABNT, 2023). A cibersegurança, com foco mais restrito, visa à proteção de ativos digitais e infraestrutura contra incidentes no ciberespaço (ABNT, 2023).

Já a privacidade, direito fundamental à intimidade e vida privada, relaciona-se à proteção de dados pessoais, assegurando aos titulares controle sobre seu uso, conforme a LGPD (Brasil, 2018). A maturidade em segurança da informação envolve a avaliação do nível de proteção dos sistemas no contexto da privacidade e cibersegurança (Brasil - Secretaria de Governo Digital, 2023).

No setor público federal, mais especificamente no poder executivo, essa maturidade é medida pelo Programa de Privacidade e Segurança da Informação – PPSI –, conduzido pela Secretaria de Governo Digital (Brasil, 2023). Baseado no *CIS Critical Security Controls* (Center for Internet Security, s.d.), o PPSI orienta órgãos públicos a adotar inicialmente medidas do nível IG1, voltadas à proteção básica. Após atingi-lo, devem avaliar a criticidade de seus sistemas para definir a necessidade de avançar aos níveis IG2 e IG3 (Brasil, 2023).

2.2. Responsabilidade e responsabilização

No âmbito jurídico-administrativo brasileiro, responsabilidade é o dever legal do agente público no exercício de suas funções, enquanto responsabilização corresponde à apuração e eventual aplicação de sanções pelo descumprimento. A Lei nº 8.443/1992 exige nexo causal e dolo ou culpa para caracterizá-la (Presidência da República, 1992). O Decreto nº 9.203/2017 reforça essa noção ao atribuir à alta administração o dever de avaliar, direcionar e monitorar a governança (Brasil, 2017).

A alta administração abrange a autoridade máxima e os dirigentes superiores responsáveis por avaliar, direcionar e monitorar a atuação institucional (Tribunal de Contas da União, 2020). No Executivo federal, conforme o Decreto nº 9.203/2017, compõem esse grupo os Ministros de Estado, ocupantes de cargos de natureza especial ou de nível 6 do

DAS, além de presidentes e diretores de autarquias, fundações e entidades equivalentes (Brasil, 2017).

Nesse arranjo, suas atribuições conectam-se diretamente à formulação de políticas, à definição de objetivos e ao direcionamento estratégico das organizações. Em complemento, a governança pública estabelece o sistema pelo qual as instituições são dirigidas, monitoradas e avaliadas, integrando liderança, estratégias e mecanismos de controle para ampliar a efetividade de serviços e políticas (Tribunal de Contas da União, 2020).

A partir desse fundamento de governança, o *compliance* opera como o conjunto de práticas de conformidade a normas, regulamentos e códigos de conduta, voltadas à prevenção de desvios, fraudes e riscos reputacionais (Silva & Santos, 2020). Por sua vez, o *accountability* fecha o ciclo ao impor a obrigação de responder pelos atos administrativos (*answerability*) e ao viabilizar a aplicação de sanções diante do descumprimento de padrões de integridade (OECD, 2020).

2.3. Trabalhos Relacionados

A literatura sobre segurança da informação no setor público brasileiro aponta desafios multifatoriais que afetam diretrizes e rotinas administrativas. Normativos de cibersegurança e privacidade, pela alta abstração, frequentemente resultam em conformidade apenas formal (Nakamura, 2024). A LGPD exemplifica esse cenário, com aplicação lenta, fragmentada e em maturidade inicial (Brasil – Presidência da República, 2018; Serpa, 2024).

Somam-se a falta de políticas institucionais consolidadas e a baixa maturidade de governança, que fragmentam a organização e dificultam a integração entre riscos cibernéticos e corporativos (Santos & Silva, 2021; Alves, Georg & Nunes, 2022). O quadro inclui imaturidade na gestão de riscos digitais em várias esferas da administração pública (Ribeiro & Segatto, 2025).

No plano regulatório, o país acumulou leis sobre o ciberespaço com baixa coerência sistêmica e difícil implementação (Goldoni, Rodrigues & Medeiros, 2023). Persistem falhas de governança e falta de priorização pela alta gestão, refletidas em restrições de pessoal, orçamento e investimentos, o que fragiliza controles (Laginestra, 2021). Em paralelo, muitos órgãos operam com equipamentos defasados, softwares desatualizados e sem processos de TI padronizados, ampliando a superfície de ataque e vulnerabilidades (Alves, Georg & Nunes, 2022).

3. Metodologia

O presente estudo é de natureza qualitativo-exploratória, pois visa identificar os o problema e construir hipóteses a partir da análise da literatura científica (Gil, 2017). A abordagem é quantitativa, utilizando a Teoria do Enfoque Meta Analítico Consolidado (TEMAC) (Mariano & Rocha, 2017). O TEMAC estabelece três etapas claramente definidas:

1. Preparação da pesquisa consiste na construção dos termos (*string*) da pesquisa de busca com palavras-chave que abordem o tema de forma mais apropriada.

Também deve-se definir o campo espaço-tempo da pesquisa, as bases de dados e as áreas de conhecimento que serão utilizadas (Mariano & Rocha, 2017).

2. Apresentação e inter-relação dos dados: nesta etapa, utiliza-se bibliometria para análise relacional entre os registros, com ênfase nas revistas mais relevantes, evolução temática e autores mais citados (Mariano & Rocha, 2017).
3. Detalhamento, modelo integrador e validação por evidências: na terceira etapa são necessárias análises mais profundas que permitam compreender melhor o tema, como a identificação de coautoria e cocitação, as principais abordagens, linhas de pesquisas e comparação dos resultados entre as fontes (Mariano & Rocha, 2017).

4. Resultado e discussões

4.1. Etapa 1: Preparação da pesquisa

Esta etapa visa estabelecer os parâmetros da busca. A busca bibliográfica foi conduzida entre os dias 10/07/2025 e 25/07/2025. Usaram-se as bases Web of Science, reconhecida pela completude (Mariano & Rocha, 2017) e Scopus, pela abrangência e precisão (Adriaanse & Rensleigh, 2013).

Como resultado, foram encontrados 32 resultados no Web of Science e 58 no Scopus. Usaram-se termos em português e inglês para identificar estudos sobre responsabilização da alta administração na segurança da informação. A pesquisa não impôs restrições de período ou área e considerou juntos os termos “segurança da informação”, “cibersegurança” e “privacidade”, devido à sua interdependência conceitual.

Base de dados	Termos Pesquisados	Resultados
Web Of Science	TI=(“public sector” OR “government agencies” OR “public administration” OR “government institutions” OR “public service” OR “civil service” OR “municipal government” OR “federal government” OR “Brazilian public sector”) AND TS=(“cybersecurity” OR “information security” OR “IT security” OR “data protection” OR “information systems security”) AND TS=(“top management” OR “senior management” OR executives OR “organizational leadership” OR board OR governance OR accountability OR responsibility OR “strategic oversight” OR “management support” OR “executive commitment”)	32
Scopus	TITLE(“public sector” OR “government agencies” OR “public administration” OR “government institutions” OR “public service” OR “civil service” OR “municipal government” OR “federal government” OR “Brazilian public sector”) AND TITLE-ABS-KEY(“cybersecurity” OR “information security” OR “IT security” OR “data protection” OR “information systems security”) AND TITLE-ABS-KEY(“top management” OR “senior management” OR executives OR “organizational leadership” OR board OR governance OR accountability OR responsibility OR “strategic oversight” OR “management support” OR “executive commitment”)	58

Tabela 1 - Termos pesquisados na pesquisa avançada
Fonte: elaborada pelos autores

4.2.Etapa 2: Apresentação e inter-relação dos dados

Após remover duplicados, identificaram-se 73 registros para análise. A Figura 1 mostra evolução das publicações em WoS e Scopus, com crescimento após 2016 e maior volume no Scopus. Os picos de citações ocorreram em 2015 e 2023, impulsionados por marcos como as revelações de Snowden, o caso Cambridge Analytica e, no Brasil, a LGPD.

Ao reunir Scopus e WoS, identificaram-se os trabalhos mais citados (Tabela 2). Seus temas distintos mas complementares, abordam: governança de grandes dados e falhas de auditoria; riscos legais e responsabilização em decisões automatizadas; governança cibernética; relação governo eletrônico, cibersegurança e corrupção; e ameaças automatizadas à legitimidade democrática, com sugestões para mitigar esses riscos.

Segundo as bases, 43 países publicaram sobre o tema; a Tabela 3 mostra os 10 com mais trabalhos. O Brasil ocupa a 7ª posição, com 2 artigos. Um estudo (Georg, Rodrigues, Alves, Silveira Júnior & Nunes, 2022) destaca os desafios de segurança cibernética federal, com ausência de governança e distanciamento da alta gestão. Outro (Canedo et al., 2022) analisou a conformidade de um órgão federal com a LGPD, ressaltando a necessidade de envolvimento interdepartamental.

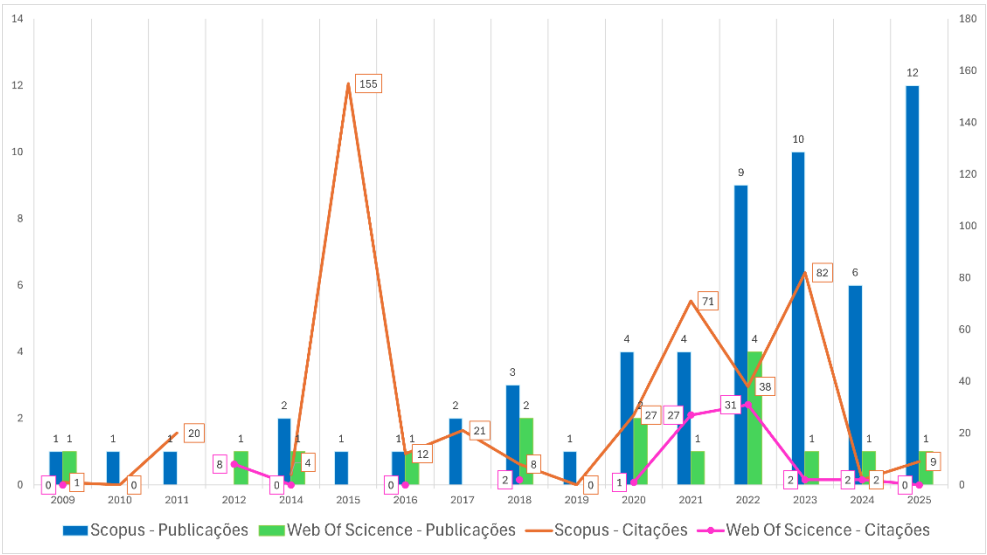


Figura 1 – Número de publicações e citações por ano

No	Título	Autores	Citações
1	Government data does not mean data governance: Lessons learned from a public sector application audit	(Thompson, Ravindran, & Nicosia, 2015)	155
2	Administrative due process when using automated decision-making in public administration: some notes from a Finnish perspective	(Suksi, 2021)	55

No	Título	Autores	Citações
3	The Purpose of Cybersecurity Governance in the Digital Transformation of Public Services and Protecting the Digital Environment	(Mijwil, Filali, Aljanabi, Bounabi, & Al-Shahwani, 2023)	37
4	Nexus of E-government, cybersecurity and corruption on public service (PSS) sustainability in Asian economies using fixed-effect and random forest algorithm	(Abbas, Qaisar, Xu, & Sun, 2022)	31
5	The legitimacy gap of algorithmic decision-making in the public sector: Why it arises and how to address it	(König & Wenzelburger, 2021)	27

Tabela 2 – Trabalhos mais citados

Nº	Países	Publicações	%
1	Indonésia	6	8
2	EUA	4	5
3	Malásia	3	4
4	Espanha	3	4
5	Reino Unido	3	4
6	Albânia	2	3
7	Brasil	2	3
8	China	2	3
9	República Tcheca	2	3
10	Finlândia	2	3
11	Outros	45	60
	Total	74	100

Tabela 3 – Número de Publicações por país

Quanto às revistas mais citadas (Tabela 4), destaca-se *Government Information Quarterly* com 155 citações. Em seguida estão *Artificial Intelligence and Law* (55) e *Mesopotamian Journal of CyberSecurity* (37). Já *Online Information Review* (31) e *Technology in Society* (27) mantêm relevância, embora em menor escala.

Fonte da Publicação	Citações	IF	SCR	SJR	Quartil
Government Information Quarterly	155	10	11,1	2,861	Q1
Artificial Intelligence and Law	55	3,1	-	0,815	Q1
Mesopotamian Journal of Cybersecurity	37	10,22	-	0,758	Q1
Online Information Review	31	3,5	3,7	1,025	Q1
Technology in Society	27	12,5	11,5	2,559	Q1

Tabela 4 – Revistas com maior número de citações

A Figura 2 apresenta a frequência das palavras-chave nos artigos selecionados, organizada em diagrama que realça, por meio de maior destaque visual, os termos mais recorrentes.



Figura 2 – Frequência de palavras-chave

Da nuvem foram excluídos os termos da estratégia de busca (Tabela 1). As palavras mais frequentes, em azul escuro, incluem “data”, “digital”, “governance”, “cloud” e “technology”, indicando eixos centrais das pesquisas.

4.3. Etapa 3: Detalhamento, modelo integrador e validação por evidências

Os dados das bases WoS e Scopus foram reunidos em base única e analisados com o VosViewer (Centre for Science and Technology Studies, Leiden University, 2023 v.1.6.20). O mapeamento bibliográfico (Figura 3) mostra um cluster denso entre Mijwil et al. (2023), Pakhnenko & Kuan (2023) e Barbala et al. (2023), indicando forte proximidade temática. Apesar de variados, esses trabalhos convergem em governança, ética e privacidade em contextos digitais. Os mapas de calor e redes de coautoria, Figura 4.

Aplicou-se filtro mínimo de três citações entre 151 autores destacou 59 de maior impacto, incluindo Afrune, Paolo e Angelleli, Mario, ligados à University of Salento (Catalano et al., 2021). Outro grupo importante está ligado à Holistic AI e UCL, com integrantes como Chaudhry A. e Kazim E. (Kingsman et al., 2021).

Figura 5 apresenta mapa de cocitação com filtro de cinco citações, revelando baixa interconexão e nós dispersos. Destaca-se Balaji, K. (Balaji, 2025), em área de maior intensidade, enquanto Mijwil et al. (Mijwil et al., 2023) aparecem em regiões isoladas. A ausência de agrupamentos sugere pouca articulação e diversidade temática.

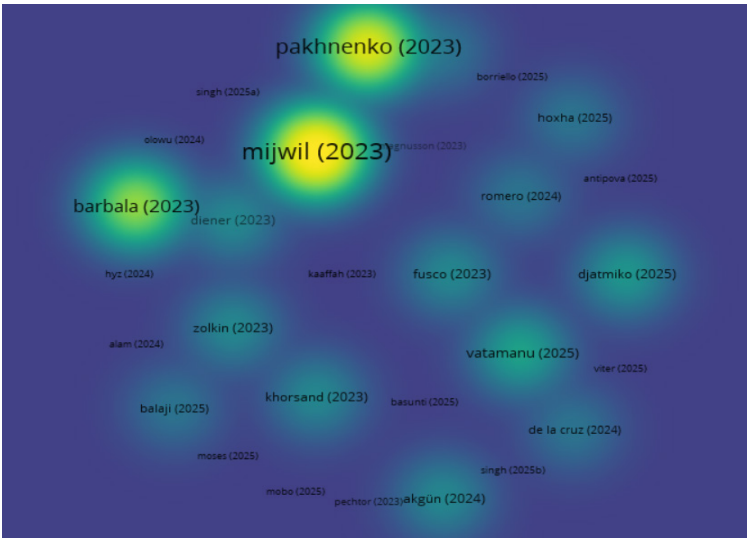


Figura 3 – Mapa de densidade de bibliographic coupling

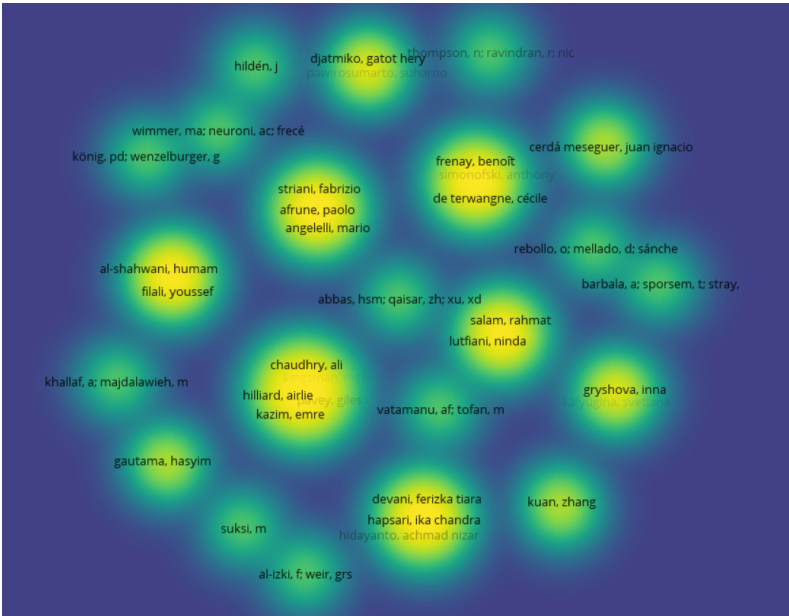


Figura 4 – Mapa de densidade de Coautoria

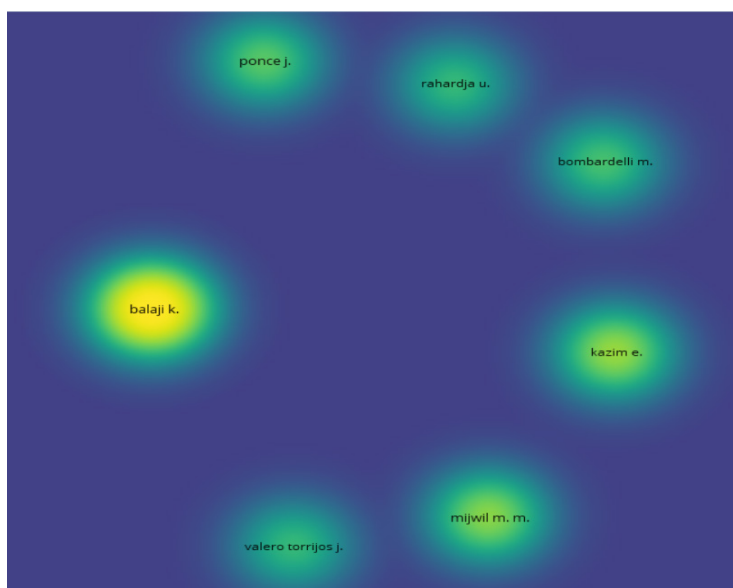


Figura 5 – Mapa de densidade de Cocitação

4.4. Governança e *accountability* em perspectivas internacionais

A literatura internacional mostra a ausência de atribuição clara de responsabilidade à alta administração em segurança da informação, aspecto também presente no Brasil. Mesmo em países com maior maturidade digital, prevalece abordagem genérica sobre deveres e mecanismos de prestação de contas (Romanovská & Pitner, 2022).

Na Austrália, a estratégia nacional não inclui governos locais e aborda superficialmente estados e territórios; na Bélgica, a indefinição regional reforça fragilidade de governança. Os autores sugerem um framework multinível para distribuir responsabilidades entre esferas governamentais e ampliar a capacidade de resposta.

Pesquisas em municípios suecos relacionam falhas de governança a incidentes como Equifax, Capital One e SolarWinds, que afetaram diretamente a resiliência organizacional (Magnusson, Dalipi & Elm, 2023). Para Suksi (2021), é indispensável definir de forma inequívoca a responsabilidade legal de decisores, inclusive em sistemas automatizados.

No Brasil, o debate surge em propostas como o PL nº 2.338/2023 (Brasil, 2023), que institui o Marco Legal da Inteligência Artificial, e o PL nº 2.630/2020 (Brasil, 2020), voltado às redes sociais.

Para Vatamanu & Tofan (2025), estruturas de compliance e *accountability* são essenciais à transparência e ao uso responsável da IA alcançando, inclusive, a alta gestão.

4.5. Desafios de implementação: Auditorias, maturidade e diagnóstico brasileiro

Embora normas alinhadas a padrões internacionais existam, falhas de implementação prejudicam a responsabilização e o controle dos ativos de informação. Auditorias australianas indicaram falta de rastreabilidade em dados sensíveis, mais devido à ausência de governança e accountability do que a limitações técnicas (Thompson, Ravindran & Nicosia, 2015).

No Brasil, a governança cibernética ainda é incipiente. Dirigentes permanecem afastados das decisões de TI (Georg, Rodrigues, Alves, Silveira Júnior & Nunes, 2022).

O BID registra que 30% dos governos estaduais não dispõem de analistas de cibersegurança e apenas 22% contam com um CISO, demonstrando fragilidade na liderança institucional (Santiago Paz, 2025).

De acordo com o Tribunal de Contas da União (2024), o SISP também enfrenta governança limitada. Nenhum dos 229 órgãos avaliados implementa integralmente o IG1; 14 aplicam mais de 70% e duas excedem 90%.

4.6. Intervenções regulatórias e recomendações estratégicas

Um estudo do leste europeu defende a criação de um sistema de proteção cibernética no setor público, alinhado à OTAN e à União Europeia, abrangendo todos os níveis da administração (Viter et al., 2025). Para ser efetivo, exige reformas normativas e transformações na cultura gerencial, com destaque para mecanismos de responsabilização.

Segundo os autores, é necessário “implementar controle e responsabilização por violações à segurança da informação em todos os níveis da administração pública”. A ausência de clareza normativa permanece como entrave central, mantendo riscos como corrupção e perda de confiança.

No Brasil, auditorias do TCU apontam fragilidades semelhantes. Os Acórdãos 2.387/2024 e 2.430/2024 indicam insuficiências normativas que dificultam responsabilizar gestores de alto escalão (Tribunal de Contas da União, 2024). O primeiro evidencia a falta de atribuição legal sobre gestão de riscos cibernéticos; o segundo mostra que a Política Nacional de Cibersegurança carece de coordenação eficaz.

Na Consulta nº 382507/2025, o TCU reforçou que a responsabilização depende da Lei nº 8.443/1992 (Presidência da República, 1992), que exige nexo causal, dolo ou culpa. Recomendou ainda explicitar o papel da alta gestão no PPSI (Brasil, 2023) e no SISP (Brasil, 2011).

4.7. Engajamento da alta administração e cultura organizacional

O estudo de De La Cruz et al. (2024) indica que o engajamento da alta administração é determinante para o êxito de sistemas de análise de dados em cibersegurança no setor público, pois o apoio da liderança — priorização, endosso e alocação de recursos — integra a segurança digital à estratégia e à cultura organizacional.

Em consonância, a literatura ressalta que alinhamento e recursos, por si, são insuficientes; é o envolvimento persistente da alta gestão, ao longo de todo o ciclo de políticas — formulação, sustentação, monitoramento e aprimoramento — que institucionaliza a segurança da informação como valor organizacional. Por fim, reforça-se a necessidade de governança contínua conduzida por gestores experientes, uma vez que fragilidades estruturais reduzem a efetividade das políticas e comprometem respostas coordenadas a ameaças (Mijwil et al., 2023).

5. Conclusão

Este artigo analisou as barreiras à implementação da segurança da informação no serviço público brasileiro, com foco na responsabilização da alta gestão, utilizando a Teoria do Enfoque Meta-analítico Consolidado (TEMAC) como método de revisão sistemática. A abordagem permitiu identificar padrões recorrentes e seus impactos.

Constatou-se que, no contexto internacional, a responsabilização da alta gestão segue abstrata e pouco vinculada a mecanismos legais. O *accountability* é reconhecido como elemento essencial de ética e transparência, devendo abranger todos os níveis hierárquicos. A falta de mecanismos formais expõe riscos institucionais como corrupção, mau uso de recursos e fragilidade de controles.

No Brasil, observou-se baixa participação da alta administração em decisões de segurança, ausência de lideranças técnicas e percepção limitada sobre sua relevância estratégica. Acórdãos do TCU evidenciam fragilidades normativas que dificultam a responsabilização direta de gestores, sugerindo a necessidade de explicitar atribuições e exigir liderança ativa da alta gestão na condução dos riscos cibernéticos, fortalecendo a governança pública.

O estudo contribui para mapear entraves normativos e institucionais ligados à responsabilidade da alta gestão em segurança da informação, oferecendo subsídios para pesquisas sobre modelos regulatórios e de conformidade. Reconhece-se, contudo, a limitação da amostra, restrita a periódicos indexados em Scopus e WoS.

Como agenda futura, sugerem-se estudos de caso longitudinais para avaliar o impacto da responsabilização executiva sobre a maturidade em segurança da informação, além da criação de um *framework* que defina inequivocamente os deveres da alta administração. Complementarmente, a criação de indicadores, poderia ser usada como forma de medição de aderência ao framework. Dentre outros aspectos poderia ser medidos o grau de explicitação de responsabilidade da alta gestão nos normativos, regularidade de sua participação em decisões de segurança e incorporação de metas de cibersegurança em suas avaliações de desempenho.

Referências

- ABNT. (2023). ABNT NBR ISO/IEC 27032:2023 — Segurança cibernética — Diretrizes para segurança na Internet. Rio de Janeiro: ABNT.
- ABNT. (2023). ISO/IEC 27000:2023 — Information technology — Security techniques — Information security management systems — Overview and vocabulary. Geneva: ISO.

- Adriaanse, L., & Rensleigh, C. (2013). Web of Science, Scopus and Google Scholar: A content comprehensiveness comparison. doi:10.1108/EL 12 2011 0174
- Alves, R. S., Georg, M. A., & Nunes, R. R. (2023). Judiciário sob ataque hacker: riscos de negócio para segurança cibernética em tribunais brasileiros. *Revista Ibérica de Sistemas e Tecnologias de Informação*. doi:10.5281/zenodo.8032915.
- Balaji, K. (2025). E-government and e-governance: Driving digital transformation in public administration. Em I. Global (Ed.), *Public Governance Practices in the Age of AI* (pp. 23-44). doi:10.4018/979-8-3693-9286-7.ch002
- Barbala, A., Sporse, T., & Stray, V. (2023). Data-Driven Development in Public Sector: How Agile Product Teams Maneuver Data Privacy Regulations. 24th International Conference on Agile Software Development (XP 2023). 475, pp. 165-180. Amsterdam, Netherlands: Springer International Publishing AG. doi:10.1007/978-3-031-33976-9_11.
- Brasil. (2018). Lei nº 13.709, de 14 de agosto de 2018 – Lei Geral de Proteção de Dados Pessoais (LGPD). Fonte: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm
- Brasil. (2023). Guia de Referência para Elaboração do Plano de Proteção de Segurança da Informação – PPSI. Brasília. <http://bit.ly/4nQOGM>.
- Brasil. (2011). Decreto nº 7.579, de 11 de outubro de 2011. Fonte: Planalto: https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/decreto/d7579.htm.
- Brasil. (2017). Decreto nº 9.203, de 22 de novembro de 2017. Dispõe sobre a política de governança da administração pública federal direta, autárquica e fundacional. Fonte: <http://bit.ly/4pDVNSB>.
- Brasil. (2020). Projeto de Lei nº 2.630, de 2020 – Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet. Câmara dos Deputados, Brasília. Fonte: <http://bit.ly/4pIkeyj>.
- Brasil. (2023). Projeto de Lei nº 2.338, de 2023 – Marco Legal da Inteligência Artificial. Câmara dos Deputados, Brasília. Fonte: <https://www25.senado.leg.br/web/atividade/materias/-/materia/157233>.
- Brasil. Ministério da Gestão e da Inovação em Serviços Públicos. (2023). SGD/MGI nº 852, de 28 de março de 2023. Fonte: Governo Federal: <http://bit.ly/3IuPYGz>.
- Canedo, E. D., Ribeiro, V. C., Cerqueira, A. J., Gravina, R. M., Camões, R., Dos Reis, V. E., De Sousa, R. T. (2022). Evaluating and Evolving the Compliance to the Brazilian General Data Protection Law in a Federal Government Agency. *Lecture Notes in Business Information Processing*, 455. doi:10.1007/978-3-031-08965-7_1.
- Catalano, C., Afrune, P., Angelelli, M., Maglio, G., Striani, F., & Tommasi, F. (2021). Security testing reuse enhancing active cyber defence in public administration., (pp. 120-132). Virtual, Online.

- Center for Internet Security. (s.d., s.d., acesso em 2025). CIS Critical Security Controls. Fonte: Center for Internet Security: <https://www.cisecurity.org/controls>.
- Centre for Science and Technology Studies, Leiden University. (2023 v.1.6.20). VOSviewer: Visualizing Scientific Landscapes. Fonte: <https://www.vosviewer.com/>.
- De La Cruz, E., Gonaygunta, H., Oni, O., Meduri, S. S., Nadella, G. S., & De La Cruz, A. M. (2024). Cybersecurity Data Analytics System Success: An Exploratory Study on U.S. Government Agencies. Preprint – ResearchGate. doi:10.1109/iSemantic63362.2024.10762173.
- Ferreira, L. V. (2022). Cybersecurity and Ransomware in the Brazilian Government. *InterAção*, v. 15, pp. 107–123.
- Georg, M. A., Rodrigues, W. M., Alves, C. A., Silveira Júnior, A., & Nunes, R. R. (nov de 2022). Os desafios da Segurança Cibernética no setor público federal do Brasil: estudo sob a ótica de gestores de tecnologia da informação. *Revista Ibérica de Sistemas e Tecnologias de Informação*, pp. 602-616. doi:10.5281/zenodo.7855320.
- Gil, A. C. (2017). *Métodos e técnicas de pesquisa social* (6 ed.). São Paulo: Atlas.
- Goldoni, L. R., Rodrigues, K. F., & Medeiros, B. P. (2023). Qual é o futuro da governança de cibersegurança no Brasil? *SciELO*. doi:10.12660/cgpc.v29.90972
- Kingsman, N., Kazim, E., Chaudhry, A., Hilliard, A., Koshiyama, A., Polle, R., . . . Mohammed, H. U. (2021). Exploring the Governance of Algorithmic Decision-Making: A Multi-Stakeholder Perspective on Auditability. 2021 AAAI/ACM Conference on AI, Ethics, and Society (AIES 2021), (pp. 162-172). Virtual, online.
- Laginestra, A. (2021). *Cibersegurança no Território Brasileiro: Impacto da LGPD e normas de Segurança da Informação na Defesa Nacional* (1 ed.). Rio de Janeiro: Biblioteca de Segurança.
- Magnusson, L., Dalipi, F., & Elm, P. (2023). Cybersecurity compliance in the public sector: are the best security practices properly addressed? *Em Communications in Computer and Information Science* (pp. 370–384). doi: https://doi.org/10.1007/978-3-031-36001-5_28.
- Mariano, A. M., & Rocha, M. S. (2017). Revisão da Literatura: Apresentação de uma Abordagem Integradora. XXVI Congresso Internacional AEDEM | 2017 AEDEM International Conference - Economy, Business and Uncertainty: ideas for a European and Mediterranean industrial policy? (pp. 427-443). Reggio Calabria: AEDEM.
- Mijwil, M. M., Filali, Y., Aljanabi, M., Bounabi, M., & Al-Shahwani, H. (2023). The Purpose of Cybersecurity Governance in the Digital Transformation of Public Services and Protecting the Digital Environment. (M. A. Press, Ed.) *Mesopotamian Journal of CyberSecurity*, pp. 1-6. doi:10.58496/MJCS/2023/001.

- Nakamura, E. T. (2024). O papel da segurança cibernética no universo digital: a importância do fator humano. Rio de Janeiro: Instituto de Pesquisa Econômica Aplicada (Ipea). doi:<http://dx.doi.org/10.38116/9786556350660cap9>.
- OECD – Organisation for Economic Co-operation and Development. (2020). OECD Public Integrity Handbook. Paris: OECD Publishing.
- Pakhnenko, O., & Kuan, Z. (2023). Ethics of Digital Innovation in Public Administration. *Business Ethics and Leadership*, pp. 113-121. doi: 10.21272/bel.7(1).113-121.2023.
- Presidência da República. (1992). Lei nº 8.443, de 16 de julho de 1992. Fonte: Lei Orgânica do Tribunal de Contas da União: https://www.planalto.gov.br/ccivil_03/leis/l8443.htm.
- Ribeiro, M. M., & Segatto, C. I. (2025). Inteligência artificial nas organizações públicas brasileiras: heterogeneidades e capacidades em tecnologia da informação. *Revista de Administração Pública (RAP)*,. doi:10.1590/0034-761220240066x.
- Romanovská, F., & Pitner, T. (2022). Multi-Level Cybersecurity Governance Frameworks for Public Administration. *proceedings of IDIMT-2022*, pp. 277-284. doi:DOI: 10.35011/IDIMT-2022-277.
- Santiago Paz, M. L. (2025). Cibersegurança nos estados brasileiros: diagnóstico e recomendações. Banco Interamericano de Desenvolvimento.
- Santos, R. B., & Silva, T. B. (2021). Gestão da segurança da informação e comunicações: análise ergonômica para avaliação de comportamentos inseguros. *Revista Brasileira de Segurança da Informação*. doi:10.20396/rdbci.v19i00.8665529.
- Serpa, D. (2024). Transformação digital da inteligência nacional brasileira. *Revista Brasileira de Inteligência (ABIN)*.
- Silva, M. A., & Santos, F. R. (2020). Fundamentos da Governança, Riscos e Compliance. São Paulo: Atlas.
- Suksi, M. (2021). Administrative due process when using automated decision-making in public administration: some notes from a Finnish perspective. *Artificial Intelligence and Law*, 29, pp. 87-110. doi:10.1007/s10506-020-09269-x.
- Thompson, N., Ravindran, R., & Nicosia, S. (2015). Government data does not mean data governance: Lessons learned from a public sector application audit. *Government Information Quarterly*, 32, pp. 316-322. doi:10.1016/j.giq.2015.05.001.
- Tribunal de Contas da União. (2020). Referencial Básico de Governança Organizacional: aplicações ao setor público. TCU, Brasília. Fonte: <http://bit.ly/4pFhZvT>.
- Tribunal de Contas da União. (2024). Acórdão nº 2387/2024 – Plenário. Brasília: Tribunal de Contas da União. Fonte: <http://bit.ly/46Is3eT>
- Tribunal de Contas da União. (2024). Acórdão nº 2430/2024 – Plenário. Brasília: Tribunal de Contas da União.

- Tribunal de Contas da União. (2025). Resposta à manifestação nº 382507 enviada à Ouvidoria do TCU pelos autores. Brasília: TCU. Fonte: <https://pesquisa.apps.tcu.gov.br/pesquisa/acordao-completo>.
- Vatamanu, A. F., & Tofan, M. (2025). Integrating Artificial Intelligence into Public Administration: Challenges and Vulnerabilities. *Administrative Sciences*, 15. doi:10.3390/admsci15040149.
- Viter, D., Guzonova, V., Shevchuk, V., Molochko, T., & Kutova, M. (2025). Mechanisms Of Public Administration To Ensure National And Information Security. *Management Theory and Studies for Rural Business and Infrastructure Development*, 47(2), 260-269.

Desafios e Oportunidades na Implantação de Sistemas Embarcados para o Controle do Ambiente em Aviários: Uma Revisão Sistemática de Literatura

Daniel de Oliveira Cunha¹, Leonardo da Silva Lisboa¹,
Marcos do Livramento Amaral¹, Uenderson Mendes do Nascimento¹,
Felipe Gonçalves dos Santos¹, Wanderson Jean Conceição Silva¹

cacor.2023121tads00030@aluno.ifpi.edu.br; cacor.2022121lfs00002@aluno.ifpi.edu.br;
cacor.2022121tads00015@aluno.ifpi.edu.br; cacor.2021121tads00031@aluno.ifpi.edu.br;
felipe.santos@ifpi.edu.br; wanderson.jean@ifpi.edu.br

¹ Instituto Federal do Piauí do Piauí (IFPI) – Campus Corrente, 64980-000, Corrente, Brasil

Pages: 114-128

Resumo: O controle do ambiente em instalações avícolas é uma variável crucial que determina a produtividade e o bem-estar das aves. Este processo é complexo, enfrentando desafios operacionais como as variações climáticas, a alta densidade populacional e a presença corrosiva de poeira e amônia, que afetam a saúde dos animais e a vida útil dos equipamentos. A adoção de sistemas embarcados e Internet das Coisas (IoT) emerge como uma solução tecnológica promissora, oferecendo monitoramento contínuo, automação de ajustes e integração de dados em redes de comunicação. Esta revisão sistemática de literatura (RSL) analisou estudos publicados entre 2020 e 2025 para mapear os principais desafios do controle ambiental, os benefícios da tecnologia embarcada, a adequação de sensores específicos e as ferramentas de desenvolvimento mais utilizadas. Os resultados indicam um campo em rápida evolução, com um aumento significativo de publicações. No entanto, persistem lacunas em áreas como a durabilidade dos sensores em ambientes hostis, a padronização de protocolos de comunicação e a avaliação de soluções em escala industrial. O artigo conclui que o avanço e a validação de tecnologias mais robustas e integradas são fundamentais para maximizar a produtividade, a sustentabilidade e os padrões de bem-estar animal na avicultura moderna.

Palavras-chave: avicultura; controle ambiental; sistemas embarcados; sensores; automação.

Challenges and Opportunities in the Implementation of Embedded Systems for Environmental Control in Poultry Houses: A Systematic Literature Review

Abstract: Environmental control in poultry facilities is a crucial factor that determines productivity and animal welfare. This process is complex, facing operational challenges such as climate variability, high stocking density, and the

corrosive presence of dust and ammonia, which affect both animal health and the lifespan of equipment. The adoption of embedded systems and the Internet of Things (IoT) emerges as a promising technological solution, offering continuous monitoring, automated adjustments, and data integration through communication networks. This systematic literature review (SLR) analyzed studies published between 2020 and 2025 to map the main challenges of environmental control, the benefits of embedded technologies, the suitability of specific sensors, and the most commonly used development tools. The results indicate a rapidly evolving field, with a significant increase in publications. However, gaps remain in areas such as sensor durability in harsh environments, the standardization of communication protocols, and the evaluation of solutions at an industrial scale. The article concludes that the advancement and validation of more robust and integrated technologies are essential to maximize productivity, sustainability, and animal welfare standards in modern poultry farming.

Keywords: poultry farming; environmental control; embedded systems; sensors; automation.

1. Introdução

O setor da avicultura moderna está sob constante pressão para maximizar a produtividade e a qualidade do produto, mantendo elevados padrões de bem-estar animal em um mercado global altamente competitivo. Neste cenário, a gestão precisa do ambiente interno dos aviários é um fator decisivo. Variáveis como **temperatura, umidade relativa, ventilação, iluminação e qualidade do ar** (especialmente a concentração de gases nocivos) influenciam diretamente a taxa de crescimento das aves, a conversão alimentar e sua saúde geral (Pereira et al., 2020). Contudo, a manutenção de condições ambientais ideais é dificultada por fatores externos (variações climáticas) e internos (alta densidade populacional, acúmulo de poeira e liberação de amônia), que podem degradar o ambiente e comprometer a longevidade dos equipamentos (Terence et al., 2024; Rathod et al., 2022).

Historicamente, o controle ambiental tem sido realizado por métodos manuais ou por sistemas de automação simples e reativos. A emergência da **tecnologia embarcada**, aliada à **Internet das Coisas (IoT)**, transformou esse panorama. A integração de microcontroladores de baixo custo, sensores inteligentes e atuadores permite o **monitoramento em tempo real** e o **controle automático** das condições (Amr et al., 2022). Tais sistemas fornecem aos produtores dados precisos e acionáveis, facilitando decisões rápidas e eficientes. Além de otimizar as condições de criação, essas inovações têm o potencial de aumentar a produtividade, reduzir o desperdício de recursos e otimizar a gestão, fortalecendo a competitividade do setor (Fahrurrozi et al., 2024; Putri et al., 2024).

Neste contexto de rápida evolução tecnológica, torna-se essencial mapear o conhecimento acumulado. Esta revisão sistemática da literatura (RSL) visa: (1) Identificar os desafios mais persistentes no controle ambiental de aviários; (2) Avaliar os benefícios tangíveis da implementação de sistemas embarcados; (3) Determinar quais sensores são mais adequados para ambientes severos; e (4) Explorar as plataformas

de desenvolvimento e linguagens de programação mais apropriadas para a área. O trabalho se propõe a oferecer um panorama aprofundado e atualizado (2020-2025) que sirva de subsídio para novas pesquisas e para a validação de soluções tecnológicas robustas no setor avícola.

2. Questões de Pesquisa

Este estudo tem como objetivo central investigar a aplicação de sistemas embarcados no manejo avícola, buscando responder à questão principal (QP): O que a literatura revela sobre a aplicação de sistemas embarcados no manejo de aves?

A partir desta, definiram-se as seguintes questões secundárias para guiar a análise dos artigos:

- **QP1:** Qual a quantidade de artigos publicados por ano sobre esse assunto?
- **QP2:** Quais países têm se destacado na produção científica sobre o uso de sistemas embarcados nas áreas de saúde, segurança e criação animal?
- **QP3:** Quais eventos científicos e periódicos concentram a maior parte das publicações sobre sistemas embarcados aplicados a essas três áreas?
- **QP4:** Quais os desafios no controle ambiental são mais recorrentes em aviários?
- **QP5:** A implementação de sistemas embarcados em aviários pode trazer benefícios na produção e no bem-estar animal?
- **QP6:** Quais sensores de temperatura são mais precisos e duráveis em ambientes com alta concentração de poeira e amônia?
- **QP7:** Quais ferramentas de desenvolvimento e linguagens são mais adequadas para sistemas embarcados em aviários?

3. Seleção dos Artigos

A seleção de artigos relevantes seguiu um rigoroso processo de Mapeamento Sistemático da Literatura (MSL), com foco em garantir a pertinência e a qualidade técnica dos estudos.

3.1. Bases de Dados e String de Busca

As bases de dados Scopus, Web of Science (WoS) e ACM Digital Library foram escolhidas pela sua relevância na indexação de publicações em Engenharia, Computação, IoT e Sistemas Embarcados.

A string de busca inicial foi refinada com o auxílio da ferramenta VOSviewer, que permitiu identificar novas associações semânticas e palavras-chave emergentes. O resultado final desse processo resultou na seguinte string aprimorada:

“IoT” OR “Internet of Things” OR “Embedded Systems” OR “Embedded Technology” OR “Embedded Computing” OR “Real-Time Systems” AND (“Animal Monitoring” OR “Livestock Tracking” OR “Poultry Farming” OR “Smart Farming” OR “Livestock Management” OR “Precision Agriculture” OR “AgriTech” OR “Smart Agriculture” OR “Agribusiness” OR “Digital Farming” OR “Agricultural Technology”)

3.2. Critérios de Inclusão e Exclusão

Para manter a relevância técnica da pesquisa e focar estritamente na aplicação de engenharia e ciência da computação, foram aplicados os seguintes critérios:

Critérios de Inclusão (CI):

- CI1: Os artigos devem estar relacionados exclusivamente às áreas de computação e engenharia.
- CI2: Os artigos devem ter data de publicação entre os anos de 2020 e 2025.
- CI3: Os artigos devem estar escritos em inglês.
- CI4: Os artigos devem ser do tipo “Journal Article” (Artigo de Periódico) ou “Conference Paper” (Artigo de Conferência).

O processo de triagem inicial, realizado entre 1 e 10 de junho de 2025, recuperou um total de 1.703 registros. Após a remoção de **854 duplicatas**, o processo de filtragem seguiu dois ciclos com critérios de exclusão:

3.2.1. Primeiro Ciclo

CE1: Descarte de artigos não relacionados a computação/engenharia, que abordavam apenas agricultura (plantio/irrigação), ou que não envolviam o uso de sistemas baseados em IoT/embarcados.

Resultado: O ciclo foi concluído com 136 artigos.

3.2.2. Segundo Ciclo

CE2: Exclusão de estudos que não mencionavam o controle de temperatura, amônia e umidade no ambiente de criação de animais.

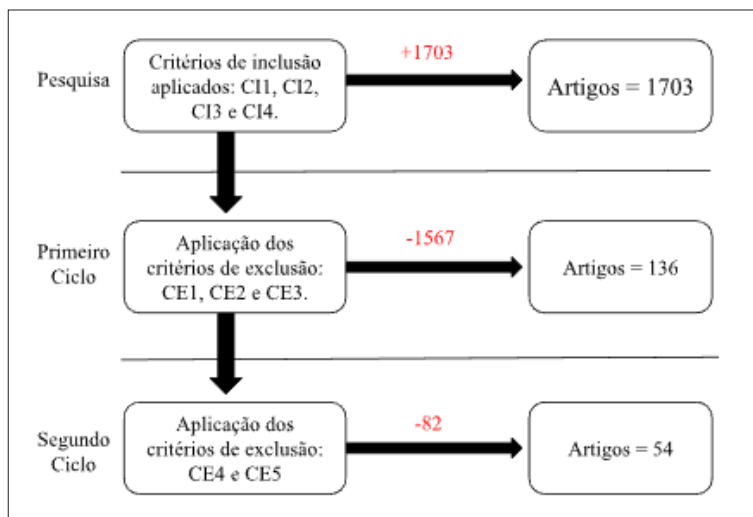


Figura 1 – Distribuição da Quantidade de Artigos por Base de Dados (Incluir a Tabela I aqui: Scopus 830, WoS 853, ACM 20. Total: 1703)

CE3: Exclusão de estudos que utilizavam processamento de imagem para outros fins (ex.: lavouras, ambientes urbanos, etc.) sem relação direta com a ambiência animal.

Resultado: 82 artigos foram descartados. A amostra final resultou em 54 artigos que tratam do tema de pesquisa proposto.

4. Resultados e Análise

Esta seção detalha a análise dos 54 artigos selecionados, com foco na resposta às questões de pesquisa (QP1 a QP7), fornecendo uma base robusta para a discussão.

4.1. QP1: Tendência Temporal da Produção Científica

A Figura 1 apresenta a distribuição da quantidade de artigos publicados entre os anos de 2021 e 2025. Observa-se um crescimento entre 2021 e 2022, seguido por uma queda expressiva em 2023 e uma recuperação significativa em 2024 e 2025.

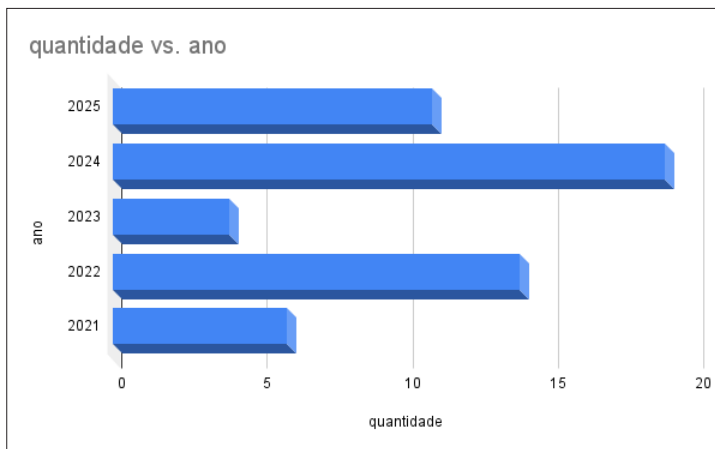


Figura 2 – Quantidade de artigos publicados por ano

Os dados mostram que, apesar da queda acentuada em 2023, houve um crescimento expressivo no ano seguinte, atingindo o pico em 2024. Em 2025, até o momento da coleta, o número de publicações se mantém elevado, o que pode indicar a continuidade do interesse e da relevância do tema na comunidade científica.

4.2. QP2: Quais países têm se destacado na produção científica sobre o uso de sistemas embarcados nas áreas de saúde, segurança e criação animal?

apresenta a distribuição geográfica das publicações relacionadas ao tema, permitindo visualizar, de forma espacial, a presença e a intensidade da pesquisa em diferentes regiões do mundo. Essa representação cartográfica é uma ferramenta valiosa para compreender como o conhecimento científico na área de sistemas embarcados está sendo produzido e difundido globalmente.

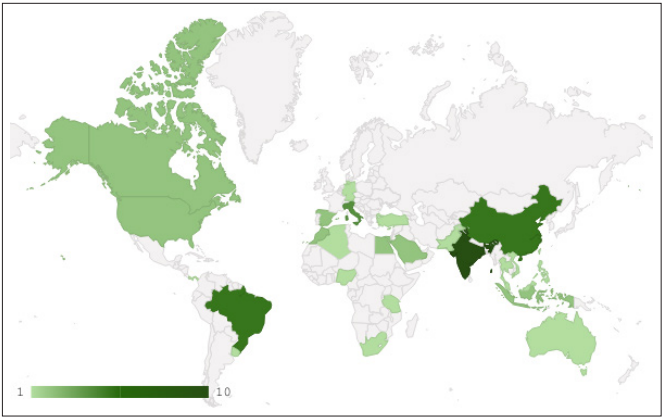


Figura 3 – Quantidade de artigos publicados por ano

A **Ásia** é o continente com maior concentração de estudos, destacando-se a **Índia** e a **China**. Na América, o **Brasil** é o país de maior destaque, seguido pelos polos tradicionais dos **Estados Unidos** e **Canadá**. A Tabela I quantifica essa liderança.

País	Quantidade
Índia	10
Brasil	5
China	5
Itália	4
Estados Unidos	2
Canadá	2
Egito	2
Indonésia	2
Marrocos	2
Espanha	2
Arábia Saudita	2
Vietnã	1
Turquia	1
Uruguai	1
Paquistão	1
Filipinas	1
África do Sul	1
Panamá	1
Taiwan	1
Malásia	1
Argélia	1

País	Quantidade
<i>Tailândia</i>	1
<i>Tanzânia</i>	1
<i>Austrália</i>	1
<i>Nigéria</i>	1
<i>Alemanha</i>	1
<i>Chipre</i>	1

Tabela 1 – Quantidade de artigos publicados por País

A liderança da Índia (10 artigos), seguida por Brasil e China (5 artigos cada), sugere que o investimento em tecnologia embarcada é uma prioridade para economias que buscam modernizar e otimizar o agronegócio em larga escala. A presença de países emergentes e em desenvolvimento evidencia que a tecnologia de baixo custo (como IoT) é crucial para a superação de barreiras infraestruturais.

4.3.QP3: Quais eventos científicos e periódicos concentram a maior parte das publicações sobre sistemas embarcados aplicados a essas três áreas?

A Tabela 2 lista os principais veículos de disseminação do conhecimento, reforçando o caráter interdisciplinar do tema, que transita entre Engenharia e Agricultura de Precisão.

Periódico	Quantidade
<i>Sensors</i>	5
<i>Computers and Electronics in Agriculture</i>	3
<i>Scalable Computing: Practice and Experience</i>	3
<i>Internet of ThingSs (Netherlands)</i>	2
<i>Indonesian Journal of Electrical Engineering and Computer Science</i>	2

Tabela 2 – Periódicos Científicos com Mínimo de Duas Publicações.

O periódico *Sensors* se destaca como o principal canal, seguido por *Computers and Electronics in Agriculture* e *Scalable Computing: Practice and Experience*. A proeminência dessas revistas sinaliza que as pesquisas se concentram em soluções práticas que envolvem a coleta robusta de dados (sensores) e a aplicação de tecnologias de processamento de dados (computação escalável) no contexto agropecuário.

4.4.QP4: Quais os desafios no controle ambiental mais recorrentes em aviários?

Os artigos são unânimes em apontar que os principais desafios no controle ambiental em aviários estão relacionados à natureza corrosiva e dinâmica do ambiente (Balthazar et al., 2024; Dollente & Hanbal, 2024; Pereira et al., 2020):

1. **Alta Concentração de Poeira e Amônia (NH₃):** Esses elementos são citados como os maiores obstáculos operacionais. Eles afetam diretamente a saúde das aves e, crucialmente, a durabilidade e precisão dos sensores, exigindo equipamentos mais robustos e manutenção constante (Terence et al., 2024).
2. **Variações Climáticas e Estresse Térmico:** A manutenção de um ambiente térmico ideal é essencial, mas a rápida oscilação de temperatura e umidade demanda uma resposta automática e em tempo real dos sistemas embarcados para evitar perdas produtivas e o estresse das aves (Chung et al., 2020).
3. **Conectividade em Ambientes Fechados:** Estruturas metálicas e ambientes fechados comprometem a transmissão de dados, afetando a confiabilidade do sistema (Tran et al., 2024). Isso exige o uso de protocolos de comunicação otimizados e uma distribuição estratégica dos pontos de coleta de dados no galpão.

4.5.QP5: A implementação de sistemas embarcados em aviários pode trazer benefícios na produção e no bem-estar animal?

A literatura confirma que a adoção de sistemas embarcados resulta em ganhos multifacetados, transformando a avicultura de forma positiva (Si et al., 2022). Os principais benefícios incluem:

- **Automação e Eficiência:** O monitoramento contínuo de variáveis e a automação de ajustes (ventilação, aquecimento) reduzem o erro humano e aumentam a eficiência geral da produção (Fahrurrozi et al., 2024; Song et al., 2024).
- **Melhoria do Bem-Estar Animal e Produtividade:** A manutenção de um ambiente estável minimiza o estresse térmico e melhora a conversão alimentar, resultando em aves mais saudáveis e maior lucratividade (Bhaskaran et al., 2024).
- **Gestão Baseada em Dados:** Sistemas integrados a IoT permitem decisões preditivas, possibilitando a antecipação de problemas (preventivo) e a redução de custos operacionais através da otimização do uso de energia e insumos (Guo et al., 2024; Rathod et al., 2022).

4.6.QP6: Quais sensores de temperatura são mais precisos e duráveis em ambientes com alta concentração de poeira e amônia?

A escolha de sensores que combinam precisão e durabilidade em ambientes com alta poeira e amônia é crucial. Os estudos destacam que o sensor **HDC1080** se sobressai pela precisão na medição de temperatura e umidade, enquanto o **MQ-135** é eficaz na detecção de amônia. Outro sensor citado é o **CCS811**, utilizado para monitoramento de CO₂ (Balthazar et al., 2024). Em testes práticos, o **DHT22** demonstrou boa precisão (Fahrurrozi et al., 2024; Pereira et al., 2020; Dollente & Hanbal, 2024), e o **MQ-137** demonstrou capacidade de detectar níveis de amônia de até 60 ppm (Fahrurrozi et al., 2024; Pereira et al., 2020; Dollente & Hanbal, 2024).

Além disso, o **SHT30** oferece precisão de $\pm 0.3^{\circ}\text{C}$ e teve sua durabilidade garantida por ter sido protegido contra poeira e radiação com encapsulamento e escudos físicos (Xu et al., 2023). A literatura enfatiza que a durabilidade não depende apenas do sensor, mas

sim do encapsulamento, sendo fundamental para prolongar a vida útil dos componentes eletrônicos a adoção de **proteções físicas**, como caixas à prova d'água e filtros (Xu et al., 2023; Biheng et al., 2025).

4.7. QP7: Quais ferramentas de desenvolvimento e linguagens são mais utilizadas para sistemas embarcados em aviários?

Os artigos apontam que plataformas de hardware como Arduino, Raspberry Pi, ESP32 e STM32 são amplamente utilizadas para o desenvolvimento de sistemas embarcados, devido ao seu baixo custo e alta disponibilidade de fóruns e materiais de apoio (Neethirajan et al., 2023; Bhaskaran et al., 2024; Chen et al., 2024; Si et al., 2022; Xu et al., 2023; S et al., 2015).

Linguagens como C e C++ são comuns na programação de tais microcontroladores (Neethirajan et al., 2023; Bhaskaran et al., 2024; Biheng et al., 2025; S et al., 2015), enquanto Python é mais usado em integrações com inteligência artificial e análise de dados (Fahrurrozi et al., 2024; Neethirajan et al., 2023; Bhaskaran et al., 2024; Li et al., 2024; Swain et al., 2024).

Para garantir que a troca de informações seja feita de forma eficiente, são recomendados protocolos como o MQTT, HTTP, CoAP e AMQP (Tran et al., 2024; Silva et al., 2025; Rathod et al., 2022; Bhaskaran et al., 2024). Já sobre serviços em nuvem, são citados AWS, Azure, ThingSpeak e Blink, pois permitem o armazenamento, processamento e uma visualização de dados de forma remota (Rathod et al., 2022; Biheng et al., 2025; Putri et al., 2024; Becoea et al., 2024). Alguns estudos também destacam o uso de interfaces web desenvolvidas em HTML e JavaScript (Fahrurrozi et al., 2024; Balthazar et al., 2024), além da integração com dashboards para facilitar a gestão do aviário. A escolha das tecnologias deve considerar o equilíbrio entre custo, robustez e escalabilidade exigidos pelo sistema (Silva et al., 2025; Balthazar et al., 2024).

5. Discussão

A análise da literatura científica revela que a aplicação de sistemas embarcados no manejo avícola é um campo de pesquisa robusto e em franca expansão. O pico de publicações em 2024 e o protagonismo de nações como Índia, Brasil e China sublinham a importância estratégica da avicultura de precisão para as economias globais e emergentes. A disseminação do conhecimento em periódicos de alto impacto confirma o rigor técnico das soluções propostas.

Respondendo à questão principal, a literatura estabelece que os sistemas embarcados são um pilar essencial para a modernização da avicultura, atuando como a principal ferramenta para mitigar os severos desafios ambientais. O foco das soluções está na automação do controle de gases (amônia) e temperatura, utilizando uma combinação de sensores específicos e plataformas de baixo custo. Esta abordagem resulta em benefícios mensuráveis, como a melhoria da produtividade e o aumento do bem-estar animal por meio de um ambiente otimizado e estável.

Existe uma clara correlação entre os desafios identificados e as tecnologias propostas. O problema da durabilidade dos sensores em ambientes corrosivos é o ponto de maior

convergência, levando os pesquisadores a investir em soluções de proteção física e na escolha de componentes mais robustos. Da mesma forma, a necessidade de resposta em tempo real às variações climáticas impulsiona a adoção de protocolos de baixa latência como o MQTT, permitindo que a infraestrutura de hardware execute ações imediatas, o que diretamente contribui para o aumento da produtividade. O debate sobre Edge, Fog e Cloud Computing (Oliveira et al., 2024) também mostra a preocupação em garantir a disponibilidade do sistema, mesmo com falhas de conectividade, um desafio central em áreas rurais.

Embora a pesquisa tenha avançado significativamente na validação de protótipos, a RSL revela importantes lacunas que representam oportunidades futuras. Há uma carência de estudos de validação em escala comercial que monitorem a precisão e a necessidade de calibração dos sensores em períodos prolongados (meses ou anos) sob condições reais de aviário

Além disso, a maioria dos estudos foca na coleta e visualização de dados. Há uma oportunidade clara para a integração com Inteligência Artificial (IA), visando o desenvolvimento de modelos preditivos e algoritmos de aprendizado de máquina (usando Python) que possam prever o estresse térmico ou o aumento da concentração de amônia antes que atinjam níveis críticos (Kow et al., 2025; Ozbilge et al., 2020). Por fim, a diversidade de protocolos e plataformas em uso dificulta a interoperabilidade entre diferentes sistemas. A padronização de comunicação e o desenvolvimento de arquiteturas de referência para a avicultura poderiam acelerar a adoção da tecnologia (Becoea et al., 2024).

6. Conclusão

Em síntese, a incorporação de sistemas embarcados e IoT no manejo avícola marca uma transição fundamental de práticas tradicionais para uma gestão de precisão, baseada em dados e automatizada. A produção científica global reflete e impulsiona essa tendência, oferecendo soluções concretas para os desafios ambientais e operacionais.

O artigo demonstra que as plataformas de desenvolvimento de baixo custo e a integração com tecnologias de computação em nuvem formam a base para um futuro onde os aviários funcionarão como ecossistemas inteligentes. Esses sistemas serão capazes de ajustar o ambiente em tempo real para maximizar a produção e garantir o bem-estar animal de forma contínua.

A principal contribuição desta revisão é a consolidação de um mapa tecnológico para a avicultura: sabemos quais sensores usar (DHT22, SHT30, MQ-137), quais plataformas são acessíveis (ESP32, Arduino) e quais protocolos (MQTT) garantem a comunicação. No entanto, o desafio da robustez dos sensores em ambientes hostis e a necessidade de escalabilidade industrial persistem como barreiras a serem transpostas. A inovação tecnológica, portanto, não é apenas um suporte operacional, mas a estratégia central para construir um setor avícola mais eficiente, lucrativo e alinhado com padrões éticos de sustentabilidade.

Referências

- Ahiara, W. C., Udeani, H. U., Okey, D. O., & Ihekweaba, C. (2022). Near Field Communication Intelligent Remote Livestock Monitoring System (Nigeria). *NIPES Journal of Science and Technology Research*, 4(2), 197-209. <https://doi.org/10.37933/nipes.e/4.2.2022.20>
- Amr, M. E., Al-Awamry, A. A., Elmenyaw, M. A., & Eldien, A. S. T. (2022). Design and Implementation of a Low-cost IoT Node for Data Processing, Case Study: Smart Agriculture. *Journal of Communications*, 17(2), 99-109. <https://doi.org/10.12720/jcm.17.2.99-109>
- Arshad, J., Rehman, A. U., Othman, M. T. B., Ahmad, M., Tariq, H. B., Khalid, M. A., Moosa, M. A. R., Shafiq, M., & Hamam, H. (2022). Deployment of Wireless Sensor Network and IoT Platform to Implement an Intelligent Animal Monitoring System. *Sustainability (Switzerland)*, 14(10), 6249. <https://doi.org/10.3390/su14106249>
- Baig, Z. T., & Shastry, C. (2022). SMART HYBRIDIZED ROUTING PROTOCOL FOR ANIMAL MONITORING AND TRACKING APPLICATIONS. *Scalable Computing: Practice and Experience*, 23(4), 339-349. <https://doi.org/10.12694/scpe.v23i4.2040>
- Balan, T., Dumitru, C., Dudnik, G., Alessi, E., Lesecq, S., Correvon, M., Passaniti, F., & Licciardello, A. (2020). Smart Multi-Sensor Platform for Analytics and Social Decision Support in Agriculture. *Sensors*, 20(15), 4127. <https://doi.org/10.3390/s20154127>
- Barbosa, V., Sabino, A., Lima, L. N., Brito, C., Feitosa, L., Andrade, E., & Silva, F. A. (2024). Dependability Evaluation of a Smart Poultry Monitoring System with Disaster Recovery Mechanism. *Journal of the Brazilian Computer Society*, 30(1). <https://doi.org/10.5753/jbcs.2024.3863>
- Bauer, J., & Aschenbruck, N. (2020). Towards a low-cost RSSI-based crop monitoring. *ACM Transactions on Internet of ThingSs*, 1(4). <https://doi.org/10.1145/3393667>
- Becoea, J. P., Grane, M., Miguez, M., & Arnaud, A. (2024). LoRa, Sigfox, and NB-IoT: An Empirical Comparison for IoT LPWAN Technologies in the Agribusiness. *IEEE Embedded Systems Letters*, 16(3), 283-286. <https://doi.org/10.1109/LES.2024.3394446>
- Bhamidipati, K., Anuradha, G., Swaminathan, B., & Muppidi, S. (2023). Smart IoT Application in Soil Moisture and Heat Level Prediction Using Sine Cosine-Horse Herd Optimized Deep Learning. *International Journal on Artificial Intelligence Tools*, 32(6). <https://doi.org/10.1142/S0218213023500252>
- Bhaskaran, H. S., Gordon, M., & Neethirajan, S. (2024). Development of a Cloud-Based IoT System for Livestock Health Monitoring Using AWS and Python. *bioRxiv*. <https://doi.org/10.1101/2024.06.08.598087>

- Biheng, A., Tu, C., Owolawi, P. A., Du Plessis, D., & Du, S. (2025). Network Architecture of a Fog–Cloud-Based Smart Farming System. *IoT*, 6(1), 17. <https://doi.org/10.3390/iot6010017>
- Boonsong, W. (2022). Animal Identification and Performance Analysis Based on RFIDIoT Smart Farming Applications. *Inżynieria Rolnicza*, 98(2), 6-9. <https://doi.org/10.15199/48.2022.02.02>
- Chaudhry, A., Mumtaz, R., Zaidi, S., Tahir, M., & School, S. (2020). Internet of ThingSs (IoT) and Machine Learning (ML) enabled Livestock Monitoring. *2020 International Conference on HONET*, 151-155. <https://doi.org/10.1109/HONET50430.2020.9322666>
- Chen, S., Li, D., Liu, J., & Wei, R. (2024, Janeiro). Raspberry Pi Based Intelligent Greenhouse Internet of ThingSs Platform. *Proceedings of the 2023 5th International Conference on Internet of ThingSs, Automation and Artificial Intelligence*, 787-791. <https://doi.org/10.1145/3653081.3653213>
- Chung, H., Li, J., Kim, Y., Van Os, J. M. C., Brounts, S. H., & Choi, C. Y. (2020). Using implantable biosensors and wearable scanners to monitor dairy cattle's core body temperature in real-time. *Computers and Electronics in Agriculture*, 174, 105453. <https://doi.org/10.1016/j.compag.2020.105453>
- Codeluppi, G., Davoli, L., & Ferrari, G. (2021). Forecasting air temperature on edge devices with embedded Ai. *Sensors*, 21(12), 3973. <https://doi.org/10.3390/s21123973>
- Cruz, E., Hidalgo-Rodriguez, M., Acosta-Reyes, A. M., Rangel, J. C., Boniche, K., & Gonzalez-Olivardia, F. (2025). ACMSPT: Automated Counting and Monitoring System for Poultry Tracking. *AgriEngineering*, 7(3), 86. <https://doi.org/10.3390/agriengineering7030086>
- da Rocha Balthazar, G., Silveira, R., & Silva, I. (2024). Use of Multi-Agent Systems and the Internet of ThingSs to Monitor the Environment of Commercial Broiler Poultry Houses through Specific Air Enthalpy. *Journal of Animal Behaviour and Biometeorology*, 12, 2024012. <https://doi.org/10.31893/jabb.2024012>
- Dollente, E., & Hanbal, I. (2024). IoT-based Decision Support System for Poultry Care for Ilocos Sur Polytechnic State College. *Nanotechnology Perceptions*, 20(S2), 149-161. <https://doi.org/10.62441/nano-ntp.v20iS2.13>
- Durairaj, K., Kumar, D. V., & Kanagachidambaresan, G. R. (2024). SMART COW HEALTH MONITORING SYSTEM: AN IOT BASED COW HOOOF DETECTION AND HEALTHCARE ALERT SYSTEM BY USING LSTM NETWORK. *Scalable Computing: Practice and Experience*, 25(3), 1702-1715. <https://doi.org/10.12694/scpe.v25i3.2785>
- Eladl, S. G., Haikal, A. Y., Saafan, M. M., & ZainEldin, H. Y. (2024). A proposed plant classification framework for smart agricultural applications using UAV images and artificial intelligence techniques. *Alexandria Engineering Journal*, 109, 466-481. <https://doi.org/10.1016/j.aej.2024.08.076>

- El Moutaouakil, K., Jdi, H., Jabir, B., & Noureddine, F. (2023). Digital Farming: A Survey on IoT-based Cattle Monitoring Systems and Dashboards. *Agris on-line Papers in Economics and Informatics*, 15(2), 31-39. <https://doi.org/10.7160/aol.2022.150203>
- Fahrurrozi, I., Wahyono, Sari, Y., Sari, A. K., Usuman, I., & Ariyadi, B. (2024). Integrating Random Forest Model and Internet of ThingSs-Based Sensor for Smart Poultry Farm Monitoring System. *Indonesian Journal of Electrical Engineering and Computer Science*, 33(2), 1283-1292. <https://doi.org/10.11591/ijeecs.v33.i2.pp1283-1292>
- Ferrag, M. A., Shu, L., Friha, O., & Yang, X. (2022). Cyber Security Intrusion Detection for Agriculture 4.0: Machine Learning-Based Solutions, Datasets, and Future Directions. *IEEE/CAA Journal of Automatica Sinica*, 9(3), 407-436. <https://doi.org/10.1109/JAS.2021.1004344>
- Guo, L., Yassin, I. M., & Rizman, Z. I. (2024). Design and Application of Agricultural Integrated Management System Based on TLINK. *Journal of Studies in Science and Engineering*, 4(2), 159-189. <https://doi.org/10.53898/josse2024429>
- Kandepan, D., & Venkatesan, D. K. (2022). A Novel Approach to Diagnose the Animal Health Continuous Monitoring Using IoT Based Sensory Data. *Instrumentation Mesure Métrologie*, 21(5), 159-170. <https://doi.org/10.18280/i2m.210501>
- Kasera, R., Nath, S., Das, B., Kumar, A., & Acharjee, T. (2025). IoT Enabled Smart Agriculture System for Detection and Classification of Tomato and Brinjal Plant Leaves Disease. *Scalable Computing: Practice and Experience*, 26(1), 96-113. <https://doi.org/10.12694/scpe.v26i1.3826>
- Kow, P.-Y., Wang, Y.-T., Chang, Y.-W., Lee, M.-H., Yao, M.-H., Chang, L.-C., & Chang, F.-J. (2025). AI-driven weather downscaling for smart agriculture using autoencoders and transformers. *Computers and Electronics in Agriculture*, 232, 110129. <https://doi.org/10.1016/j.compag.2025.110129>
- Li, G., Wu, C., An, Q., Peng, H., Alazab, M., Wang, F., Li, Y., & Liu, Z. (2024). Privacy-Enhanced Digital Twin Framework for Smart Livestock Management: A Federated Learning Approach with Privacy-Preserving Hybrid Aggregation (PPHA). *Information Sciences*, 665, 119311. <https://doi.org/10.1016/j.ins.2024.119311>
- Lufyagila, B., Machuve, D., & Clemen, T. (2022). IoT-powered system for environmental conditions monitoring in poultry house: A case of Tanzania. *Information Technology for Development*, 14(4), 1020-1031. <https://doi.org/10.1080/20421338.2021.1924348>
- Manikandan, R., Ranganathan, G., & Bindhu, V. (2022). Intelligent Computing and Control Framework for Smart Automated System. *Intelligent Automation and Soft Computing*, 33(1), 173-189. <https://doi.org/10.32604/iasec.2022.023922>
- Martínez, F., Romaine, J. B., Johnson, P., Cardona Ruiz, A., & Millán Gata, P. (2025). Novel Fusion Technique for High-Performance Automated Crop Edge Detection in Smart Agriculture. *IEEE Access*, 13, 22429-22445. <https://doi.org/10.1109/ACCESS.2025.3536701>

- Mohana Priya, S., Kanimozhi, N., Nandhagopal, S., & Saveetha, P. (2020). Precision agriculture using IOT varied sensors: A gateway management system. *Journal of Advanced Research in Dynamical and Control Systems*, 12(1), 56-65. <https://doi.org/10.5373/JARDCS/V12I1/20201008>
- Morchid, A., Oughannou, Z., El Alami, R., Qjidaa, H., Ouazzani Jamil, M., & Khalid, H. M. (2024). Integrated Internet of ThingSs (iot) solutions for early fire detection in smart agriculture. *Results in Engineering*, 24, 103392. <https://doi.org/10.1016/j.rineng.2024.103392>
- Neethirajan, S. (2023). Artificial Intelligence and Sensor Technologies in Dairy Livestock Export: Charting a Digital Transformation. *Sensors*, 23(16), 7045. <https://doi.org/10.3390/s23167045>
- Oguz, F., Ekersular, M., Sunnetci, K., & Alkan, A. (2024). Enabling Smart Agriculture: An IoT-Based Framework for Real-Time Monitoring and Analysis of Agricultural Data. *Agricultural Research*, 13, 1-12. <https://doi.org/10.1007/s40003-024-00705-x>
- Ojo, M. O., Viola, I., Baratta, M., & Giordano, S. (2022). Practical experiences of a smart livestock location monitoring system leveraging GNSS, LoRaWAN and cloud services. *Sensors*, 22(1), 273. <https://doi.org/10.3390/s22010273>
- Ojo, M. O., Viola, I., Miretti, S., Martignani, E., Giordano, S., & Baratta, M. (2024). A Deep Learning Approach for Accurate Path Loss Prediction in LoRaWAN Livestock Monitoring. *Sensors*, 24(10), 2991. <https://doi.org/10.3390/s24102991>
- Oliveira, F., Pereira, P., Dantas, J., Araujo, J., & Maciel, P. (2024). Dependability Evaluation of a Smart Poultry House: Addressing Availability Issues Through the Edge, Fog, and Cloud Computing. *IEEE Transactions on Industrial Informatics*, 20(2), 1304-1312. <https://doi.org/10.1109/TII.2023.3275656>
- Özbilge, E., Kirsal, Y., & Çağlar, E. (2020). Modelling and analysis of IoT technology using neural networks in agriculture environment. *International Journal of Computers, Communications and Control*, 15(3). <https://doi.org/10.15837/IJCCC.2020.3.3885>
- Patrizi, G., Bartolini, A., Ciani, L., Gallo, V., Sommella, P., & Carratu, M. (2022). A Virtual Soil Moisture Sensor for Smart Farming Using Deep Learning. *IEEE Transactions on Instrumentation and Measurement*, 71. <https://doi.org/10.1109/TIM.2022.3196446>
- Pereira, W. F., da Silva Fonseca, L., Putti, F. F., Góes, B. C., & Naves, L. P. (2020). Environmental monitoring in a poultry farm using an instrument developed with the Internet of ThingSs concept. *Computers and Electronics in Agriculture*, 170, 105257. <https://doi.org/10.1016/j.compag.2020.105257>
- Putri, R. E., Lestari, N. U., Ifmalinda, Arlius, F., Putri, I., & Hasan, A. (2024). Monitoring and Controlling System of Smart Mini Greenhouse Based on Internet of ThingSs (IoT) for Spinach Plant (*Amaranthus* sp.). *International Journal on Advanced Science, Engineering and Information Technology*, 14(1), 131-136. <https://doi.org/10.18517/ijaseit.14.1.18408>

- Rathod, M. L., Shivaputra, A., Umadevi, H., Nagamani, K., Periyasamy, S., & Chelladurai, S. J. S. (2022). Cloud Computing and Networking for SmartFarm AgriTech. *Journal of Nanomaterials*, 2022, 6491747. <https://doi.org/10.1155/2022/6491747>
- Sabrina, F., Sohail, S., Farid, F., Jahan, S., Ahamed, F., & Gordon, S. (2022). An Interpretable Artificial Intelligence Based Smart Agriculture System. *Computers, Materials & Continua*, 72(2), 3777-3797. <https://doi.org/10.32604/cmc.2022.026363>
- Si, L. F., Li, M. Y., & He, L. (2022). Farmland monitoring and livestock management based on Internet of ThingSs. *Internet of ThingSs (Netherlands)*, 19, 100581. <https://doi.org/10.1016/j.iot.2022.100581>
- Silva, C., Yanagi Junior, T., Bettio, R., & Bahuti, M. (2025). Design of Wireless Web-Based Multiplatform System for Thermal Environmental Control of Broiler Facilities Using Fuzzy Set Theory. *Anais da Academia Brasileira de Ciências*, 97. <https://doi.org/10.1590/0001-3765202520240032>
- Song, Y., Bi, J., & Wang, X. (2024). Design and Implementation of Intelligent Monitoring System for Agricultural Environment in IoT. *Internet of ThingSs (Netherlands)*, 25, 101029. <https://doi.org/10.1016/j.iot.2023.101029>
- Swain, S. (2024). Smart Livestock Management: Integrating IoT for Cattle Health Diagnosis and Disease Prediction through Machine Learning. *Indonesian Journal of Electrical Engineering and Computer Science*, 34(2), 1192-1203. <https://doi.org/10.11591/ijeecs.v34.i2.pp1192-1203>
- Terence, S., Immaculate, J., Raj, A., & Nadarajan, J. (2024). Systematic Review on Internet of ThingSs in Smart Livestock Management Systems. *Sustainability*, 16(10), 4073. <https://doi.org/10.3390/su16104073>
- Torres-Sanz, V., Sanguesa, J. A., Martinez, F. J., Garrido, P., & Calafate, C. T. (2025). An autoconfiguration strategy for very large scale long range wide area network deployments in smart cities. *IET Smart Cities*, 7(1). <https://doi.org/10.1049/smc2.12096>
- Tran, K., Pham, A., Phươg, N., & Dang Thi, P. (2024). Analysis and Performance Comparison of IoT Message Transfer Protocols Applying in Real Photovoltaic System. *International Journal of Networked and Distributed Computing*, 12, 1-13. <https://doi.org/10.1007/s44227-024-00021-4>
- Xu, Z., Yang, J., Zhou, H., & Hou, Y. (2023). A Wireless Remote Monitoring Sensor for Agricultural Environment Based on NB-IoT. *INMATEH Agricultural Engineering*, 285-294. <https://doi.org/10.35633/inmateh-69-26>

Sistemas Embarcados e Visão Computacional para Segurança Residencial: Uma Revisão da Literatura

Joniel Mendes da Silva¹, João André Borges Santana¹, Heitor Garcez Martins Louzeiro¹, Wanderson Jean Conceição Silva¹, Felipe Goncalves dos Santos¹

Jonielmendes237@gmail.com; joaoandresantana38@gmail.com; heitorlouzeiro2019@gmail.com; wanderson.jean11@gmail.com; fgsantos.ti@gmail.com

¹ Instituto Federal de Educação, Ciência e Tecnologia do Piauí – IFPI, Piauí, Brasil

Pages: 129-141

Resumo: Este artigo apresenta uma revisão sistemática da literatura sobre sistemas embarcados e visão computacional aplicados à segurança residencial (2020-2025). Analisando 63 publicações selecionadas de bases como Scopus e ACM, o estudo identifica tendências e desafios. Os resultados revelam que a visão computacional é a tecnologia-chave, permitindo que os sistemas interpretem eventos. As aplicações mais recorrentes são controle de acesso por reconhecimento facial, detecção de intrusos e monitoramento de incêndios. Índia (38,10%), Paquistão (14,29%) e China (11,11%) lideram a produção científica. Esta revisão consolida o estado da arte, destacando o papel da análise visual na automação da segurança e apontando direções para futuras pesquisas.

Palavras-chave: Sistemas Embarcados; Segurança Residencial; Visão Computacional; Automação Residencial; Reconhecimento Facial.

Embedded Systems and Computer Vision for Home Security: A Literature Review

Abstract: This article presents a systematic literature review on embedded systems and computer vision applied to residential security (2020-2025). Analyzing 63 selected publications from databases such as Scopus and ACM, the study identifies trends and challenges. The results reveal that computer vision is the key technology, enabling systems to interpret events. The most recurrent applications are access control via facial recognition, intruder detection, and fire monitoring. India (38.10%), Pakistan (14.29%), and China (11.11%) lead in scientific production. This review consolidates the state-of-the-art, highlighting the role of visual analysis in security automation and pointing toward future research directions.

Keywords: Embedded Systems; Home Security; Computer Vision; Home Automation; Facial Recognition.

1. Introdução

A integração crescente da tecnologia no cotidiano impulsiona o desenvolvimento de soluções inteligentes em diversas esferas sociais. A segurança, em particular, emerge como um dos setores mais transformados pelos avanços em sistemas embarcados e Internet das Coisas (IoT). Dispositivos compactos, acessíveis e interconectados viabilizam a criação de mecanismos aprimorados de monitoramento, autenticação e resposta automatizada, expandindo significativamente a abrangência e a eficácia das aplicações destinadas à proteção de indivíduos e ambientes.

Nesse cenário, os sistemas embarcados aplicados à segurança residencial desempenham um papel crucial, ao permitirem a união de sensores, câmeras, atuadores e módulos de comunicação em soluções de baixo custo e alta confiabilidade. Tais capacidades possibilitam desde o controle de acesso por biometria e reconhecimento facial até a detecção de intrusos e o monitoramento remoto em tempo real, impulsionando a adoção de conceitos como casas inteligentes e automação residencial.

Para tornar esses sistemas verdadeiramente “inteligentes”, a visão computacional surge como uma tecnologia-chave, permitindo que eles não apenas detectem eventos, mas também os interpretem. Em vez de depender exclusivamente de sensores simples, as soluções modernas utilizam câmeras e algoritmos avançados para realizar tarefas complexas, como reconhecimento facial, detecção de objetos e análise de comportamento. Essa capacidade de análise visual transforma um simples dispositivo de monitoramento em um agente proativo de segurança, capaz de diferenciar entre um residente e um intruso (Taiwo; Ezugwu, 2021) ou identificar um foco de incêndio antes que ele se alastre (Abdusalomov et al., 2024). A aplicação de modelos de aprendizado de máquina, desde classificadores como Support Vector Machine (SVM), até redes neurais convolucionais (CNNs) profundas, tornou esses sistemas mais precisos e confiáveis (Wu et al., 2021; EL Fadel, 2025; Lima et al., 2021), consolidando o reconhecimento facial como uma aplicação dominante na área.

Diante deste panorama, torna-se fundamental investigar como a literatura científica tem abordado o tema nos últimos cinco anos, a fim de identificar tendências, aplicações predominantes e desafios persistentes. Este estudo, portanto, busca delinear o estado da arte e as lacunas de pesquisa existentes no campo do uso e desenvolvimento de sistemas embarcados na área de segurança. Os cabeçalhos e rodapés serão incluídos pelos editores.

2. Questões de pesquisa

Neste trabalho, foi realizada uma revisão da literatura com ênfase em sistemas embarcados, na área de segurança. O processo foi estruturado em etapas, incluindo a definição das perguntas de pesquisa, a seleção de estudos relevantes, a análise dos trabalhos selecionados e a posterior resposta às perguntas propostas (Kitchenham; Charters, 2007). A questão principal desta revisão (QP) é: O que a literatura, dos últimos cinco anos, diz sobre o uso e o desenvolvimento de sistemas embarcados na área de segurança? Com o intuito de fornecer uma resposta mais abrangente à questão central, foram formuladas as seguintes subquestões: PPO1 De que forma sistemas embarcados têm sido utilizados em soluções de segurança residencial?, PPO2: Quais países têm se destacado na produção científica sobre o uso de sistemas embarcados na

área de segurança?, PPO3: Quais eventos científicos e periódicos concentram a maior parte das publicações sobre sistemas embarcados aplicados à área de segurança?, PPO4: Quais aplicações de sistemas embarcados em segurança são mais recorrentes?, uais as linguagens mais utilizadas na área de sistemas embarcados e segurança?

3. Seleção dos artigos relevantes

A seleção de artigos relevantes é uma fase essencial no processo de revisão da literatura, envolvendo a construção de uma *string* de busca, escolha de bases de dados e a definição de critérios de elegibilidade de inclusão e exclusão que servirão como filtro para a relevância dos resultados.

(“IoT” OR “Internet of Things” OR “Smart Devices” OR “Connected Devices”) AND (“security” OR “data protection” OR “privacy”) AND (“Embedded systems” OR “Microcontroller”) AND “computer vision” OR “Smart alarm system”)

A fase inicial da pesquisa buscou abranger a intersecção entre sistemas embarcados, Internet das Coisas e segurança. Com base nisso, a primeira *string* de busca foi formulada para capturar esses três pilares de forma ampla.

No entanto, os resultados preliminares revelaram que a abordagem era genérica demais. O termo segurança se mostrou um conceito muito amplo, trazendo um volume excessivo de publicações de áreas distintas. Para analisar essa dispersão e identificar um foco mais coeso, utilizou-se a ferramenta VOSviewer. A análise de ocorrência de palavras-chave nos artigos recuperados confirmou a falta de um núcleo temático claro e a necessidade de contextualizar a busca.

A partir dessa análise, a estratégia foi refinar o foco para um domínio de aplicação específico: a automação residencial. Decisão essa que permitiu contextualizar a busca, e termos como *computer vision* e *smart alarm system* foram combinados com *smart home* e *home automation*, direcionando a pesquisa para soluções de segurança no ambiente doméstico.

Esse processo de afunilamento, validado pela análise no VOSviewer, resultou na seguinte *string* de busca final, mais precisa e alinhada aos objetivos do trabalho:

“Embedded system” OR “smart home” OR “smart device” AND “home automation” AND “computer vision” OR “Smart alarm system”

Para a realização deste mapeamento, foram selecionadas as bases de dados Scopus, Web of Science e ACM. A escolha se justifica por serem repositórios que agregam as publicações das principais conferências e periódicos nas áreas de Internet das Coisas, visão computacional e microcontroladores, garantindo assim uma cobertura abrangente e relevante da literatura técnica sobre o tema.

Como critérios de inclusão e exclusão, foram definidas algumas métricas a serem aplicadas durante a busca nas bases. Os critérios de inclusão foram:

CI1: Artigos relacionados exclusivamente às áreas de computação e engenharia;

CI2: Artigos com data de publicação entre os anos de 2020 e 2025;

CI3: Artigos escritos em inglês;

CI4: Artigos do tipo *Journal Article* (Artigo de Periódico) ou *Conference Paper* (Artigo de Conferência).

A etapa de busca de artigos foi feita entre maio e junho de 2025, resultando em um levantamento inicial de 1.876 publicações. Conforme detalhado na Tabela 1, os registros foram extraídos das bases Scopus (816), Web of Science (1.001) e ACM (59). Deste montante, 2 artigos duplicados foram removidos.

Fonte	Número de artigos
Scopus	816
Web of Science	1.001
ACM Digital Library	59
Total encontrado	1.876
Duplicatas removidas	-2
Quantidade final	1.874

Tabela 1 – Quantidade de artigos por banco de dados

3.1. Primeiro Ciclo

Com o total de 1.876 artigos da busca inicial e já sem as duplicatas, foi elaborada uma planilha para gerenciar os metadados (fonte, título, resumo, ano, autores, etc.). A partir daí, iniciou-se a primeira filtragem com base no critério de exclusão. EC1: Artigos que não tratam de segurança residencial com visão computacional. O resultado foi a exclusão da maioria dos artigos, cujo foco estava em áreas não relacionadas como cibersegurança e protocolos de criptografia. Assim, este primeiro ciclo de seleção foi finalizado com 194 artigos, que seguiram para a análise seguinte.

3.2. Segundo ciclo

O segundo ciclo de seleção dos artigos relevantes consistiu na aplicação de critérios de exclusão adicionais, visando refinar ainda mais a base de trabalhos. Além dos critérios já aplicados, foram utilizados:

EC2: Artigos focados em Interação Humano-Computador (IHC) para controle do ambiente.

EC3: Artigos cuja aplicação final não seja segurança física (mas sim conforto, conveniência, saúde, energia, etc.).

Com a aplicação desses critérios, um total de 131 artigos foram excluídos dos 194 que haviam restado após o primeiro ciclo. Tais artigos foram descartados por não se alinharem diretamente ao escopo da pesquisa. Após esse processo de filtragem, foram selecionados 63 artigos que tratam especificamente do tema proposto, finalizando assim o ciclo de busca e seleção.

4. Resultado e Análise

Essa Seção irá responder as cinco perguntas de pesquisa com base nos 63 artigos selecionados.

PP01: De que forma sistemas embarcados têm sido utilizados em soluções de segurança residencial? Os sistemas embarcados vêm sendo utilizados na segurança residencial para integrar sensores, câmeras e módulos de comunicação, permitindo monitoramento e resposta a diferentes situações. As soluções variam desde o controle de acesso por reconhecimento facial e a detecção de intrusos por sensores de movimento, até a identificação de riscos como incêndios. Esses sistemas podem acionar fechaduras eletrônicas, disparar alarmes e enviar alertas para acompanhamento remoto em aplicativos móveis, criando um ecossistema de segurança proativo (Yedulapuram et al., 2020).

PP02: Quais países têm se destacado na produção científica sobre o uso de sistemas embarcados na área de segurança? Durante o processo de seleção, 25 países e 6 continentes foram identificados: Oceania, Ásia, América do Sul, América do Norte, Europa e África. A Figura 1 mostra a distribuição geográfica das publicações.

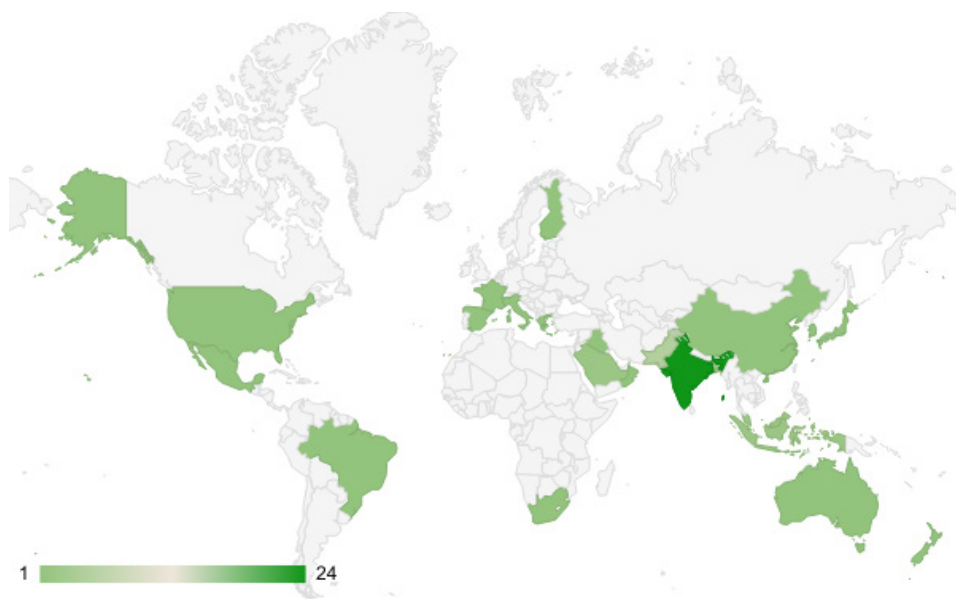


Figura 1 – Distribuição entre países de publicações relacionadas a área.

Dentre os 25 países, os países com maior produção são a Índia, Paquistão e China. Também aparecem, embora em menor número, países como Coreia, Bangladesh, Malásia e Arábia Saudita.

A Figura 1 mostra o nível superior da Índia, que produziu 24 dos 63 artigos, 38,10%. Outros países que também se destacaram nos últimos cinco anos incluem: Iraque, Finlândia, Itália e Brasil.

PP03: Quais eventos científicos e periódicos concentram a maior parte das publicações sobre sistemas embarcados aplicados à área de segurança?

Dos 63 artigos filtrados na etapa de seleção, foi possível categorizar a totalidade das publicações entre conferências e periódicos. A análise demonstrou que 65,1% correspondem a publicações em conferências (41 artigos), enquanto 34,9% foram publicados em periódicos (22 artigos). Observa-se, portanto, uma predominância significativa de publicações em anais de eventos e conferências na área pesquisada.

Dentre as conferências, destacam-se a série Lecture Notes in Electrical Engineering (LNEE), que reúne diversos eventos internacionais, e conferências IEEE como o International Conference on Communication Systems, Computing and IT Applications (CSCITA), o International Congress of Advanced Technology and Engineering (ICOTEN) e o International Conference on Robotics, Automation, Artificial Intelligence and Internet of Things (RAAICON).

No caso dos periódicos, o maior destaque é o Electronics, responsável por 5 publicações, seguido por títulos como IEEE Access e Applied Soft Computing, reconhecidos pelo rigor e impacto na área.

PP04: Quais aplicações de sistemas embarcados em segurança são mais recorrentes? As aplicações mais recorrentes em sistemas embarcados para segurança concentraram-se em mecanismos de autenticação por reconhecimento facial, frequentemente integrados a fechaduras eletrônicas. Também se destacaram soluções de detecção de movimento com geração de alertas e controle de acesso, aplicadas a cenários de vigilância e monitoramento residencial.

Outra abordagem observada foi a detecção de incêndio, integrando sensores de temperatura, gás e umidade para antecipar riscos e acionar alarmes de forma preventiva. Esses métodos representam os focos predominantes nas implementações analisadas, indicando as estratégias mais frequentes para controle integrado em residências ou outros ambientes.

PP05: Quais as linguagens de programação mais utilizadas na área de sistemas embarcados e segurança? Algumas linguagens foram mais utilizadas, destacando-se a linguagem Python, amplamente utilizada em projetos que envolvem detecção de intrusos com câmeras, reconhecimento facial e aplicações de aprendizado de máquina. A escolha do Python está fortemente associada à sua compatibilidade com bibliotecas como OpenCV, TensorFlow e frameworks de deep learning, além da facilidade de implementação em plataformas como o Raspberry Pi (Yedulapuram et al., 2020; Banerjee et al., 2022).

Além disso, C e C++ também foram linguagens usadas, especialmente em soluções baseadas em microcontroladores, como o Arduino, e em sistemas embarcados que exigem resposta em tempo real (Taiwo; Ezugwu, 2021). O uso de C foi mais comum em aplicações que utilizam sensores e atuadores simples. Já o C++ apareceu em sistemas

mais robustos, com uso de câmeras e implementação de algoritmos mais complexos, como o YOLO para detecção de objetos.

5. Discussão

A análise dos resultados obtidos nesta revisão de literatura revela tendências significativas e levanta questões importantes sobre o estado atual e futuro dos sistemas embarcados aplicados à segurança residencial. Os achados demonstram uma evolução tecnológica no campo, com implicações tanto para a comunidade acadêmica quanto para a indústria.

As tecnologias emergentes mostram uma integração crescente de técnicas de inteligência artificial, especialmente visão computacional, em sistemas embarcados para segurança residencial (Wu et al., 2021). O reconhecimento facial emerge como a aplicação dominante, representando quase 30\% das implementações analisadas, o que indica o avanço tecnológico, como também a aceitação comercial desta solução (Das et al., 2022). Esta predominância sugere que a tecnologia de reconhecimento facial atingiu um nível de confiabilidade e custo-benefício adequado para aplicações residenciais, superando limitações técnicas anteriores relacionadas à precisão e velocidade de processamento em dispositivos com recursos limitados (Banerjee et al., 2022).

A distribuição geográfica da pesquisa revela padrões interessantes que merecem análise detalhada. A Índia em primeiro lugar em produção científica, que reflete o investimento significativo do país em tecnologias emergentes e Internet das Coisas. Esta liderança pode ser atribuída a fatores como políticas governamentais de incentivo à inovação tecnológica, crescimento do mercado interno de tecnologia, e a presença de uma base sólida de instituições de pesquisa e desenvolvimento. A forte presença de países asiáticos, incluindo Paquistão e China, sugere uma concentração regional de pesquisa, possivelmente impulsionada pela demanda crescente por soluções de segurança residencial em mercados emergentes, onde o crescimento urbano acelerado e o aumento do poder aquisitivo criam oportunidades para adoção de tecnologias de segurança avançadas.

Apesar dos avanços identificados, persistem desafios importantes que limitam a adoção generalizada destas tecnologias. As questões de privacidade e proteção de dados emergem como preocupações centrais, especialmente em sistemas biométricos que coletam e processam informações pessoais sensíveis (EL FADEL, 2025). A falta de regulamentação adequada e padrões de segurança específicos para estes sistemas cria incertezas para usuários finais.

A revisão identificou lacunas importantes na literatura que representam oportunidades significativas para pesquisas futuras. A segurança cibernética dos próprios sistemas embarcados recebe atenção limitada nos estudos analisados, apesar de constituir um aspecto crítico considerando que estes dispositivos frequentemente operam conectados à internet. A eficiência energética, embora mencionada em alguns estudos, carece de pesquisas sistemáticas sobre otimização de consumo em dispositivos IoT para segurança. A integração com smart grids e redes elétricas inteligentes representa uma área pouco explorada, mas com potencial significativo para sistemas de segurança mais eficientes e sustentáveis. A ausência de estudos econômicos comparativos entre diferentes soluções

limita a compreensão sobre viabilidade comercial e adoção em diferentes contextos socioeconômicos.

As implicações práticas dos resultados sugerem que a adoção de sistemas embarcados em segurança residencial está em fase de consolidação, com soluções comercialmente viáveis disponíveis no mercado. No entanto, a necessidade de padronização e regulamentação adequada permanece como barreira significativa para adoção em massa. A educação dos usuários sobre benefícios, limitações e questões de privacidade também constitui aspecto importante para aceitação mais ampla destas tecnologias.

Esta revisão apresenta limitações que devem ser consideradas na interpretação dos resultados. O foco em publicações em inglês pode ter excluído pesquisas relevantes conduzidas em outros idiomas, especialmente considerando a concentração geográfica identificada em países asiáticos. O período de análise, limitado aos últimos cinco anos, embora adequado para capturar desenvolvimentos recentes, pode não refletir tendências de longo prazo. A concentração em bases de dados específicas, apesar de sua relevância e abrangência, pode ter omitido publicações importantes disponíveis em outros repositórios especializados. Estas limitações sugerem oportunidades para estudos futuros que expandam o escopo temporal, linguístico e de fontes de dados para uma compreensão mais completa do campo.

6. Conclusão

Os resultados demonstram uma evolução significativa de sistemas simples para soluções que integram inteligência artificial e visão computacional. O reconhecimento facial emergiu como aplicação dominante, seguido por detecção de movimento e detecção de incêndio, indicando avanço tecnológico e viabilidade comercial dessas soluções.

A concentração da produção científica em países asiáticos, liderada pela Índia, Paquistão e China, mostra investimentos estratégicos em tecnologias emergentes e demanda crescente por segurança em mercados em desenvolvimento. A diversidade de veículos de publicação evidencia a natureza interdisciplinar do campo, combinando engenharia eletrônica, ciência da computação e automação residencial.

Apesar dos avanços, persistem desafios importantes como questões de privacidade em sistemas biométricos, falta de padronização entre fabricantes e limitações de eficiência energética. A pesquisa identificou lacunas significativas, especialmente em segurança cibernética dos dispositivos embarcados, otimização energética e integração com infraestruturas inteligentes.

Os achados indicam que o campo está preparado para uma expansão, condicionada à superação dos desafios identificados através de regulamentação adequada, padronização técnica e educação dos usuários. Esta revisão contribui para o conhecimento científico ao mapear sistematicamente o estado da arte, identificando tendências e direções futuras que podem orientar pesquisadores, desenvolvedores e formuladores de políticas.

Para trabalhos futuros, recomenda-se a expansão do escopo temporal e linguístico da revisão, além de estudos empíricos focados nas lacunas identificadas, particularmente em segurança cibernética e análises econômicas, contribuindo para o avanço contínuo do conhecimento na área.

Referências

- Abdusalomov, A., Umirzakova, S., Safarov, F., Mirzakhilov, S., Egamberdiev, N., & Cho, Y.-I. (2024). A multi-scale approach to early fire detection in smart homes. *Electronics*, 13(22).
- Ahmed, S. M., Aditi, J., Afreen, A., Tharun, G., & Gunjan, V. K. (2023). Face recognition based home security system to detect usual/unusual person using iot. *Lecture Notes in Electrical Engineering*, 1038, 187-198.
- Ahmed, S. U., Khalid, H., Affan, M., Khan, T. A., & Ahmad, M. (2020). Smart surveillance and tracking system. *Proceedings - 2020 23rd IEEE International Multi-Topic Conference, INMIC 2020*.
- Alanazi, A., Al-Hadhrami, T., Saeed, F., & Awuson-David, K. (2021). Wireless remote control-security system for entrances (wrc-sse). *2021 International Congress of Advanced Technology and Engineering, ICOTEN 2021*.
- Alshar'e, M., Al Nasar, M. R., Kumar, R., Sharma, M., Dharamvir, & Tripathi, V. (2022). A face recognition method in machine learning (ml) for enhancing security in smart home. *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering, ICACITE 2022*, 1081-1086.
- Anandan, D., R, S., P, S., A, S., & B, S. (2024). Enhanced approach of face recognition in home security using deep learning techniques. *[s.n.]*, 1032-1037.
- Ayad, R., & Al-Khalidi, F. Q. (2023). Convolutional neural network (cnn) model to mobile remote surveillance system for home security. *International Journal of Computing and Digital Systems*, 13(1), 1329-1336.
- Azhar, M. H., & Jalal, A. (2024). Human-human interaction recognition using mask r-cnn and multi-class svm. *2024 3rd International Conference on Emerging Trends in Electrical, Control, and Telecommunication Engineering (ETECTE)*, 1-6.
- Banerjee, P., Datta, P., Pal, S., Chakraborty, S., Roy, A., Poddar, S., Dhali, S., & Ghosh, A. (2022). Home security system using raspberrypi. *Lecture Notes in Electrical Engineering*, 820, 167-176.
- Rocha, Á. (2012). Framework for a Global Quality Evaluation of a Website. *Online Information Review*, 36(3), 374-382. <https://www.abc.com>
- Bera, T. K., & Acharjya, P. P. (2023). *An investigation into the use of deep learning to recognize human activity*.
- Bruce, X., Liu, Y. Y., & Chan, K. C. C. (2020). A survey of sensor modalities for human activity recognition. *IC3K 2020 - Proceedings of the 12th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, 1, 282-294.
- Buddhi, D., Singh, R., & Gehlot, A. (2022). Computer assistance based home security analysis. *International Interdisciplinary Humanitarian Conference for Sustainability, IIHC 2022 - Proceedings*, 1435-1438.

- Condés, I., Fernández-Conde, J., Perdices, E., & Cañas, J. M. (2023). Robust person identification and following in a mobile robot based on deep learning and optical tracking. *Electronics*, 12(21).
- Costa e Silva, E. J., Guedes, K. Y. G. V., de Almeida Nava Alves, P. A. A. da S., Ribeiro, P. R. de A., & Barradas Filho, A. O. (2025). Systematic review of radio wave techniques for indoor positioning systems. *IEEE Latin America Transactions*, 23(3), 205-215.
- Dalal, S., Dahiya, N., Verma, A., Faujdar, N., Rathee, S., Jaglan, V., Rani, U., & Le, D.-N. (2025). A deep learning framework with learning without forgetting for intelligent surveillance in iot-enabled home environments in smart cities. *Recent Advances in Computer Science and Communications*, 18.
- Dalal, S., Lilhore, U. K., Sharma, N., Arora, S., Simaiya, S., Ayadi, M., Almujaally, N. A., & Ksibi, A. (2024). Improving smart home surveillance through yolo model with transfer learning and quantization for enhanced accuracy and efficiency. *PeerJ Computer Science*, 10, e1939.
- Das, H., & Meher, P. (2024). *Android Application-Based Security Surveillance Implementing Machine Learning*, 1-8.
- Das, P., Asif, N. A., Hasan, M. M., Abhi, S. H., Tatha, M. J., & Bristi, S. D. (2022). Intelligent door controller using deep learning-based network pruned face recognition. *Proceedings of 2022 25th International Conference on Computer and Information Technology, ICCIT 2022*, 120-124.
- Deb, A., Islam, A., Nath, A., Haque, M., Roy, R., Ul Islam, M. J., & Shahnaz, C. (2024). Vois+: A video over intercom service with home automation features for advanced security and safety. *2024 IEEE 3rd International Conference on Robotics, Automation, Artificial-Intelligence and Internet-of-Things (RAAICON)*, 299-302.
- Dewantara, B. S. B., Firmansyah, M. R., Sari, D. M., Setiawardhana, Harsono, T., & Aqil, P. R. (2022). Door lock access control using fusion of face and fingerprint recognition. *2022 IEEE Creative Communication and Innovative Technology, ICCIT 2022*.
- Di Benedetto, M., Carrara, F., Ciampi, L., Falchi, F., Gennaro, C., & Amato, G. (2022). An embedded toolset for human activity monitoring in critical environments. *Expert Systems with Applications*, 199, 117125.
- Egashira, T., Nishikawa, H., Kong, X., & Tomiyama, H. (2021). A home security camera system with container-based resource allocation on raspberry pi. *2021 International Conference on Electronics, Information, and Communication, ICEIC 2021*.
- Ekle, O. A., & Eberle, W. (2024). Anomaly detection in dynamic graphs: A comprehensive survey. *ACM Transactions on Knowledge Discovery from Data*, 18.
- EL Fadel, N. (2025). Facial recognition algorithms: A systematic literature review. *Journal of Imaging*, 11(2).
- Flores-Salgado, B., Gonzalez-Ambriz, S.-J., Martínez-García-Moreno, C.-A., & Beltrán, J. (2024). Iot-based system for campus community security. *Internet of Things*, 26, 101179.

- Ghafoor, S., Khan, K. B., Tahir, M. R., & Mustafa, M. (2020). Home automation security system based on face detection and recognition using iot. *Communications in Computer and Information Science*, 1198, 67-78.
- Hossain, M. A., Hossain, M. I., Hossain, M. D., & Huh, E.-N. (2022). DFC-D: A dynamic weight-based multiple features combination for real-time moving object detection. *Multimedia Tools and Applications*, 81, 32549-32580.
- Jain, P., Potter, H., Lee, A. J., & Mosse, D. (2020). Mafia: Multi-layered architecture for iot-based authentication. *Proceedings - 2020 2nd IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications, TPS-ISA 2020*, 199-208.
- Jenifa, G., Indu, S., Jeevitha, C., & Kiruthika, V. (2021). *Visitor/Intruder Monitoring System Using Machine Learning*.
- Jion, M. S. A., & Ahmad, M. (2024). A smart and secured office system using iot. *2024 International Conference on Advances in Computing, Communication, Electrical, and Smart Systems (iCACCESS)*, 1-6.
- Kalbande, V., Tikale, G., Agrawal, R., Sorte, S., Thakre, L., & Khekare, G. (2022). Design and implementation of motion sensing security system. *3rd International Conference on Electronics and Sustainable Communication Systems, ICESC 2022 - Proceedings*, 612-617.
- Kang, M., Lee, Y., & Park, M. (2020). Energy efficiency of machine learning in embedded systems using neuromorphic hardware. *Electronics*, 9(7).
- Kitchenham, B., & Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. Technical Report EBSE 2007-001, Keele University and Durham University Joint Report.
- Kaushik, H., Kumar, T., & Bhalla, K. (2022). iSecureHome: A deep fusion framework for surveillance of smart homes using real-time emotion recognition. *Applied Soft Computing*, 122.
- Lakshman, S. A., Akash, S., Cynthia, J., Gautam, R., & Ebenezer, D. (2024). Architecture and applications of iot devices in socially relevant fields. *SN Computer Science*, 5.
- Lima, G. A., Cotrin, R. O., Belan, P. A., & Araújo, S. A. de. (2021). Sistema de Visão Computacional para Identificação Automática de Potenciais Focos do Mosquito *Aedes aegypti* Usando Drones. *RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação*, 43, 93-109.
- Lu, X., Wang, L., Lin, C., Fan, X., Han, B., Han, X., & Qin, Z. (2024). Autodlar: A semi-supervised cross-modal contact-free human activity recognition system. *ACM Transactions on Sensor Networks*, 20.
- Mariappan, M., Thong, L. K., & Muthukaruppan, K. (2020). A design methodology of an embedded motion-detecting video surveillance system. *International Journal of Integrated Engineering*, 12(2), 55-69.

- Paikaray, D., & Parikh, S. (2022). A new way of smart home security using ml face recognition. *Proceedings of the 2022 11th International Conference on System Modeling and Advancement in Research Trends, SMART 2022*, 1628-1632.
- Papaleonidas, A., Psathas, A. P., & Iliadis, L. (2022). High accuracy human activity recognition using machine learning and wearable devices' raw signals. *Journal of Information and Telecommunication*, 6(3), 237-253.
- Patil, V. K., Hadawale, O., Pawar, V. R., & Gijre, M. (2021). Emotion linked aiot based cognitive home automation system with sensovisual method. *2021 IEEE Pune Section International Conference, PuneCon 2021*.
- Pavithra, M., Surya Karthikesh, P., Jahnnavi, B., Navyalokesh, M., C, R., & Krishna, K. (2024). Implementation of enhanced security system using roboflow. *[s.n.]*, 1-5.
- Rao, N., Rajan, M. H., Rebello, K., & Gharat, S. (2020). Energy-efficient face recognition authentication system using human detection iot modules. *2020 3rd International Conference on Communication Systems, Computing and IT Applications, CSCITA 2020 - Proceedings*, 125-129.
- Ravikumar, P., N, A. C., & G, H. (2024). Smart lock security system. *[s.n.]*, 1-4.
- Razzaq, M., Medina, J., Cleland, I., Nugent, C., Akhtar, U., Ali, B., Rehman, U. U., & Lee, S. (2020). umodt: An unobtrusive multi-occupant detection and tracking using robust kalman filter for real-time activity recognition. *Multimedia Systems*, 26.
- Rehman, Z., Tariq, N., Moqurrab, S. A., Yoo, J., & Srivastava, G. (2024). Machine learning and internet of things applications in enterprise architectures: Solutions, challenges, and open issues. *Expert Systems*, 41(1), e13467.
- Sabit, H. (2025). Artificial intelligence-based smart security system using internet of things for smart home applications. *Electronics*, 14(3).
- Salami, D., Hasibi, R., Palipana, S., Popovski, P., Michoel, T., & Sigg, S. (2023). Teslarapture: A lightweight gesture recognition system from mmwave radar sparse point clouds. *IEEE Transactions on Mobile Computing*, 22(8), 4946-4960.
- Sharma, V., Gupta, M., Kumar, A., & Mishra, D. (2021). Video processing using deep learning techniques: A systematic literature review. *IEEE Access*, 9, 139489-139507.
- Soji, E., Gnanamalar, S., Arumugam, N., Priscila, S., Selvam, N., & Rajest, S. (2024). *AI-Driven Computer Vision for Intelligent Home Automation and Surveillance Systems*.
- Surantha, N., Mertha, I. M. P., Widjaja, J. A., & Andersen, P. D. (2023). Smart home security system for intruder detection using yolo v3 algorithm. *ICIC Express Letters, Part B: Applications*, 14(3), 295-302.
- Swami, S., Singh, R., Gehlot, A., Kaur, J., Sharma, D., Sharma, S., & Singh, D. (2025). Fire detection and surveillance system with cloud-based alert to enhance safety in commercials and home. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, 23, 185.

- Taiwo, O., & Ezugwu, A. E. (2021). Internet of things-based intelligent smart home control system. *Security and Communication Networks*, 2021.
- Upadhyay, J., Deb, D., & Rawat, A. (2020). Design of smart door closer system with image classification over wlan. *Wireless Personal Communications*, 111(3), 1941-1953.
- Usman Ullah, A. W., Shah, J. A., Kadir, K., & Wahid, A. (2021). Development of smart surveillance system using cloud for security application. *Conference Record - IEEE Instrumentation and Measurement Technology Conference*, 2021-May.
- V, A. S., Jha, N., Begum, K., & Brahmani, K. (2024). *Smart Warehouse Management System*, 365-374.
- Valarmathi, V., Sathya, T., Nirmala, J., Sivarajeswari, S., Prasanth, R. M., & Srihari, V. (2022). Design and implementation of secured contactless doorbell using iot. *2022 1st International Conference on Computer, Power and Communications, ICCPC 2022 - Proceedings*, 187-191.
- Waseem, M., Khowaja, S. A., Ayyasamy, R. K., & Bashir, F. (2020). Face recognition for smart door lock system using hierarchical network. *2020 International Conference on Computational Intelligence, ICCI 2020*, 51-56.
- Wu, W., Su, D., Yuan, B., & Li, Y. (2021). Intelligent security monitoring system based on risc-v soc. *Electronics*, 10(11).
- Xu, W., & Zhai, Y. (2023). Design and implementation of home video surveillance systems based on iot location service. *International Journal of Information Technologies and Systems Approach*, 16, 1-18.
- Yang, A., Zhang, C., Chen, Y., Zhuansun, Y., & Liu, H. (2020). Security and privacy of smart home systems based on the internet of things and stereo matching algorithms. *IEEE Internet of Things Journal*, 7(4), 2521-2530.
- Yar, H., Khan, Z. A., Rida, I., Ullah, W., Kim, M. J., & Baik, S. W. (2024). An efficient deep learning architecture for effective fire detection in smart surveillance. *Image and Vision Computing*, 145.
- Yedulapuram, S., Arabelli, R., Mahender, K., & Sidhardha, C. (2020). Automatic door lock system by face recognition. *IOP Conference Series: Materials Science and Engineering*, 981(3).
- Zafar, R., Zaib, S., & Asif, M. (2020). False fire alarm detection using data mining techniques. *International Journal of Decision Support System Technology*, 12(4), 21-35.

Modelo de aprendizaje automático para la recomendación de carreras en jóvenes preuniversitarios

Andree Carrillo Guevara¹, Jose Avila Rabanal¹, Juan Mansilla-Lopez¹, Ricardo Loza²

u20211b589@upc.edu.pe; u202112411@upc.edu.pe; pcsijman@upc.edu.pe;
ricardo.loza@sheridancollege.ca

¹ Universidad Peruana de Ciencias Aplicadas, Facultad de Ingeniería, Lima, Perú.

² Shreridan College, Ontario, Canadá

Pages: 142-156

Resumen: La deserción universitaria es una problemática relevante en el Perú, influenciada en gran medida por una elección vocacional inadecuada en la etapa preuniversitaria. Este trabajo propone un modelo de recomendación de carreras basado en aprendizaje automático para brindar orientación vocacional personalizada a estudiantes de secundaria y contribuir a prevenir el abandono universitario. Se desarrolló un sistema con Random Forest que integra variables sociodemográficas, académicas y vocacionales obtenidas mediante encuestas aplicadas a más de 600 estudiantes de universidades de Lima Metropolitana. El modelo alcanzó métricas superiores al 78% en precisión, recall y F1-score, demostrando buena capacidad predictiva. Asimismo, se emplearon técnicas psicométricas, como el alfa de Cronbach ($\alpha=0.710$), para garantizar la calidad de los datos. La propuesta ofrece recomendaciones contextualizadas y personalizadas, aportando una solución concreta a la orientación vocacional como estrategia para reducir la deserción en la educación superior.

Palabras-clave: deserción; aprendizaje automático; recomendación de carreras.

Machine learning model for career recommendation for pre-university students

Abstract: University dropout is a significant issue in Peru, largely influenced by inadequate vocational choice during the pre-university stage. This work proposes a career recommendation model based on machine learning to provide personalized vocational guidance to high school students and help prevent university attrition. A system was developed using Random Forest, integrating sociodemographic, academic, and vocational variables collected through surveys applied to more than 600 students from universities in Metropolitan Lima. The model achieved metrics above 78% in precision, recall, and F1-score, demonstrating strong predictive performance. Psychometric techniques, such as Cronbach's alpha ($\alpha=0.710$), were also applied to ensure data quality. The proposal offers contextualized and

personalized career recommendations, providing a concrete solution for vocational guidance as a key strategy to reduce dropout rates in higher education.

Keywords: dropout; machine learning; career recommendation.

1. Introducción

La deserción universitaria es una problemática de gran relevancia en el contexto educativo peruano y latinoamericano, afectando directamente no solo a los estudiantes, sino también a las instituciones educativas, a la sociedad y al desarrollo económico del país. Según Villarreal et al. (2024), este fenómeno responde a una variedad de factores personales, económicos, familiares, institucionales y, en particular, vocacionales, lo que evidencia la necesidad urgente de implementar soluciones efectivas y preventivas que acompañen al estudiante desde su transición del colegio a la universidad.

Diversos estudios revelan que una de las principales causas del abandono universitario está vinculada a una inadecuada elección de carrera, influenciada por la falta de orientación vocacional oportuna y herramientas adecuadas que permitan al estudiante alinear sus intereses, habilidades y expectativas con una trayectoria académica coherente (Escalante et al., 2023; Villarreal et al., 2024). En el Perú, La UNITEC (Tolentino, 2023) señala que el 42% de los estudiantes en Perú deja la universidad en los dos primeros ciclos debido a una elección vocacional deficiente, lo cual refuerza la idea de que la orientación profesional insuficiente es un factor crítico.

Esta situación se agrava en contextos socioeconómicos vulnerables, donde las limitaciones financieras y el bajo nivel educativo de los padres condicionan negativamente el proceso de elección vocacional (Villarreal et al., 2024; INEI, 2024). Por otro lado, la presión social y familiar —como lo mencionan Aguilar (2024)— contribuye a que muchos adolescentes elijan carreras no por vocación, sino por expectativas externas, lo que a largo plazo afecta su motivación y desempeño académico.

En el contexto peruano, Huayta y Pacheco (2024) indican que solo 4 de cada 10 egresados de secundaria acceden a la educación universitaria, lo que evidencia que más del 50 % de los estudiantes no continúan con su formación académica por múltiples factores. Entre ellos, la falta de orientación vocacional destaca como una de las principales causas, representando más del 50 %, según Gutiérrez et al. (2023). Esta situación limita significativamente el desarrollo del capital humano y afecta negativamente el progreso del país.

Ante esta situación, los investigadores han propuesto el uso de machine learning para recomendar carreras universitarias de forma más precisa. Cisneros (2023) implementó modelos de aprendizaje automático como Random Forest y Árboles de Decisión en 350 estudiantes de primer ciclo en México. Llauró et al. (2023) utilizaron la metodología de Diseño Centrado en el Usuario identificando que la edad, el rendimiento académico previo y la motivación del estudiante con la carrera son factores influyentes en la deserción. Majjate et al. (2024) desarrollaron un sistema web predictivo que incluye un módulo para recomendar carreras populares y otro para estimar la probabilidad de ingreso a universidades. Siswipraptini et al. (2024) proponen un modelo de recomendación de carreras para estudiantes de TI en Indonesia basado en EDM, Teoría Fundamentada

y un Naïve Bayes personalizado. Song et al. (2024) demuestra que el uso de Machine Learning mejora la precisión en la predicción vocacional frente a métodos tradicionales.

Soluciones como la propuesta por Llauro et al. (2023) se enfocan en reducir la deserción universitaria prediciendo qué estudiantes están en riesgo, pero no en prevenirla a través de una orientación vocacional adecuada. Las pocas que ofrecen orientación como las de Cisneros (2023) y Majate et al. (2024) no lo hacen de forma personalizada ni proporcionan un contexto completo. Generalmente, se limitan a sugerir una carrera sin brindar información adicional sobre ella ni sobre las universidades cercanas donde se puede estudiar.

Ante esta situación se propone la siguiente pregunta de investigación: ¿Puede un modelo de aprendizaje automático de recomendación de carreras puede mejorar la orientación vocacional de los estudiantes preuniversitarios?

La presente investigación propone un modelo de recomendación de carreras universitarias basado en Random Forest. El algoritmo utiliza información académica, personal y vocacional del estudiante, apoyándose de evaluaciones psicométricas, para sugerir opciones de carrera acordes a su perfil e intereses. Así, ofrece una orientación vocacional personalizada que facilita una toma de decisiones más informada previo al inicio de la educación superior.

El artículo está estructurado de la siguiente manera: luego de esta introducción, en la Sección 2 se presenta los trabajos relacionados. La Sección 3 describe las características del modelo propuesto. En la Sección 4 se detalla la metodología de implementación, incluyendo los datos utilizados, el modelo, el algoritmo y los resultados. Finalmente, en la Sección 5 se presentan las conclusiones del estudio.

2. Trabajos Relacionados

Diversos estudios recientes han desarrollado soluciones basadas en Machine Learning (ML), minería de datos y psicometría avanzada para mejorar la orientación vocacional, predecir el rendimiento académico y prevenir la deserción universitaria. Estas investigaciones demuestran que la integración de modelos predictivos, análisis de comportamiento estudiantil y sistemas de recomendación contribuye a una toma de decisiones más informada por parte de los estudiantes y a una mejor identificación de riesgos institucionales.

En el ámbito de los sistemas inteligentes de recomendación y acompañamiento académico, Atalla et al. (2023) diseñaron un sistema automatizado fundamentado en teoría de grafos y modelos como Random Forest y Redes Bayesianas, logrando una precisión del 86% en la predicción del rendimiento académico y generando rutas formativas personalizadas. En una línea similar, Assiri et al. (2024) implementaron un modelo de asignación vocacional empleando K-Nearest Neighbors (KNN), alcanzando una precisión del 100% al clasificar estudiantes según sus habilidades. Rico et al. (2024) propusieron un modelo conceptual para estimar el rendimiento académico mediante Regresión Lineal, Árboles de Decisión y Random Forest, obteniendo una mejora del 27% frente a modelos tradicionales.

La literatura también destaca el rol del Learning Analytics. Gedrimiene et al. (2023) identificaron que solo el 45% de los estudiantes confía en los sistemas de recomendación, lo que evidencia la necesidad de modelos más explicables. Gutierrez et al. (2023) integraron análisis de supervivencia y Deep Learning (especialmente redes neuronales convolucionales, CNN) logrando 98.7% de exactitud en la predicción del abandono. Zhang (2024) utilizó modelos profundos, como LSTM (Long Short-Term Memory) y CNN, para mejorar la planificación de carrera a partir de patrones detectados en datos estudiantiles. Vives et al. (2024) combinaron LSTM con GAN (Redes Generativas Adversarias) y técnicas de balanceo como SMOTE, alcanzando 98.3% de precisión en la predicción del desempeño en cursos introductorios. Urbina y Méndez (2022) aplicaron minería de datos y Random Forest para identificar factores de deserción, logrando 94.34% de precisión.

La deserción universitaria es ampliamente reconocida como un fenómeno multicausal. Fernández et al. (2024) identificaron como principales predictores la baja motivación, limitada integración académica y problemas emocionales. Arias et al. (2024) encontraron que el bajo rendimiento y las dificultades económicas son los factores más influyentes en Colombia. Valencia et al. (2023) destacaron la falta de apoyo familiar, mientras que Villarreal et al. (2024), mediante AutoML en universidades peruanas, señalaron que la carga emocional, el rendimiento y la situación económica explican gran parte del abandono. Maphosa et al. (2024) confirmaron la influencia de la mala elección vocacional, mientras que Goran et al. (2024) utilizaron inteligencia artificial explicable para resaltar la relevancia del rendimiento académico y las condiciones económicas. Alvarado et al. (2022) encontraron que el promedio, el avance curricular y el nivel socioeconómico son factores decisivos. Fantinelli et al. (2023) mostraron que la concordancia entre intereses del estudiante y su familia reduce el riesgo de deserción.

En el campo psicométrico aplicado a la orientación vocacional, López et al. (2023) demostraron que la satisfacción social universitaria modera significativamente la intención de abandono. Domingues et al. (2024) validaron el Career Indecision Profile (CIP-65), señalando que el neuroticismo incrementa la indecisión vocacional. Mbanga et al. (2024), mediante un estudio longitudinal, hallaron que la motivación autónoma disminuye la indecisión. Acosta et al. (2024) validaron un instrumento basado en inteligencias múltiples para identificar habilidades vocacionales en estudiantes de educación media. Langer et al. (2024) confirmaron la fiabilidad del Cuestionario de Aceptación y Acción-II (AAQ-II) en universitarios, mientras que Kranz et al. (2023) y Wang et al. (2024) desarrollaron herramientas basadas en teoría de respuesta al ítem (incluyendo MIRT) para medir competencias y preparación profesional. Santana et al. (2023) evidenciaron que los estereotipos de género afectan negativamente la autoeficacia y claridad vocacional.

Finalmente, se observa un creciente uso de sistemas conversacionales y recomendadores avanzados. Valtolina et al. (2024) propusieron un sistema basado en RASA, un framework open-source para chatbots impulsado por ML, combinado con BERT, un modelo de lenguaje basado en Transformers que permite comprensión contextual profunda. Mahale et al. (2024), Paranjape et al. (2024), Siddique et al. (2023), Li y Niu (2023), Xu et al. (2024) y Hamdollahi y Hashemzadeh (2023) desarrollaron sistemas en otros dominios —agricultura, cine, energía, educación y alimentación— demostrando la

versatilidad de modelos como Random Forest, XGBoost, LSTM y redes convolucionales, cuyas técnicas y arquitecturas pueden adaptarse para la recomendación vocacional.

3. Método

Nuestro estudio propone un modelo de recomendación de carreras universitarias basado en variables demográficas, económicas, vocacionales y educativas. Como se observa en la Figura 1, la metodología seguida comprende la extracción, limpieza y transformación de los datos, seguida del modelado y evaluación de los resultados. El objetivo es identificar el modelo que ofrezca la mayor precisión en la recomendación personalizada de carreras para estudiantes de secundaria.

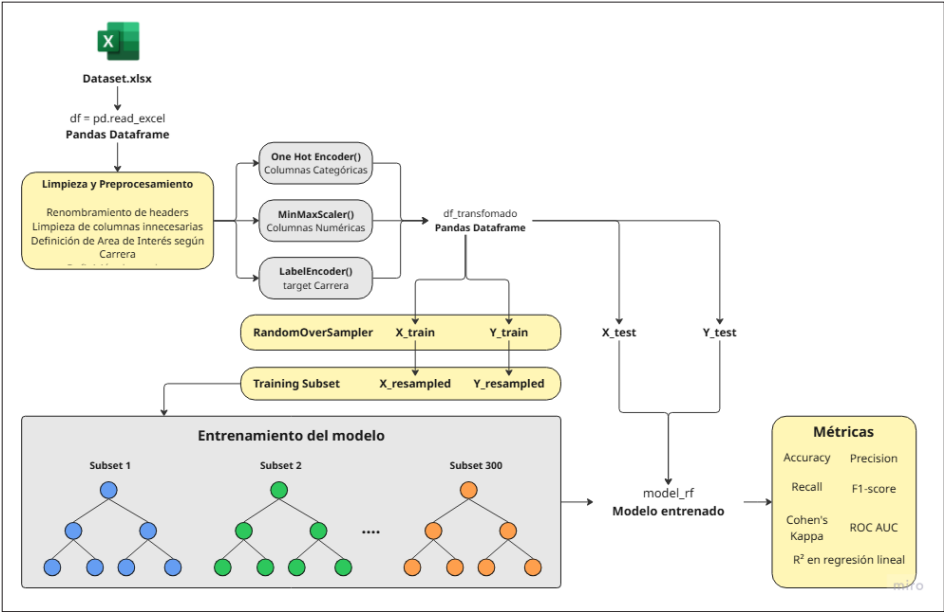


Figura 1 – Proceso de Entrenamiento del Modelo.

3.1. Extracción

Durante el semestre 2025-I se aplicó una encuesta a estudiantes universitarios a partir del quinto ciclo (enlace: <https://forms.office.com/r/CpkPbfFq6Q>), con el objetivo de analizar las características que influyeron en la elección de su carrera y su permanencia en ella. La recolección se llevó a cabo en ocho universidades con mayor población estudiantil de Lima Metropolitana, seleccionadas por su representatividad académica. Se obtuvieron más de 600 respuestas válidas, abarcando 30 carreras agrupadas en 11 facultades. El cuestionario incluyó variables demográficas, académicas, económicas y vocacionales, permitiendo construir un conjunto de datos diverso y útil para el entrenamiento del modelo.

Este instrumento de recolección y la metodología empleada fueron validados y aprobados por nuestra asesora especializada, garantizando la pertinencia y solidez académica del estudio. Su supervisión abarcó desde el diseño del cuestionario hasta la estrategia de muestreo, asegurando la calidad y fiabilidad de los datos obtenidos para los objetivos del proyecto.

Categoría	Variable	Tipo
Personal	Age, Gender	Numeric/Categorical
Academic	Preferred Courses (1–3), Type of School, Major	Categorical
Soft Skills	Empathy, Listening, Problem-solving, Communication, Teamwork	Numeric
Socioeconomic	District, Tuition Fee	Categorical / Numeric
Socioeconómica	Mensualidad universitaria	Numérico

Tabla 1 – Variables del conjunto de datos

3.2. Limpieza

A partir de las variables iniciales, se generaron columnas derivadas esenciales: el área de interés asociada a cada carrera y el nivel socioeconómico correspondiente al distrito de residencia del estudiante. La categorización del área de interés para cada carrera se realizó mediante una clasificación preestablecida, detallada y visualizada en la Tabla 2, que agrupa las diversas disciplinas académicas en bloques temáticos afines (e.g., Ingenierías, Humanidades, Salud, etc.).

Área de interés	Carreras asociadas
<i>I</i>	Ingeniería de sistemas de Información, Ingeniería Civil, Ingeniería Industrial, Ingeniería Mecatrónica, Ingeniería Eléctrica, Ingeniería de Minas, Ingeniería Biomédica, Ingeniería de Telecomunicaciones
<i>E</i>	Ingeniería Ambiental, Ingeniería Agroindustrial
<i>A</i>	Arquitectura, Diseño Gráfico, Gastronomía, Arte, Diseño de Modas
<i>C</i>	Economía, Marketing, Administración, Ingeniería de Gestión Empresarial, Contabilidad
<i>H</i>	Derecho, Ciencias de la comunicación, Educación, Psicología, Turismo
<i>S</i>	Odontología, Medicina, Medicina Veterinaria, Enfermería

Tabla 2 – Relación carrera y área de interés

Paralelamente, el nivel socioeconómico fue determinado en función del distrito de procedencia, utilizando clasificaciones socioeconómicas reconocidas para Lima Metropolitana, cuya distribución se ilustra en la Tabla 3. Estas variables secundarias son fundamentales para analizar la interacción entre las aptitudes vocacionales, identificadas por el test de CHASIDE, y las condiciones socioeconómicas del entorno del estudiante.

Nivel Socioeconómico (NSE)	Distritos Asociados
A	San Isidro, La Molina, Miraflores, San Borja
B	Jesús María, Magdalena del Mar, Santiago de Surco, Lince, Barranco
C	Pueblo Libre, San Miguel, Chorrillos, Cercado de Lima, Rímac, Santa Anita, Breña, Surquillo, San Luis
D	Comas, Los Olivos, Independencia, Villa el Salvador, San Juan de Miraflores, El Agustino, La Victoria, Lurín, Callao, Lurigancho
E	Villa María del Triunfo, San Juan de Lurigancho, Ate, Puente Piedra, Carabaylo, San Martín de Porres, Pachacámac

Tabla 3 – Relación nivel socioeconómico y distrito

En el marco del preprocesamiento de los datos, se llevaron a cabo diversas optimizaciones en el conjunto de información. Se procedió a la estandarización de los encabezados de las columnas para garantizar uniformidad y facilitar el manejo de los datos. Asimismo, se implementó un criterio de filtrado para incluir únicamente a estudiantes que hubieran completado al menos el cuarto ciclo académico. Finalmente, se eliminaron aquellas columnas que, tras una evaluación de su relevancia, no mostraron un impacto significativo en la capacidad de predicción de la variable objetivo.

La fiabilidad de un instrumento de medición es fundamental para garantizar que los datos recopilados sean consistentes y estables. Cuando se trabaja con un conjunto de variables numéricas que se espera midan un mismo constructo subyacente, el coeficiente Alfa de Cronbach (α) es la herramienta estadística más utilizada para evaluar la consistencia interna (Cronbach, 1951).

Este coeficiente indica qué tan bien se correlacionan los ítems entre sí, reflejando la coherencia de las respuestas a lo largo de la escala. Un Alfa de Cronbach elevado sugiere que los ítems están midiendo de manera consistente el mismo concepto.

La consistencia interna de la escala de habilidades blandas (empatía, escucha, resolución de problemas, comunicación y trabajo en equipo) fue validada mediante el Alfa de Cronbach, obteniéndose $\alpha=0.710$, lo cual se considera aceptable (George & Mallery, 2010).

4.3. Transformación

Para preparar los datos de entrada al modelo de Machine Learning, se aplicaron transformaciones específicas. Las variables categóricas fueron procesadas con un OneHotEncoder, que las convierte en columnas binarias (0 o 1). Esto evita que el modelo infiera un orden o magnitud incorrecta entre categorías nominales.

Las variables numéricas se escalaron utilizando un MinMaxScaler. Este proceso normaliza los datos a un rango específico (entre 0 y 1), lo cual es crucial para evitar que variables con valores más grandes dominen el entrenamiento y para acelerar la convergencia de algoritmos sensibles a la escala. Finalmente, la variable objetivo, siendo categórica, se

transformó a valores numéricos enteros con un LabelEncoder, ya que la mayoría de los algoritmos de clasificación requieren un formato numérico para la etiqueta de salida. Todos estos transformadores fueron ajustados a los datos y guardados para asegurar la consistencia en el preprocesamiento de futuros datos.

4.4. Entrenamiento

Para el entrenamiento del modelo, el conjunto de datos se subdividió aleatoriamente en dos particiones: 70% para entrenamiento y 30% para prueba, una división crucial para evaluar la capacidad de generalización del modelo y mitigar el sobreajuste. Posteriormente, se abordó el desbalance de clases aplicando técnicas de sobremuestreo (Oversampling) para igualar la representación de las clases minoritarias con las mayoritarias, lo que ayuda a prevenir el sesgo del modelo y a mejorar su rendimiento.

El entrenamiento se realizó utilizando el algoritmo Random Forest, un método de aprendizaje en conjunto que construye múltiples árboles de decisión. Los hiperparámetros específicos utilizados para la configuración del modelo Random Forest fueron los siguientes:

- `n_estimators=300`: Este parámetro define la cantidad de árboles de decisión en el bosque. Un número mayor de árboles generalmente mejora la precisión y estabilidad del modelo.
- `max_depth=20`: Este parámetro limita la profundidad máxima de cada árbol a 20 niveles. Limitar la profundidad es clave para prevenir el sobreajuste y asegurar la generalización.
- `min_samples_leaf=1`: Este parámetro define el número mínimo de muestras requeridas para estar en un nodo hoja. Su valor de 1 indica que los nodos hoja pueden tener una sola muestra.

La aleatoriedad inherente del Random Forest proviene de dos fuentes: cada árbol se entrena con una muestra aleatoria de registros (bagging) y en cada división de nodo se considera un subconjunto aleatorio de características. Esta doble aleatoriedad asegura que los 300 árboles generados sean diversos e independientes, lo que confiere al modelo una mayor robustez y reduce la correlación entre las predicciones individuales.

4.5. Evaluación

La evaluación del desempeño del modelo se llevó a cabo utilizando el subconjunto de pruebas, garantizando así una validación imparcial de su capacidad de generalización. Tras generar las predicciones con el modelo entrenado, se visualizaron y evaluaron los resultados obtenidos.

En la Figura 2 se muestra una comparación entre los resultados reales del conjunto de prueba y las predicciones realizadas por el modelo, utilizando una matriz de confusión. Esta herramienta permite identificar con mayor precisión en qué casos el modelo acierta y en cuáles se equivoca. De esta manera, obtenemos una visión más detallada de su desempeño, ya que no solo vemos el porcentaje de aciertos generales, sino también los tipos de errores que comete.

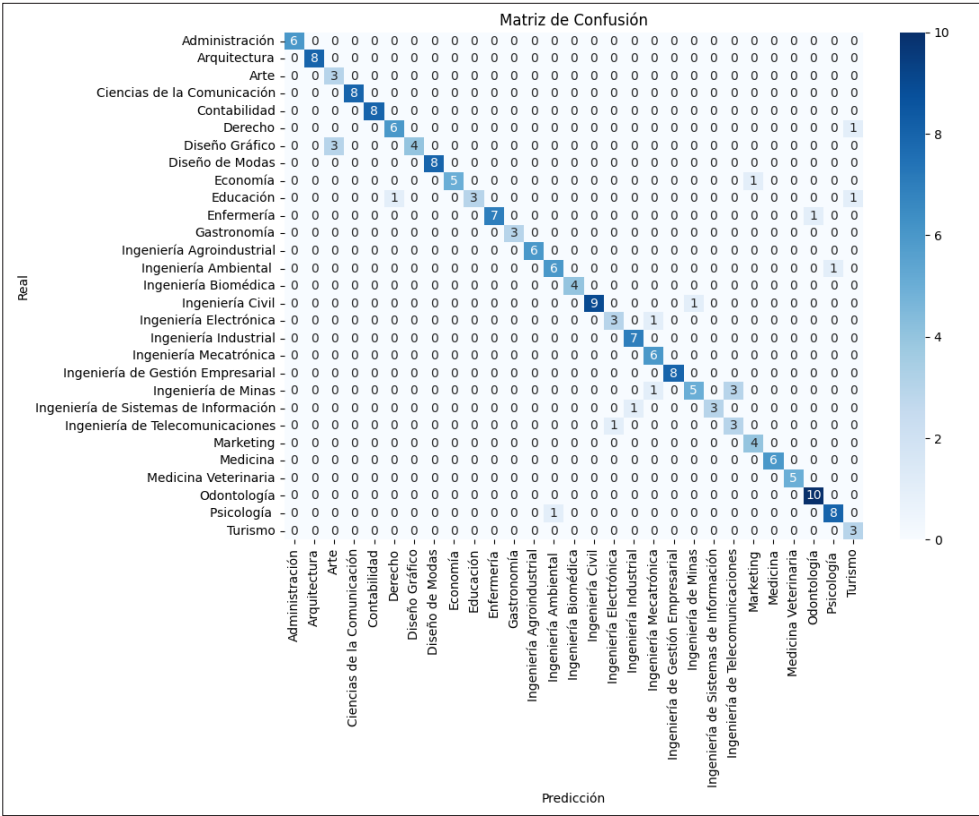


Figura 2 - Matriz de confusión

Asimismo, el análisis de la importancia de las características, una propiedad intrínseca del modelo Random Forest, fue crucial para interpretar la contribución de cada variable predictora. La Figura 3 demuestra que, de las variables evaluadas, las áreas de interés obtenidas a partir del test CHASIDE y los cursos preferidos de los estudiantes presentaron el mayor peso predictivo. Este hallazgo demuestra la importancia de los intereses vocacionales para la predicción de carreras.

Asimismo, como se detalla en la Tabla 4, los resultados obtenidos para las métricas clave oscilaron consistentemente entre el 78% y el 80%. Esta baja variabilidad y los altos valores en métricas como Accuracy, Precision, Recall, F1-score y Cohen's Kappa demuestran la fiabilidad del modelo, indicando un equilibrio sólido y una alta capacidad predictiva sobre datos no vistos. Este enfoque de evaluación se alinea con las recomendaciones metodológicas señaladas por Borja et al. (2024) respecto a la estandarización e interpretación coherente de métricas en clasificadores de Machine Learning y Deep Learning.

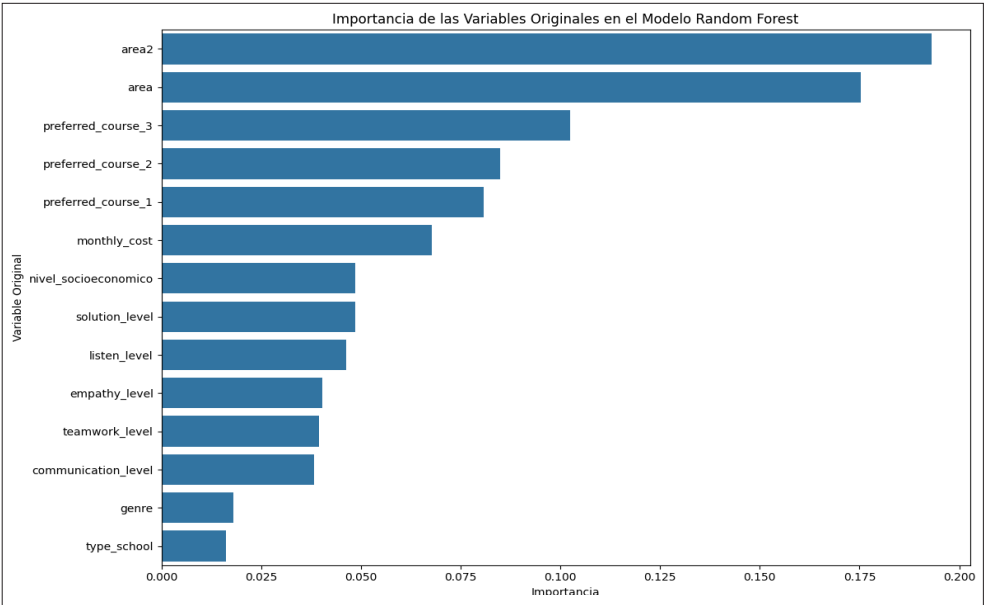


Figura 3 – Importancia de las variables

Con el objetivo de validar la efectividad del modelo propuesto, se entrenaron modelos adicionales utilizando los mismos datos y bajo las mismas condiciones experimentales. En esta etapa comparativa se incluyeron algoritmos clásicos de clasificación supervisada como Árbol de Decisión y K-Nearest Neighbors (KNN), los cuales fueron seleccionados por su amplia utilización en tareas de clasificación multiclase. Para garantizar una evaluación equitativa, a cada modelo se le aplicaron exactamente las mismas métricas de rendimiento, siguiendo las prácticas de evaluación homogénea recomendadas en la literatura (Borja et al., 2024). Los resultados obtenidos revelaron una diferencia significativa en el desempeño, evidenciando que el modelo basado en Random Forest superó considerablemente a los demás en términos de capacidad predictiva y estabilidad. En la siguiente tabla se detallan los resultados de esta comparación:

	Random Forest	Árbol de Decisión	KNN
Accuracy	90.16%	72.13%	69.95%
Precision	90.07%	71.34%	68.95%
Recall	90.31%	72.54%	70.08%
F1-Score	88.94%	70.25%	67.44%
Cohen’s Kappa	89.78%	71.06%	68.79%
ROC AUC	99.87%	95.98%	96.13%
R2 en regresión lineal	85.08%	59.23%	46.59%

Tabla 4 – Comparación de Algoritmos de Clasificación Supervisada

5. Conclusiones

El presente estudio demostró que el modelo de recomendación de carreras universitarias basado en Random Forest alcanzó un desempeño superior frente a otros algoritmos clásicos como Árbol de Decisión y K-Nearest Neighbors, logrando métricas mayores al 90 % en precisión, recall y F1-score. Estos resultados validan la capacidad del modelo para ofrecer recomendaciones confiables y personalizadas a estudiantes de secundaria.

A diferencia de propuestas previas que solo predicen riesgos de deserción o sugieren carreras de forma general, este trabajo aporta un enfoque integral que considera variables sociodemográficas, académicas y vocacionales, fortalecidas con instrumentos psicométricos. Con ello, se contribuye a la prevención de la deserción universitaria desde la etapa preuniversitaria, ofreciendo una herramienta de apoyo para la toma de decisiones académicas.

El impacto de esta propuesta radica en su aplicabilidad en el contexto educativo peruano, donde la falta de orientación vocacional constituye un factor crítico en el abandono universitario. La implementación de sistemas como el aquí presentado puede ayudar a los colegios y universidades a guiar de manera más efectiva a los jóvenes en su elección profesional.

Referencias

- Acosta, C., Peñaranda, E., Monsalvo, J., & Hernández, I. (2024). Instrumento innovador para la Orientación Vocacional: Cuestionario de Habilidades enfocado a las Inteligencias Múltiples. *Revista De Ciencias Sociales*, 30(3), 158–172. <https://doi.org/10.31876/rsc.v30i3.42656>
- Aguilar, A. (2024). La encrucijada juvenil: vocación vs. presión familiar en la elección de la carrera universitaria. *Infobae*. <https://www.infobae.com/peru/2024/04/06/la-encrucijada-juvenil-vocacion-vs-presion-familiar-en-la-eleccion-de-carrera-universitaria-segun-los-especialistas/>
- Alvarado, J., Mejía, P., Masetto, A. L., Molontay, R., Hilliger, I., Hegde, V., Montemayor, J. E., Ramírez, R. A., & Ceballos, H. G. (2022). Student Dataset from Tecnológico de Monterrey in Mexico to Predict Dropout in Higher Education. *Data*, 7(9). <https://doi.org/10.3390/data7090119>
- Arias, A., Linares, M., & Héndez, N. R. (2024). Undergraduate Dropout in Colombia: A Systematic Literature Review of Causes and Solutions. *Journal of Latinos and Education*, 23(2), 612–627. <https://doi.org/10.1080/15348431.2023.2171042>
- Assiri, B., Bashraheel, M., & Alsuri, A. (2024). Enhanced Student Admission Procedures at Universities Using Data Mining and Machine Learning Techniques. *Applied Sciences*, 14(3). <https://doi.org/10.3390/app14031109>
- Atalla, S., Daradkeh, M., Gawanmeh, A., Khalil, H., Mansoor, W., Miniaoui, S., & Himeur, Y. (2023). An Intelligent Recommendation System for Automating Academic Advising Based on Curriculum Analysis and Performance Modeling. *MATHEMATICS*, 11(5). <https://doi.org/10.3390/math11051098>

- Borja-Robalino, R., Monleon-Getino, A., Monleón-Getino, A., & Rodellar, J. (2020). Estandarización de métricas de rendimiento para clasificadores Machine y Deep Learning. *Revista Ibérica de Sistemas e Tecnologias de Informação*, 30, 184–196. <https://www.researchgate.net/publication/342009715>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Cisneros, B. E. (2023). *ALGORITMO DE MACHINE LEARNING Y USO DE PROPIEDADES SEMÁNTICAS PARA LA IDENTIFICACIÓN Y SUGERENCIA VOCACIONAL*. [Tesis de maestría, Universidad Autónoma del Estado de México] <http://hdl.handle.net/20.500.11799/138519>
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334.
- Domingues, A. S. B., da Silva, J. M. T., de Albuquerque Paixão, M. P. B., & da Silva David, R. M. F. (2024). Career indecision profile-short version: validity evidence. *International Journal for Educational and Vocational Guidance*. <https://doi.org/10.1007/s10775-024-09705-3>
- Escalante, J. I., Medina, C. J., & Vásquez, A. (2023). La deserción universitaria: un problema no resuelto en el Perú. *Hacedor - AIAPÆC*, 7(1), 60–72. <https://doi.org/10.26495/rch.v7i1.2421>
- Fantinelli, S., Esposito, C., Carlucci, L., Limone, P., & Sulla, F. (2023). The Influence of Individual and Contextual Factors on the Vocational Choices of Adolescents and Their Impact on Well-Being. *Behavioral Sciences*, 13(3). <https://doi.org/10.3390/bs13030233>
- Fernández, M., Álvarez, D., Fernández, F. B., & González, E. (2024). Dropout in Andalusian universities: prediction and prevention. *Frontiers in Education*, 8. <https://doi.org/10.3389/feduc.2023.1304016>
- Gedrimiene, E., Celik, I., Mäkitalo, K., & Muukkonen, H. (2023). Transparency and Trustworthiness in User Intentions to Follow Career Recommendations from a Learning Analytics Tool. *Journal of Learning Analytics*, 10(1), 54–70. <https://doi.org/10.18608/jla.2023.7791>
- George, D., & Mallery, P. (2010). *SPSS for Windows Step by Step: A Simple Guide and Reference* (10th ed.). Pearson.
- Goran, R., Jovanovic, L., Bacanin, N., Stankovic, M., Simic, V., Antonijevic, M., & Zivkovic, M. (2024). Identifying and understanding student dropouts using metaheuristic optimized classifiers and explainable artificial intelligence techniques. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2024.3446653>
- Gutierrez, D. A., Garcia-, G., Cuadros, E., Camara, G., & Gomez, E. (2023). Supporting Decision-Making Process on Higher Education Dropout by Analyzing Academic, Socioeconomic, and Equity Factors through Machine Learning and Survival Analysis Methods in the Latin American Context. *Education Sciences*, 13(2). <https://doi.org/10.3390/educsci13020154>

- Hamdollahi Oskouei, S., & Hashemzadeh, M. (2023). FoodRecNet: a comprehensively personalized food recommender system using deep neural networks. *Knowledge and Information Systems*, 65(9), 3753–3775. <https://doi.org/10.1007/s10115-023-01897-4>
- Huayta, J., & Pacheco, A. (2024). Improving vocational guidance through an expert system: Temporary decreasing and enhancing student self-awareness. *F1000Research*, 13, 232. <https://doi.org/10.12688/f1000research.145109.1>
- Instituto Nacional de Estadística e Informática [INEI]. (2024). *Perú: Evolución de la Pobreza Monetaria, 2014-2023*.
- Kranz, A., Plappert, C., Graf, J., Abele, H., Schulz, A. A., & Wirtz, M. A. (2023). Assessment of the relevance of midwifery competencies in academic education in Germany from the midwives' perspective: A structural analysis of cross sectional survey data. *European Journal of Midwifery*, 7(September), 1–13. <https://doi.org/10.18332/ejm/169658>
- Langer, Á. I., Ponce, F. P., Ordóñez, J. L., Fuentes, R., Mac, S., Gaete, J., & Núñez, D. (2024). Psychometric evidence of the Acceptance and Action Questionnaire II (AAQ-II): an item response theory analysis in university students from Chile. *BMC Psychology*, 12(1). <https://doi.org/10.1186/s40359-024-01608-w>
- Li, M., & Niu, L. (2023). Exercise Recommendation with Preferences and Expectations Based on Ability Computation. *Computers, Materials and Continua*, 77(1), 263–284. <https://doi.org/10.32604/cmc.2023.041193>
- Llauró, A., Fonseca, D., Villegas, E., Aláez, M., & Romero, S. (2023). Improvement of Academic Analytics Processes Through the Identification of the Main Variables Affecting Early Dropout of First-Year Students in Technical Degrees. A Case Study. *INTERNATIONAL JOURNAL OF INTERACTIVE MULTIMEDIA AND ARTIFICIAL INTELLIGENCE*. <https://doi.org/10.9781/ijimai.2023.06.002>
- López, D., Álvarez, P. R., & González, N. (2023). Model for Adaptability of Students in Transition to University. *Revista Electronica de Investigacion Educativa*, 25. <https://doi.org/10.24320/REDIE.2023.25.E10.4398>
- Mahale, Y., Khan, N., Kulkarni, K., Wagle, S. A., Pareek, P., Kotecha, K., Kotecha, K., Choudhury, T., & Sharma, A. (2024). Crop recommendation and forecasting system for Maharashtra using machine learning with LSTM: a novel expectation-maximization technique. *Discover Sustainability*, 5(1). <https://doi.org/10.1007/s43621-024-00292-5>
- Majjate, H., Bellarhmouch, Y., Jeghal, A., Yahyaouy, A., Tairi, H., & Zidani, K. A. (2024). AI-Powered Academic Guidance and Counseling System Based on Student Profile and Interests. *Applied System Innovation*, 7(1). <https://doi.org/10.3390/asi7010006>
- Maphosa, M., Doorsamy, W., & Paul, B. (2024). Improving Academic Advising in Engineering Education with Machine Learning Using a Real-World Dataset. *Algorithms*, 17(2). <https://doi.org/10.3390/a17020085>

- Mbanga, R., Ratelle, C. F., & Duchesne, S. (2024). Bidirectional and longitudinal associations between academic motivation and vocational indecision. *Motivation and Emotion*, 48(4), 573–588. <https://doi.org/10.1007/s11031-024-10082-1>
- Paranjape, V., Nihalani, N., & Mishra, N. (2024). Design and Development of an Efficient Demographic-based Movie Recommender System using Hybrid Machine Learning Techniques. *INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL*, 19(4). <https://doi.org/10.15837/ijccc.2024.4.5840>
- Rico, J. R., Cachero, C., & Macià, H. (2024). Study regarding the influence of a student's personality and an LMS usage profile on learning performance using machine learning techniques. *Applied Intelligence*, 54(8), 6175–6197. <https://doi.org/10.1007/s10489-024-05483-2>
- Santana, L. E., Ruiz, Z., & Feliciano, L. (2023). Gender stereotypes and vocational variables in secondary education female students. *Revista de Educacion*, 2023(400), 247–274. <https://doi.org/10.4438/1988-592X-RE-2023-400-578>
- Siddique, M. T., Koukaras, P., Ioannidis, D., & Tjortjis, C. (2023). SmartBuild RecSys: A Recommendation System Based on the Smart Readiness Indicator for Energy Efficiency in Buildings. *Algorithms*, 16(10). <https://doi.org/10.3390/a16100482>
- Siswipraptini, P. C., Warnars, H. L. H. S., Ramadhan, A., & Budiharto, W. (2024). Personalized Career-Path Recommendation Model for Information Technology Students in Indonesia. *IEEE Access*, 12, 49092–49105. <https://doi.org/10.1109/ACCESS.2024.3381032>
- Song, Q. C., Shin, H. J., Tang, C., Hanna, A., & Behrend, T. (2024). Investigating machine learning's capacity to enhance the prediction of career choices. *PERSONNEL PSYCHOLOGY*, 77(2), 295–319. <https://doi.org/10.1111/peps.12529>
- Tolentino, S. (2023). Estas son las 10 carreras en el Perú con más deserción universitaria en el primer año de estudios. *La República*. <https://www.proquest.com/newspapers/estas-son-las-10-carreras-en-el-perú-con-más/docview/2860544769/se-2?accountid=43860>
- Urbina-Nájera, A. B., & Méndez-Ortega, L. A. (2022). Predictive Model for Taking Decision to Prevent University Dropout. *INTERNATIONAL JOURNAL OF INTERACTIVE MULTIMEDIA AND ARTIFICIAL INTELLIGENCE*, 7(4), 205–213. <https://doi.org/10.9781/ijimai.2022.01.006>
- Valencia, A., Chalela, S., Cadavid, M., Gallegos, A., Benjumea, M., & Rodríguez, D. Y. (2023). University Dropout Model for Developing Countries: A Colombian Context Approach. *BEHAVIORAL SCIENCES*, 13(5). <https://doi.org/10.3390/bs13050382>
- Valtolina, S., Anibal Matamoros, R., & Epifania, F. (2024). Design of a conversational recommender system in education. *User Modeling and User-Adapted Interaction*. <https://doi.org/10.1007/s11257-024-09397-y>
- Villarreal, H., Ángeles, J., Marín, W., & Cano, J. (2024). Modelo de clasificación para la deserción estudiantil en las universidades públicas del Perú. *Revista de Ciencias Sociales*, 30(1), 452–469. <https://doi.org/10.31876/rcs.v30i1.41667>

- Vives, L., Cabezas, I., Vives, J. C., Reyes, N. G., Aquino, J., Córdor, J. B., & Altamirano, S. F. S. (2024). Prediction of Students' Academic Performance in the Programming Fundamentals Course Using Long Short-Term Memory Neural Networks. *IEEE ACCESS*, 12, 5882–5898. <https://doi.org/10.1109/ACCESS.2024.3350163>
- Wang, P., Zheng, Y., Zhang, M., Yin, K., Geng, F., Zheng, F., Ma, J., & Wu, X. (2024). Methods for measuring career readiness of high school students: based on multidimensional item response theory and text mining. *Humanities and Social Sciences Communications*, 11(1). <https://doi.org/10.1057/s41599-024-03436-0>
- Xu, Z., Chu, C., & Song, S. (2024). An Effective Federated Recommendation Framework with Differential Privacy. *Electronics*, 13(8). <https://doi.org/10.3390/electronics13081589>
- Zhang, Y. (2024). Path of career planning and employment strategy based on deep learning in the information age. *PLoS One*, 19(10). <https://doi.org/https://doi.org/10.1371/journal.pone.0308654>

CATCH: A Nova Fronteira dos Chatbots na Gestão de Força de Trabalhos

Elon Oliveira Albuquerque¹, Bruno Justino Garcia Praciano¹,
Paulo Henrique Batista Rodrigues¹, Flávio Garcia Praciano¹,
Fábio L. Lopes de Mendonça¹, Geraldo P. Rocha Filho²

elon.albuquerque@gmail.com; bruno.justino@ieee.org; paulo.rodrigues@redes.unb.br;
flavio.praciano@redes.unb.br; fabio.mendonca@redes.unb.br; geraldo.rocha@uesb.edu.br

¹Universidade de Brasília - UnB, Brazil.

² Universidade Estadual do Sudoeste da Bahia - UESB, Brazil. Email: geraldo.rocha@uesb.edu.br

Pages: 157-169

Resumo: O Dimensionamento da Força de Trabalho (DFT) é um desafio essencial para a gestão de recursos humanos, especialmente no setor público, em que a eficiência e a precisão na alocação de pessoal são cruciais. A fase qualitativa do DFT, que depende de métodos manuais e subjetivos, enfrenta problemas significativos de imprecisão e ineficiência, criando uma lacuna na precisão das previsões e na gestão das demandas de trabalho. Em resposta a essas limitações, este trabalho propõe o CATCH – Chatbot para Automação de Tarefas e Comunicação Humanizada – como uma solução para aprimorar o DFT. O CATCH foi desenvolvido para automatizar a coleta e análise de dados qualitativos por meio de interações baseadas em inteligência artificial, o que resulta em uma abordagem mais precisa e eficiente para entender as necessidades de pessoal e otimizar a alocação de recursos. Em comparação com modelos existentes na literatura, o CATCH apresentou resultados superiores em termos de acurácia, com uma taxa de 0,9, tempo de resposta reduzido para 1,5 segundos e um detalhamento das respostas de 0,9. Esses resultados demonstram uma melhoria significativa na precisão das previsões e na eficiência do processo de DFT.

Palavras-chave: Inteligência Artificial, Dimensionamento da Força de Trabalho, Qualidade dos Serviços, Automação de Coleta de Dados, Chatbot.

CATCH: The New Frontier of Chatbots in Workforce Management

Abstract: Workforce Sizing (WFS) is an essential challenge for human resource management, especially in the public sector, where efficiency and accuracy in staff allocation are crucial. The qualitative phase of WFS, which relies on manual and subjective methods, faces significant issues of inaccuracy and inefficiency, creating a gap in forecast precision and workload management. In response to these limitations, this paper proposes CATCH – Chatbot for Automated Tasks and Humanized Communication – as a solution to enhance WFS. CATCH was developed to automate the collection and analysis of qualitative data through AI-based

interactions, resulting in a more accurate and efficient approach to understanding personnel needs and optimizing resource allocation. Compared to existing models in the literature, CATCH showed superior results in terms of accuracy, with a 0.9 rate, reduced response time to 1.5 seconds, and a response detail score of 0.9. These results demonstrate a significant improvement in forecast precision and process efficiency in WFS.

Keywords: Artificial Intelligence, Workforce Sizing, Service Quality, Data Collection Automation, Chatbot.

1. Introdução

O Dimensionamento da Força de Trabalho (DFT) representa um desafio crucial para as unidades de gestão de pessoas Marques et al. (2022); Silva et al. (2024). Este processo é fundamental para garantir que o número ideal de profissionais especializados por categoria seja definido, de modo a atender às demandas da máquina pública Serrano et al. (2021). Em um contexto de incerteza macroeconômica, adotar práticas e processos inovadores é vital. Embora a iniciativa privada frequentemente adote tais práticas, elas ainda são limitadas no setor público da Silva et al. (2019). Novas tecnologias, como *Robotic Process Automation* (RPA), análise preditiva, *chatbots* e sistemas de gestão com inteligência artificial, são essenciais para aprimorar a eficácia operacional no setor público de Oliveira Mendes et al. (2020); Serrano et al. (2017).

O processo de DFT envolve várias etapas cruciais, incluindo a previsão da demanda, a previsão de suprimento e as estratégias para equilibrá-los Chowdhury (2016). As metodologias para o DFT podem ser divididas em duas fases. A fase qualitativa foca no levantamento das atividades, geralmente por meio de entrevistas, questionários e análise de tarefas, com o objetivo de entender as necessidades e o volume de trabalho. Em contraste, a fase quantitativa avalia a produtividade e o desempenho das atividades identificadas, utilizando dados históricos e métricas para calcular o número ideal de profissionais necessários. Essas duas fases juntas permitem uma análise abrangente e uma previsão mais precisa das necessidades de pessoal.

A fase qualitativa do DFT que é o foco desta pesquisa enfrenta várias limitações. Ela depende de métodos manuais e subjetivos, o que frequentemente resulta em imprecisões e ineficiências Marques et al. (2022); Silva et al. (2024). Além disso, a variabilidade nas respostas e interpretações dos envolvidos pode levar a inconsistências nos dados coletados, prejudicando a qualidade das informações utilizadas para o planejamento Silva et al. (2024). A introdução de *chatbots* transforma significativamente essa fase ao automatizar a coleta e análise de dados sobre as demandas de trabalho e a alocação de recursos. *Chatbots* são sistemas baseados em inteligência artificial projetados para interagir com usuários, fornecer respostas automatizadas e coletar informações detalhadas sobre processos e atividades. Essa automação não só melhora a precisão das previsões como também facilita a otimização da força de trabalho, contribuindo para a redução de custos e o aumento da eficiência do processo Pu et al. (2024); Rane et al. (2024); Silva et al. (2024).

Para sanar esse problema e otimizar o processo de DFT, este artigo apresenta o CATCH - *Chatbot* para Automatização de Tarefas e Comunicação Humanizada. Diferente de outras soluções que focam apenas na automação de tarefas repetitivas ou no atendimento ao

cliente, o CATCH integra funcionalidades específicas para a análise e gestão de dados sobre a força de trabalho. O CATCH não apenas automatiza tarefas repetitivas, como responder a perguntas frequentes sobre o projeto, mas também melhora a coleta de dados qualitativos ao interagir diretamente com os funcionários e gerentes, coletando informações sobre as atividades e necessidades de pessoal. Isso permite uma análise mais precisa e em tempo real das demandas e do desempenho, facilitando a otimização da força de trabalho e contribuindo para a redução de custos e aumento da eficiência.

O restante deste artigo está organizado da seguinte forma: na Seção 2, são apresentados e discutidos os trabalhos relacionados na área. Na Seção 3, é descrito o modelo do CATCH. Na Seção 4, são apresentados os resultados obtidos para validar o CATCH. Por fim, a Seção 5 apresenta as conclusões e os trabalhos futuros.

2. Trabalhos Relacionados

Na literatura, existem trabalhos que adotam uma abordagem teórica e quantitativa para o DFT no setor público Serrano et al. (2017, 2021). Enquanto Serrano et al. (2021) foca na aplicação de modelos algébricos, estatísticos e de otimização para apoiar decisões estratégicas em cenários de incerteza econômica, Serrano et al. (2017) propõe uma metodologia baseada em modelagens matemáticas para instituições públicas federais, realizando uma análise das atividades organizacionais e o desenvolvimento de um sistema de apoio à decisão gerencial. Ambos os trabalhos compartilham uma visão teórica centrada no uso de modelos quantitativos para otimizar a gestão de recursos humanos, o que complementa a proposta do presente artigo, que descreve a aplicação prática de tecnologia, especificamente o desenvolvimento de um *chatbot* de inteligência artificial, para automatizar a coleta de dados qualitativos sobre as atividades das unidades de trabalho, otimizando processos não só no setor público.

Por outro lado, os estudos de da Silva et al. (2019) e de Oliveira Mendes et al. (2020) destacam a inovação na gestão pública, abordando a implementação de novos processos e tecnologias para melhorar a eficiência dos serviços. da Silva et al. (2019) foca na introdução de inovações no contexto da gestão pública municipal, analisando a percepção dos servidores sobre mudanças em um ambiente burocrático e legalmente rígido. Já de Oliveira Mendes et al. (2020) investiga o impacto do teletrabalho na administração pública, evidenciando os benefícios em termos de redução de custos administrativos e melhoria da qualidade de vida dos servidores. Ambos os estudos, embora focados em diferentes formas de inovação, reforçam a importância de modernizar a gestão pública. Entretanto, o presente trabalho diferencia-se por propor a automação por meio de um *chatbot*, integrando inteligência artificial para otimizar a coleta de dados e reduzir o uso de recursos humanos e financeiros no DFT.

Em Khennouche et al. (2024), os autores exploram uma taxonomia de *chatbots*, distinguindo entre modelos baseados em regras, recuperação e gerativos, como o ChatGPT. A pesquisa enfatiza a personalização e os desafios éticos na integração de *chatbots* com bases de conhecimento. Já o trabalho de Liu et al. (2024) investiga a eficácia de *chatbots* gerativos, confirmando que interações mais naturais, como as proporcionadas por esses *chatbots*, podem aumentar a eficácia das intervenções. Ambas as pesquisas destacam o potencial dos *chatbots* gerativos; porém, as limitações de ambos

os trabalhos residem na falta de foco em aplicações para a gestão da força de trabalho, como a análise de métricas de desempenho ou a otimização de processos. No contexto do DFT, tais funcionalidades são cruciais, mas não são abordadas de maneira direta nos estudos citados.

Ao contrário das abordagens supracitadas, que se concentram em métodos teóricos, quantitativos e inovações gerenciais, esta pesquisa introduz uma aplicação prática para o DFT ao propor um *chatbot* de inteligência artificial para automatizar a coleta e análise de dados qualitativos sobre as atividades das unidades de trabalho, como apresentado na Tabela 1. O objetivo é não apenas otimizar a gestão da força de trabalho, mas também reduzir significativamente o esforço humano necessário e aprimorar a precisão dos dados coletados. A pesquisa se destaca por integrar tecnologias avançadas que visam aumentar a eficiência, reduzir custos e preencher lacunas deixadas por estudos anteriores, oferecendo uma solução prática e adaptável para a análise e otimização das atividades e métricas de desempenho no setor público.

Trabalhos	Abordagem teórica e quantitativa para o DFT	Inovação na gestão pública com novas tecnologias	Automação da coleta de dados qualitativos	Redução de recursos humanos e financeiros
Serrano et al. (2021)	Sim	Não	Não	Não
Serrano et al. (2017)	Sim	Não	Não	Não
da Silva et al. (2019)	Não	Sim	Não	Não
de Oliveira Mendes et al. (2020)	Não	Sim	Não	Não
Khenrouche et al. (2024)	Não	Não	Não	Não
Liu et al. (2024)	Não	Não	Não	Não
CATCH	Sim	Sim	Sim	Sim

Tabela 1 – Sumarização das contribuições dos trabalhos relacionados e do trabalho proposto.

3. CATCH – Chatbot para Automatização de Tarefas e comunicação Humanizada

Esta seção apresenta o CATCH, um *Chatbot* para Automatização de Tarefas e Comunicação Humanizada com o intuito de otimizar o processo de DFT. Desenvolvido com base em Inteligência Artificial (IA) e em modelos de Processamento de Linguagem Natural (PLN), o CATCH visa automatizar a coleta e análise de dados qualitativos, permitindo que os gestores tomem decisões estratégicas com base em dados precisos e em tempo real. A integração dessas tecnologias busca garantir maior eficiência no gerenciamento da força de trabalho.

3.1. Arquitetura do CATCH

A arquitetura do CATCH é projetada para fornecer respostas precisas e contextualizadas através de uma estrutura modular que integra tecnologias de PLN e recuperação de informações, conforme ilustrada na Figura 1. O CATCH foi modelado com base no

ChromaDB, um banco de dados vetorial para armazenar e recuperar grandes volumes de informações contextuais, e o LLaMA3, um modelo de linguagem que, em combinação com o ChromaDB, gera respostas contextualizadas. O ChromaDB é responsável pela indexação e busca de passagens relevantes, enquanto o LLaMA3 emprega técnicas de modelagem de linguagem para fornecer respostas informadas e pertinentes ao contexto da pergunta.

A interação com o CATCH inicia quando o usuário realiza a submissão de perguntas, a qual foi implementada com o framework FastAPI (Rotulo 1, Figura 1). Esta API gerencia as solicitações e garante a comunicação fluida entre o usuário e o *backend* do sistema. A implementação do CORS (*Cross-Origin Resource Sharing*) assegura a compatibilidade do CATCH com diferentes ambientes web, permitindo uma integração transparente com diversas plataformas.

Após a validação da solicitação, a pergunta é encaminhada para o processamento de linguagem (Rotulo 2, Figura 1). Neste estágio, a pergunta é convertida em uma representação vetorial utilizando o modelo de *Embeddings SentenceTransformer*. Este modelo é especializado em capturar a semântica e o significado das questões, facilitando uma compreensão mais profunda do conteúdo. A representação vetorial resultante é então utilizada pelo ChromaDB para buscar passagens relevantes armazenadas, fornecendo uma base sólida de informações contextuais.

Com as passagens relevantes recuperadas, o CATCH utiliza o modelo de Geração Aumentada por Recuperação (RAG) para combinar essas informações com o conhecimento do modelo de linguagem LLaMA3 (Rotulo 3, Figura 1). O modelo RAG é eficaz na integração das informações recuperadas com a geração de linguagem natural, produzindo respostas enriquecidas e personalizadas. A resposta final é então apresentada ao usuário através da interface do CATCH, completando o ciclo de interação.

3.2. Gerenciamento de Consultas e Controle de Acesso no CATCH

O CATCH utiliza o FastAPI, um *framework* para construção de APIs web, com o objetivo de gerenciar de forma otimizada as requisições dos usuários. Esse *framework* foi escolhido por sua capacidade de lidar com operações assíncronas e sua compatibilidade com Python 3.7+, proporcionando um gerenciamento eficiente do ciclo de vida das requisições e melhorando a latência do sistema. A funcionalidade de submissão de consultas dos usuários (Rotulo 1, Figura 1) é implementada diretamente pelo FastAPI, permitindo o processamento paralelo de requisições em tempo real. Além disso, o FastAPI oferece suporte a recursos como a validação de dados automatizada através do Pydantic, facilitando a detecção de inconsistências nos dados enviados e a redução de falhas no processamento. O FastAPI também gera automaticamente documentação interativa compatível com *OpenAPI* e Swagger, facilitando o uso da API e a integração com outras ferramentas de desenvolvimento. Dessa forma, o sistema melhora a responsividade e mantém a consistência nas interações.

Para garantir que o CATCH seja acessível a partir de diferentes domínios web, foi implementado o CORS, um componente que controla a política de compartilhamento de recursos entre origens distintas. O CORS é responsável por configurar regras que garantem que apenas requisições de domínios autorizados sejam aceitas, bloqueando

tentativas não autorizadas que poderiam comprometer a integridade e a segurança do sistema. Esse mecanismo é utilizado em ambientes onde a API do CATCH pode ser consumida por diferentes aplicações web distribuídas, promovendo uma integração com múltiplas plataformas. O CORS, ao configurar políticas de controle de acesso, permite que o CATCH atenda a solicitações legítimas sem comprometer a segurança, promovendo interoperabilidade e acessibilidade em um ambiente web distribuído. Esse mecanismo assegura que a comunicação entre o *frontend* e o *backend* seja contínua e eficiente, independentemente da origem do acesso, permitindo que o CATCH opere de maneira coesa em diversos contextos de uso e integração.

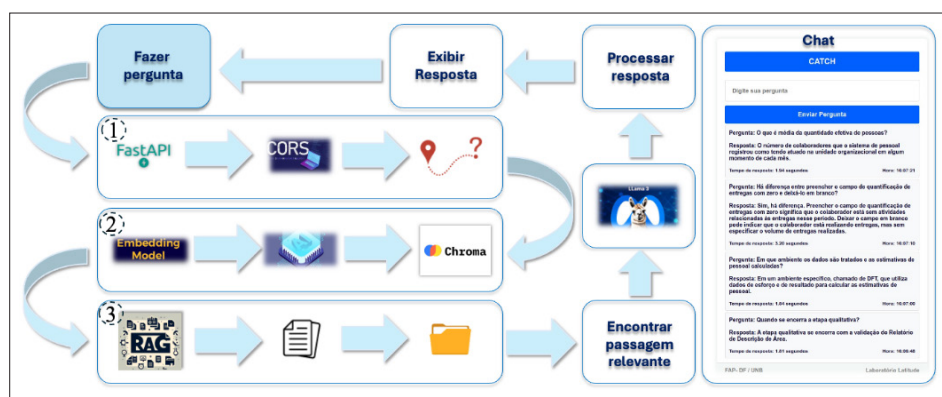


Figura 1 – Visão geral da arquitetura do CATCH

3.3. Processamento de Embedding no CATCH

No CATCH, o modelo *SentenceTransformer* é responsável pela conversão das perguntas dos usuários em representações vetoriais de alta dimensionalidade, facilitando o processo de busca e recuperação de informações contextuais. O *SentenceTransformer* é especializado em capturar a semântica e o significado das consultas, gerando vetores que encapsulam tanto o conteúdo explícito quanto o contexto implícito das perguntas. Esses vetores são posteriormente utilizados pelo ChromaDB, um banco de dados vetorial otimizado para armazenamento e recuperação de grandes volumes de dados, projetado para realizar buscas baseadas em similaridade semântica de maneira eficiente.

O ChromaDB desempenha um papel fundamental na infraestrutura do CATCH, permitindo a indexação e recuperação de passagens relevantes com base nas representações vetoriais geradas pelo *SentenceTransformer*. Este banco de dados utiliza técnicas de indexação vetorial, como árvores de busca aproximada, para garantir que as consultas sejam respondidas com baixa latência, mesmo com grandes volumes de dados armazenados. Dessa forma, o CATCH é capaz de localizar informações pertinentes com precisão e eficiência.

A integração entre o *SentenceTransformer* e o ChromaDB possibilita ao CATCH realizar uma recuperação eficiente de informações contextuais em cenários complexos. O fluxo de trabalho inicia-se com a conversão das perguntas em vetores pelo *SentenceTransformer*, que gera uma representação densa de cada consulta. Em seguida, o ChromaDB utiliza

essas representações para buscar as passagens mais relevantes, com base em medidas de similaridade semântica. Após a recuperação das passagens, o modelo RAG combina essas informações com o modelo de linguagem LLaMA3, que gera respostas contextualizadas. Esse processo de integração entre recuperação de informações e geração de linguagem natural assegura que as respostas fornecidas pelo CATCH sejam coerentes e alinhadas ao contexto original da consulta.

3.4. Mecanismo para Geração de Respostas no CATCH

No CATCH, a integração entre o modelo RAG e o LLaMA3 é fundamental para assegurar a precisão e relevância contextual das respostas geradas. Essa combinação não é trivial, pois envolve a coordenação entre a recuperação eficiente de informações e a geração de linguagem natural coerente. O RAG desempenha a função de recuperar passagens contextuais relevantes a partir de uma base de dados, enquanto o LLaMA3 utiliza essas passagens para gerar respostas que não apenas refletem o conteúdo das informações recuperadas, mas também estão alinhadas ao contexto específico da consulta do usuário. Esse processo, ilustrado no Algoritmo 1, garante que as respostas geradas pelo CATCH sejam tanto informativas quanto adequadamente contextualizadas, lidando com a complexidade e a necessidade de precisão em interações de linguagem natural.

O RAG foi modelado para combinar técnicas de recuperação de informações com geração de texto, operando em duas fases principais: a recuperação de passagens e a geração da resposta final. Na fase de recuperação, o RAG busca as informações relevantes em um banco de dados utilizando uma abordagem baseada em vetores. No CATCH, o RAG acessa o ChromaDB para recuperar passagens pertinentes que estão indexadas no banco de dados. A recuperação é realizada com base na representação vetorial da consulta do usuário, gerada pelo modelo *SentenceTransformer*. Essa representação vetorial captura a semântica e o significado da pergunta, permitindo que o RAG localize rapidamente as passagens que mais se aproximam do contexto da consulta, facilitando uma recuperação eficiente.

Após a fase de recuperação, o RAG passa para a fase de geração, em que combina as passagens recuperadas com o conhecimento do modelo de linguagem LLaMA3. O LLaMA3 utiliza as informações recuperadas para gerar uma resposta adaptada ao contexto específico da consulta. Essa abordagem de geração aumentada permite que o CATCH produza respostas que não são apenas precisas em termos de conteúdo, mas também altamente contextualizadas. O RAG garante que as passagens recuperadas sejam integradas de forma coesa e relevante, resultando em respostas que abordam diretamente a consulta do usuário, mantendo a precisão sem comprometer a fluidez.

O LLaMA3 tem um papel central na geração de respostas, empregando técnicas de PLN para interpretar as passagens recuperadas e produzir respostas que sejam gramaticalmente corretas e semanticamente adequadas. O modelo, desenvolvido com base na arquitetura *Transformer*, utiliza mecanismos de atenção multi-cabeça e codificação posicional para capturar relações de longo alcance entre as palavras, permitindo uma compreensão detalhada da estrutura textual. Pré-treinado em grandes volumes de dados, o LLaMA3 é ajustado para tarefas específicas, possibilitando uma interpretação robusta de nuances e contextos complexos, que são características essenciais para a aplicação em questão.

Algoritmo 1 Mecanismo de Geração de Respostas no CATCH**Inicialização:**

- 1: Carregar modelo de embeddings E
- 2: Carregar banco de dados vetorial D
- 3: Carregar modelo de recuperação e geração G
- 4: Definir consulta do usuário q
- 5: Definir número de passagens k e parâmetro de similaridade ϵ

Processamento da consulta:

- 1: **function** GERAREMBEDDING(q)
- 2: $embedding_q \leftarrow E.encode(q)$
- 3: **Retornar** $embedding_q$
- 4: **end function**

Recuperação de passagens relevantes:

- 1: **function** BUSCARINFORMACOES($embedding_q, D, k, \epsilon$)
- 2: $resultados \leftarrow D.query(embedding_q, k, \epsilon)$
- 3: **Retornar** $resultados$
- 4: **end function**

Geração de resposta:

- 1: **function** GERARRESPOSTA($resultados, G$)
- 2: $resposta \leftarrow G.generate(resultados)$
- 3: **Retornar** $resposta$
- 4: **end function**

Execução completa:

- 1: $embedding_q \leftarrow GERAREMBEDDING(q)$
- 2: $resultados \leftarrow BUSCARINFORMACOES(embedding_q, D, k, \epsilon)$
- 3: $resposta \leftarrow GERARRESPOSTA(resultados, G)$
- 4: **Retornar** $resposta$

No contexto do CATCH, o LLaMA3 processa as passagens recuperadas pelo RAG, utilizando técnicas de geração de texto baseadas em decodificadores para formular

respostas que sejam detalhadas e relevantes. O modelo também aplica máscaras de contexto para garantir a coerência e continuidade na geração de texto, mantendo a integridade das informações recuperadas e gerando respostas fluentes. Essas técnicas permitem que o LLaMA3 forneça respostas que não apenas respondam diretamente à pergunta, mas que também integrem e contextualizem as informações recuperadas, garantindo que a interação com o usuário seja precisa e adequada ao contexto da consulta.

4. Avaliação de Desempenho

Para avaliar o desempenho do mecanismo de geração de respostas implementado no CATCH, foram utilizados os modelos MISTRAL AI (2024) e PHI Microsoft (2024), além do baseline LLAMA 3.1 Llama (2024) para comparação. O objetivo principal foi medir a eficácia do CATCH em fornecer respostas adequadas, bem como avaliar a eficiência computacional durante o processamento das consultas.

O CATCH foi avaliado com base nas seguintes métricas:

Tempo de resposta: O tempo médio para gerar uma resposta foi registrado para cada consulta, incluindo as etapas de geração de *embeddings*, recuperação de passagens e síntese da resposta final.

Acurácia: A acurácia foi calculada como a proporção de respostas corretas (definidas como aquelas que correspondem à intenção da consulta) em relação ao total de consultas feitas.

Detalhamento: Esta métrica avalia o grau de profundidade e especificidade das respostas fornecidas, medindo o quanto o sistema consegue explorar nuances e fornecer informações adicionais relevantes para a consulta.

Os experimentos foram realizados em um servidor com Ubuntu 23.04 (Kernel 6.2.0-39-generic), equipado com processador Intel Core i7-10700F (16 núcleos, 16 threads), 32 GiB de RAM, armazenamento LVM de 921 GiB e seis GPUs NVIDIA GeForce RTX 3060 Ti com 8 GiB cada, utilizando Python 3.12.

4.1. Impacto dos Resultados Obtidos

A Figura 2 apresenta a comparação da acurácia dos modelos avaliados a partir de cinquenta questões sobre DFT, evidenciando a superioridade do CATCH, que alcançou acurácia de 0,90. Esse desempenho decorre da integração do modelo LLaMA 3 com ChromaDB e RAG, possibilitando recuperação contextual mais eficiente e respostas mais precisas. O LLaMA 3.1 obteve acurácia de 0,85, enquanto MISTRAL e PHI apresentaram desempenhos inferiores, com 0,75 e 0,70, respectivamente, indicando limitações nos mecanismos de recuperação e geração de respostas.

O gráfico de tempo de resposta (Figura 3) apresenta a comparação do desempenho dos modelos avaliados quanto ao tempo médio de processamento. O CATCH obteve o menor tempo de resposta (1,5 s), seguido pelo LLaMA 3.1 (1,7 s), enquanto o MISTRAL e o PHI apresentaram tempos superiores, de 2,5 s e 3,0 s, respectivamente. O melhor desempenho do CATCH está associado à sua arquitetura otimizada, baseada no FastAPI

e na integração eficiente entre ChromaDB, SentenceTransformer e o modelo RAG, o que reduz a latência nas etapas de recuperação e geração.

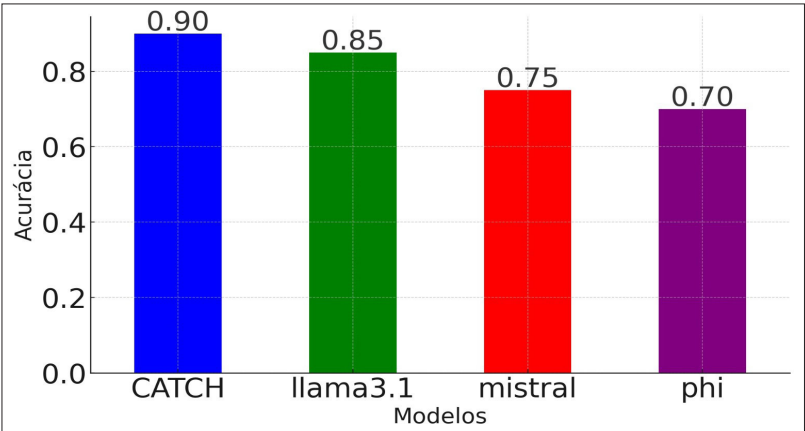


Figura 2 – Impacto da acurácia no CATCH

A Figura 4 apresenta a análise do grau de profundidade das respostas, medida pela métrica de detalhamento. O CATCH alcançou a maior pontuação (0,9), seguido pelo LLaMA 3.1 (0,85), MISTRAL (0,8) e PHI (0,7). Esse resultado evidencia a eficácia da integração entre o RAG e o LLaMA 3, aliada ao uso do ChromaDB e do SentenceTransformer, que favorece a recuperação contextual e a geração de respostas mais completas. Os demais modelos demonstraram menor nível de detalhamento, indicando limitações nos mecanismos de recuperação e combinação de informações, especialmente no caso do PHI.

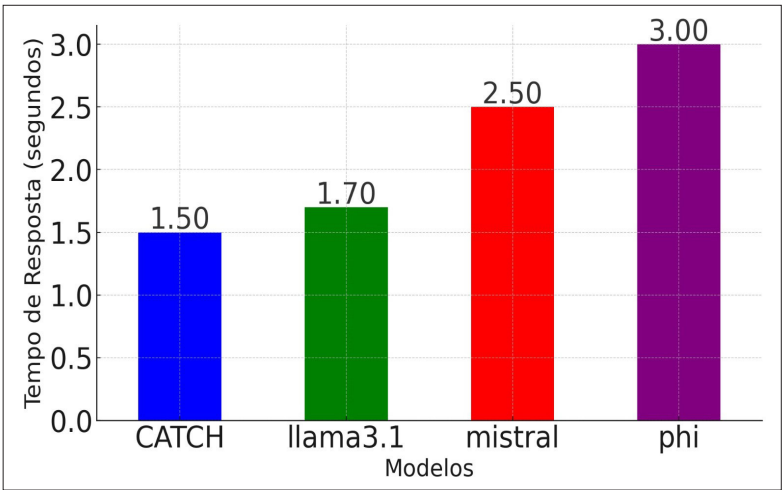


Figura 3 – Impacto do tempo de resposta no CATCH

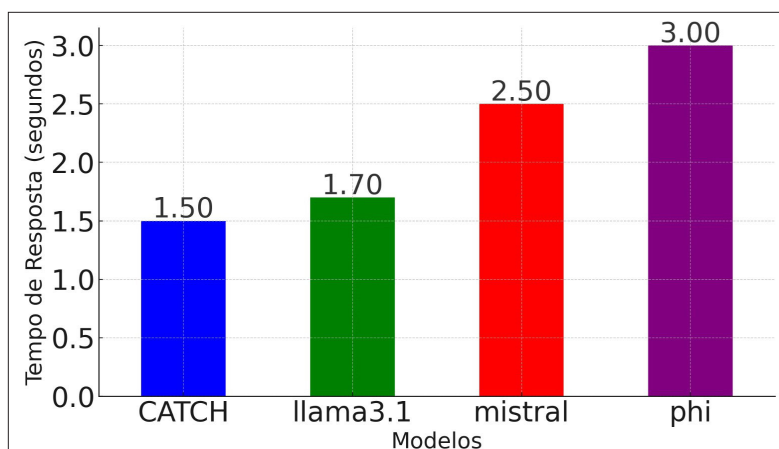


Figura 4 – Impacto do detalhamento no CATCH

4.2. Discussão de resultados

Os resultados do CATCH mostram avanços no processo de DFT, com uma acurácia de 90%, tempo de resposta de 1,5 segundos e detalhamento de 0,9 nas respostas. Esses números indicam que o CATCH supera os modelos da literatura, tais como o MISTRAL e o PHI, em precisão, rapidez e qualidade das respostas geradas. A arquitetura do CATCH, que integra o modelo LLaMA3 com o ChromaDB e o mecanismo de RAG, foi responsável pelo desempenho obtido. A capacidade de recuperar informações contextuais e gerar respostas detalhadas e alinhadas ao contexto permite que os gestores tomem decisões com base em dados, de forma mais eficiente no setor público.

O CATCH, no entanto, ainda apresenta limitações que precisam ser consideradas. Embora tenha sido eficaz no contexto público, sua aplicabilidade em outros setores e indústrias precisa ser avaliada. Além disso, como a qualidade das respostas depende da interação dos usuários com o sistema, melhorias futuras podem focar na implementação de feedback contínuo e no uso de técnicas de aprendizado adaptativo para ajustar o chatbot. Com isso, o CATCH poderá se adaptar às mudanças nas demandas de trabalho e expandir seu uso para outros contextos.

5. Conclusão e Trabalhos Futuros

O DFT no setor público enfrenta limitações relevantes na etapa qualitativa, em razão do uso de métodos manuais e subjetivos que impactam a precisão na previsão de pessoal. Este trabalho apresentou o CATCH, um chatbot baseado em IA e PLN para automatizar a coleta e análise de dados qualitativos, cujos resultados evidenciaram ganhos expressivos de eficiência e precisão, com acurácia de 90%, tempo médio de resposta de 1,5 segundos e desempenho superior a modelos como MISTRAL e PHI. Além da automação de tarefas repetitivas, o CATCH promoveu comunicação mais humanizada e suporte em tempo real à tomada de decisão estratégica, apontando como trabalhos futuros a incorporação

de aprendizado contínuo, feedback dos usuários e a expansão de sua aplicação para outros setores.

Agradecimentos

Este trabalho é apoiado pela Procuradoria Geral da Fazenda Nacional (nº PGFN 23106.148934/2019- 67). Em parte pelo CNPq – Conselho Nacional de Pesquisas (Nº PQ-2 312180/2019-5 de Cibersegurança nº 465741/2014-2), em parte pelo Ministério da Economia do Brasil (N.º DIPLA 005/2016) em parte pelo Conselho Administrativo de Defesa Econômica (Nº CADE 08700.000047/2019-14), em parte pela Advocacia Geral da União (nº AGU 697.935/2019), e em parte pela Fundação de Amparo a` Pesquisa do Distrito Federal – FAPDF.

Referencias

- AI, M. (2024). Mistral ai documentation. Acesso em 18 setembro 2024.
- Chowdhury, S. (2016). Estudos de otimização e melhoria de negócios na indústria upstream de petróleo e gás.
- da Silva, M. T., Pavan, D. P., Dechechi, E. C., da Silva Sampaio, V., and Panek, L. (2019). Dimensões da inovação no setor público: um estudo de caso nas prefeituras do oeste do paran . *Brazilian Journal of Development*, 5(11):25650–25675.
- de Oliveira Mendes, R. A., Oliveira, L. C. D., and Veiga, A. G. B. (2020). A viabilidade do teletrabalho na administra  o p blica brasileira. *Brazilian Journal of Development*, 6(3):12745–12759.
- Khennouche, F., Elmir, Y., Himeur, Y., Djebari, N., and Amira, A. (2024). Revolutionizing generative pre-trained: Insights and challenges in deploying chatgpt and generative chatbots for faqs. *Expert Systems with Applications*, 246:123224.
- Liu, I., Liu, F., Xiao, Y., Huang, Y., Wu, S., and Ni, S. (2024). Investigating the key success factors of chatbot-based positive psychology intervention with retrieval-and generative pre-trained transformer (gpt)-based chatbots. *International Journal of Human–Computer Interaction*, pages 1–12.
- Llama (2024). Llama official website. Acesso em 18 setembro 2024.
- Marques, A. L., Ferreira, L. O. G., Cavalcante, P. P. M. M., Mendes, N. C. F., da Costa, C. S., dos Santos Silv rio, J. C., Neumann, C., Cruz, E. R. N., and de Souza Barbosa, E. (2022). Gest o dos custos na administra  o p blica federal: um estudo de caso a partir das entregas. *Brazilian Journal of Development*, 8(4):25744–25768.
- Microsoft (2024). Phi-2 model. Acesso em 18 setembro 2024.
- Pu, H., Yang, X., Li, J., and Guo, R. (2024). Autorepo: A general framework for multimodal llm-based automated construction reporting. *Expert Systems with Applications*, 255:124601.

- Rane, N., Choudhary, S., and Rane, J. (2024). A new era of automation in the construction industry: Implementing leading-edge generative artificial intelligence, such as chatgpt or bard. Available at SSRN.
- Serrano, A., CUNHA, R. D., FRANCO, V. R., Assis, M., and SOUZA, F. (2017). Dimensionamento da força de trabalho aplicado a uma organização do poder executivo federal. Anais do XX SEMEAD, São Paulo.
- Serrano, A. L. M., Ferreira, L. O. G., Mendes, N. C. F., Cavalcante, P. P. M. M., and Neumann, C. (2021). Modelos empíricos aplicados a análise da capacidade produtiva: Aplicações em cenários de incertezas empirical models applied to productive capacity analysis: Applications in uncertainty scenarios. Brazilian Journal of Development, 7(12):111940–111959.
- Silva, G. W. et al. (2024). Dimensionamento da força de trabalho: desafios e possibilidades para a gestão universitária.

Mitigação de Ataques DDoS em IoT: Comparação entre Aprendizado Federado e Modelos Centralizados

Mateus Flach Romani¹, João Kleber M. dos Santos¹, Gabriel Filipe de O. Cardoso¹,
Giovani B. O. Viana¹, Francisco L. de Caldas Filho¹, Fábio L. L. de Mendonça¹

mateusfromani@gmail.com; joaokleber200110@gmail.com; gafoco9@gmail.com;
giovani.braga.oliveira.viana@gmail.com; fabio.mendonca@redes.unb.br;
franciscolopescaledas@gmail.com.

¹ Universidade de Brasília, Brasília-DF, Brasil Departamento de Engenharia Elétrica, Faculdade de Tecnologia, Universidade de Brasília, Brasília-DF, 70910-900, Brasil.

Pages: 170-183

Resumo: A crescente adoção de dispositivos de Internet das Coisas (IoT) trouxe riscos como *malwares* e ataques de negação de serviço (DoS), demandando sistemas eficazes de detecção e prevenção de intrusões (IPS). Este estudo compara técnicas de aprendizado distribuído, que analisam padrões de tráfego malicioso sem centralização de dados, com um modelo centralizado baseado em Unidade Recorrente Fechada (GRU). A pesquisa avalia as vantagens e desvantagens de cada abordagem na mitigação de ameaças à segurança. Para fornecer escalabilidade e adaptabilidade em tempo real, a computação em nuvem é integrada à solução, que é testada com a botnet Mirai. O objetivo é contribuir para o desenvolvimento de mecanismos de segurança robustos para sistemas IoT, melhorando a experiência dos usuários e a integridade dos ambientes conectados.

Palavras-chave: Federated Learning; GRU; IOT; DDOS.

Mitigation of DDoS Attacks in IoT: A Comparison Between Federated Learning and Centralized Models

Abstract: The growing adoption of Internet of Things (IoT) devices has introduced risks such as malware and denial-of-service (DoS) attacks, demanding effective intrusion detection and prevention systems (IPS). This study compares distributed learning techniques, which analyze malicious traffic patterns without centralizing data, with a centralized model based on Gated Recurrent Units (GRU). The research evaluates the advantages and drawbacks of each approach in mitigating security threats. To ensure scalability and real-time adaptability, cloud computing is integrated into the solution, which is tested using the Mirai botnet. The goal is to contribute to the development of robust security mechanisms for IoT systems, enhancing user experience and the integrity of connected environments.

Keywords: Federated Learning; GRU; IOT; DDOS.

1. Introdução

A Internet das Coisas (IoT) é definida como o uso de protocolos de comunicação da Internet em equipamentos de coleta de dados, como sensores, scanners e dispositivos de identificação por radiofrequência, com o objetivo de controlá-los remotamente (Ali & Younis, 2023). Essa tecnologia é empregada em diversas aplicações, tanto na indústria (IIoT) quanto em ambientes domésticos.

O principal risco associado às redes IoT é a presença de malware, que explora vulnerabilidades em dispositivos para realizar atividades maliciosas, incluindo ataques de negação de serviço (DoS) e negação de serviço distribuída (DDoS) (Ali & Yousafzai, 2024). Tais ataques são especialmente preocupantes em sistemas IoT devido às limitações de processamento e consumo de energia dos dispositivos, o que torna a tarefa de projetar sistemas eficazes de Detecção e Prevenção de Intrusões (IDS) um grande desafio (Bhale et al, 2023). O aumento das preocupações com a segurança em ambientes IoT tem impulsionado avanços significativos em técnicas de detecção e prevenção de intrusões (Zhang et al, 2024).

Em um cenário de redes dinâmicas, é crucial que os sistemas de detecção e prevenção de intrusões sejam capazes de identificar cada tipo de tráfego e atualizar as regras de roteamento de forma imediata. As Redes Definidas por Software (SDN) oferecem uma arquitetura programável e controlada centralmente, adaptável e gerenciável independentemente da complexidade tecnológica da rede subjacente (Xu et al, 2022).

Os mesmos mecanismos de ataque e mitigação podem ser aplicados a todos os tipos de dispositivos conectados à Internet, não apenas a computadores convencionais. A adoção massiva de dispositivos IoT, especialmente em ambientes residenciais, amplifica a necessidade de uma linha de defesa robusta, uma vez que esses dispositivos estão majoritariamente na borda das redes, aumentando a superfície de possíveis ataques e comprometendo a integridade dos dados (Barret et al, 2023).

Diante dessa situação, diversos segmentos têm adotado sistemas de automação e tecnologias associadas à Internet das Coisas, como NFC (Near Field Communication) e RFID (Radio Frequency Identification). Essas tecnologias são integradas a dispositivos eletrônicos em vários segmentos da IoT, proporcionando funcionalidades avançadas de interação e controle (Mothlag, 2012).

A abordagem sugerida visa mitigar essas ameaças ao aproveitar as capacidades do aprendizado distribuído para analisar e identificar padrões de tráfego malicioso sem a necessidade de centralizar os dados. Isso não apenas melhora a eficiência na detecção de ataques, mas também preserva a privacidade dos dados dos usuários. Complementando essa estratégia, a computação em nuvem fornece a infraestrutura necessária para escalabilidade e processamento em larga escala, garantindo que a solução possa se adaptar rapidamente às mudanças dinâmicas das redes IoT e oferecer uma resposta ágil a ameaças emergentes. Tais soluções podem influenciar positivamente a experiência dos usuários em diversas aplicações, possibilitando a realização de pagamentos, o acesso a catracas e a obtenção e transmissão de conteúdo, mantendo mecanismos de segurança eficazes para a implementação dessas funcionalidades.

Este trabalho realiza um comparativo entre diferentes abordagens de aprendizado de máquina para a identificação de ataques de negação de serviço (DoS), permitindo avaliar as vantagens e desvantagens de cada modelo. A validação dos dois modelos foi realizada em um ambiente de rede contendo uma botnet, onde foi possível gerar um ataque de negação de serviço utilizando a botnet Mirai. Nesse cenário, observou-se o comportamento de um algoritmo distribuído baseado em GRU (*Gated Recurrent Unit*).

2. Revisão da Literatura

Esta seção visa discutir e apresentar conceitos acadêmicos e industriais levando em conta a segurança em *Internet of Things Devices* (IoT devices).

O advento de dispositivos de IoT frequentemente gera debates sobre sua confiabilidade e segurança. Em uma era em que a proteção de dados é crucial, inúmeros esforços e pesquisas têm sido direcionados para promover inovações nesse campo. Esta seção será dedicada para a revisão de soluções existentes, estabelecendo comparações com a nossa proposta.

Este estudo propõe o uso de técnicas de aprendizado de máquina baseado em GRU. Tomando como exemplo sua aplicação em (Zhang et al, 2021), é proposto e desenvolvido um framework para a detecção de ataques DDoS utilizando aprendizado distribuído. O uso dessa tecnologia para treinar modelos sem a necessidade de centralizar os dados dos dispositivos demonstrou um bom desempenho na detecção de diversos tipos de ataques, além de proporcionar eficiência em termos de tempo de treinamento e uso de memória nos dispositivos IoT.

Em (Ali & Younis, 2023) é proposto a integração de tecnologias *blockchain* com aprendizado distribuído para aprimorar a segurança e privacidade em redes IoT industriais. A implementação de sistemas inteligentes de Detecção e Prevenção de Intrusões (IDS/IPS) baseados em aprendizado distribuído, aliados à tecnologia *blockchain*, oferece uma solução robusta para monitorar o tráfego de rede IoT e detectar anomalias, ao mesmo tempo em que preserva a integridade e confidencialidade dos dados. Com a análise de diversas abordagens de IDS, foi evidenciada a complexidade e variedade de métodos que podem ser utilizados para lidar com ameaças de segurança.

No artigo (de Camara et al, 2023) é construído uma arquitetura de FL agrupado. Utilizando técnicas de agrupamento baseadas em “*Fingerprinting*” de modelos, esta abordagem não só preserva a privacidade dos dados, mas também melhora significativamente a convergência do modelo global. Os resultados experimentais, utilizando uma rede emulada com 78 dispositivos IoT e IIoT, demonstram a superioridade desta técnica em comparação com abordagens não agrupadas e métodos tradicionais de detecção de anomalias, especialmente em cenários com dispositivos de comportamentos diversos.

2.1. Interoperabilidade de Sistemas IoT

A definição concreta de interoperabilidade de sistema em redes IoT vai de encontro com o uso de serviços em nuvem, como é o caso do sistema da Sensorweb, a única tecnologia de comunicação aceitável é a internet baseada em IP. Isto inclui dispositivos que se

comuniquem diretamente pela internet, ou que se comuniquem através de agregadores *gateways* conectados à mesma (Raman, 2020).

Sistemas de software contemporâneos, tais como aqueles inerentes ao paradigma da Internet das Coisas (IoT), Indústria 4.0, Cidades Inteligentes, dentre outros, são complexos quando comparados a sistemas de software tradicionais. Essa complexidade provém da inclusão de novas preocupações e características relacionadas à rede, software, hardware, sensibilidade ao contexto, interface, interoperabilidade (Raman, 2020).

Considerando o contexto de redes IoT e sua significativa relevância em um cenário de múltiplos dispositivos conectados, neste trabalho será abordado um cenário de coordenação de dispositivos e soluções de segurança que possam ser integradas e aplicadas de forma uniforme em diferentes ambientes e plataformas.

2.2. Detecção de Intrusões e Anomalias

A detecção de intrusões e a identificação de anomalias em redes IoT são desafios críticos, especialmente em ambientes heterogêneos e de grande escala. Em (Radjaa et al, 2023), é proposta uma abordagem baseada em aprendizado profundo federado para a detecção de intrusões, visando aprimorar a privacidade e a segurança nas redes IoT. Destaca-se a comparação de desempenho e acurácia entre modelo distribuído e convencional.

Já o artigo (de Camara et al, 2023), apresenta uma arquitetura de aprendizado federado agrupado, que é particularmente útil para detectar anomalias em redes IoT heterogêneas de grande escala. A proposta enfatiza a necessidade de técnicas escaláveis e adaptáveis para enfrentar a complexidade e a diversidade dos dados gerados por dispositivos IoT.

Em (Liu et al, 2023) é apresentado um modelo de aprendizado federado assíncrono para a detecção de ataques DDoS de baixa taxa, ou seja, ataques que possuem maior variância no tempo e portanto, mais difíceis de detectar. Esse modelo é projetado para lidar com a complexidade e a variação dos ataques, melhorando a capacidade de resposta.

O trabalho citado é focado especificamente em ataques LDDos (*Low-rate Distributed Denial of Service*) enquanto nossa solução proposta aborda botnets e ataques de negação de serviço de forma mais ampla. Além disso o modelo de aprendizado de máquina utilizado foi o Bi-LSTM (*Bidirectional Long Short-Term Memory*), este modelo busca capturar dependências de longo prazo e focar em partes críticas da sequência, identificando padrões sutis no tráfego de rede que podem indicar um ataque em andamento, sendo bem eficaz para o contexto de ataques LDDos. Já esta solução utiliza o modelo 1D-CNN que oferece vantagens em termos de eficiência computacional e capacidade de capturar padrões locais relevantes para a detecção de ameaças.

Os autores (Yantian et al, 2024), propõem criar um framework distribuído para detectar tráfego malicioso no acesso de dispositivos IoT em redes 5G/6G. Coloca o mecanismo de detecção no lado dos dispositivos atacantes e não dos dispositivos atacados. Para tanto usa um modelo baseado em *transformer*. Alcançou uma acurácia de detecção média de 99.2% e um F1-score de 99.2% para todos os pontos de detecção no dataset N-BaIoT.

No proposto em (Nandanwar & Katarya, 2024), foi construído um IDS baseado em CNN-GRU para identificar intrusões dentro de cenários IoT industriais. Usam como base

o e N_BaIoT dataset. Atingem uma acurácia de teste de 99.75%, uma loss de 0.0063, precisão e recall score de 99.75% e 99.74% respectivamente.

Os resultados apresentados na solução acima evidenciam a eficácia superior em relação aos métodos tradicionais de aprendizado profundo. Em comparação com o modelo proposto neste estudo, que combina detecções de intrusões baseadas em rede e no host, promovendo uma resposta proativa a ataques zero-day, observa-se uma precisão média de 89,753% que, embora a taxa média de precisão seja inferior à do modelo em comparação, o IRS implementa um sistema que agrega para a privacidade dos dados e se expande, permitindo maior escalabilidade e adaptação a um ambiente real.

Já em (Wang et al, 2024), os autores propõem um modelo que consome menos recursos para detectar tráfego malicioso. Este método faz uso da redução da dimensionalidade dos dados altamente complexos da rede, com o uso de uma rede Siamese. Usa um modelo de larga escala de transformer como instrutor para um modelo menor e de pequena escala que atuará diretamente com o dispositivo. O modelo BT-TPF atinge mais de 99% de acurácia em ambos os datasets CIC-IDS2017 e TON_IoT.

O modelo apresentado acima traz uma abordagem inovadora que se distingue pela utilização de técnicas de destilação de conhecimento e uma estrutura leve, permitindo operações em dispositivos com recursos limitados, além de apresentar desempenho superior em termos de eficiência de parâmetros e tempo de inferência. O IRS, por sua vez, propõe uma solução robusta que permite a mitigação de ataques na origem. Ambas as abordagens refletem a crescente necessidade de soluções adaptáveis e eficientes em segurança de redes, com o BT-TPF enfatizando a leveza e a rapidez, enquanto o IRS se concentra em uma arquitetura mais integrada e escalável.

3. Metodologia

3.1. Arquitetura Proposta

O modelo proposto busca comparar diferentes abordagens para a detecção de ataques de negação de serviço distribuído (DDoS). O comparativo ocorre entre a detecção de anomalias utilizando um modelo de aprendizado de máquina federado, conforme apresentado em (de Caldas et al, 2023), e um modelo de detecção de anomalias centralizado, desenvolvido no presente experimento. Ambos os modelos foram avaliados utilizando o mesmo dataset e topologia, permitindo uma análise comparativa entre o desempenho da GRU (Gated Recurrent Unit) no modelo centralizado e da 1-CNN empregada no Federated Learning.

A arquitetura utilizada neste trabalho é composta por três componentes principais, operando dentro de uma mesma Virtual Local Area Network (VLAN), que os isola da rede externa. Os componentes incluem: o Network Intrusion Detection System (NIDS); um desktop responsável por executar aplicações como Apache Kafka, Suricata, Apache Spark e o servidor Flower; e o Mirai Controller, que é um notebook atuando como Comando e Controle (CNC) para os dispositivos infectados pelo malware. Além disso, há um conjunto de Raspberry Pi, divididos em dois grupos: um grupo atua como clientes legítimos da rede, enquanto o outro foi infectado pelo malware. Todos os dispositivos

são interligados por um switch de 24 portas, com a VLAN configurada nas portas de 1 a 10. Essa arquitetura foi inspirada na proposta dos autores em (de Caldas et al, 2023). O desenho dessa arquitetura pode ser visualizado na Figura 1.

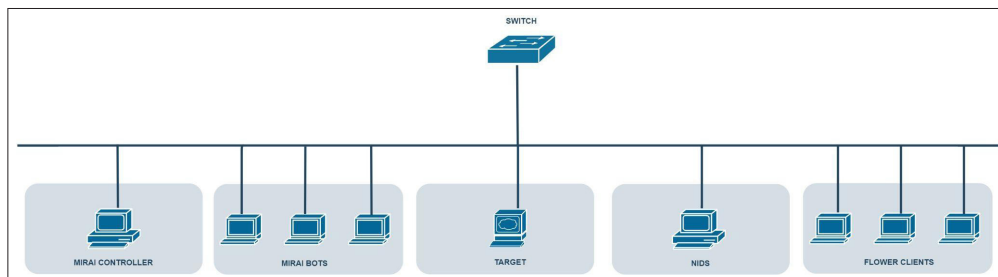


Figura 1 – Arquitetura Rede Botnet.

Uma adição ao que foi feito no artigo anterior de 2023, é o uso de um modelo centralizado, baseado em GRU, para realizar uma análise comparativa. O modelo centralizado utilizou os dados gerados da arquitetura acima para o seu treinamento.

A relação entre os componentes segue a seguinte estrutura. O Mirai Controller comanda a intrusão dos dispositivos infectados que compõem a *botnet*, na Figura 2 representados como Mirai Bots. Estes dispositivos terão como alvo um servidor web (Target). Os dados de tráfego gerados pelas tentativas de intrusão serão monitorados pelo NIDS, e utilizados para treinamento do modelo de aprendizado de máquina a fim de classificar o tipo de tráfego.

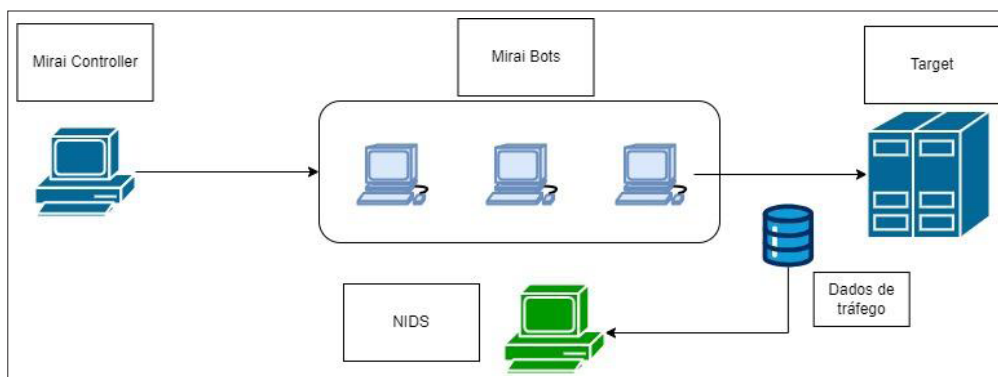


Figura 2 – Fluxograma dos Dados.

3.2. Network Intrusion Detection System (NIDS)

O NIDS é o elemento da arquitetura responsável pelas análises de dados, atuando como um IDS (*Intrusion Detection System*). O fluxograma de funcionamento básico consiste

no Suricata recebendo todo o tráfego da rede através do switch pelo *port mirroring*. Em seguida, os dados do armazenados em formato JSON, são recolhidos em tempo real, em tópicos do Apache Kafka, e enviados para consumidores como o Spark, que faz a manutenção desses dados.

3.3. Suricata

O Suricata é um software de código aberto que opera como um monitor de redes. Possui capacidade de detecção e prevenção de invasões, podendo ser configurado como um IDS (Intrusion Detection System) ou IPS (Intrusion Prevention System).

Neste trabalho, o Suricata foi utilizado como um coletor de logs de rede em tempo real por meio da interface de rede monitorada em seu arquivo de configuração.

3.4. Kafka

O Kafka é um software que permite que várias máquinas separadas, atuando como servidores ou clientes, se comuniquem através de protocolos TCP (Transmission Control Protocol) de alta performance. Ele funciona sobre um modelo de produtor-consumidor, na qual os produtores publicaram informação em um tópico e os consumidores se inscrevem nesse tópico para obtê-las em tempo real. Além disso, possui uma interface que permite a programação de *clusters*, com o uso de paralelismo.

3.5. Spark

O Spark é uma ferramenta de análise de dados, feita para atuar com um grande volume de dados. Os processamentos de dados do spark podem se dar em máquinas de um nó ou em *clusters*. Pode ser usado para consultas otimizadas (*queries*), *streaming*, criação de gráficos e aprendizado de máquina.

Um dos aspectos relevantes é o Spark Streaming, que permite manipular grande volume de dados em tempo real com baixa latência, por meio do *n-memory computing*, assim sendo uma ferramenta para análise de *Big Data*.

3.6. Modelo Centralizado

O modelo centralizado foi construído com base na arquitetura GRU (*Gated Recurrent Unit*). Os dados de treinamento utilizados foram colhidos durante o ataque de uma *botnet* real, cujos bots estavam infectados com o *malware* Mirai, e é o mesmo *dataset* utilizado em (de Caldas et al, 2023).

Os dados são pré-processados e separados em quinze colunas diferentes com os seguintes labels: 'flow_id' (identificador do fluxo de dados capturado pelo Wireshark), 'src_ip' (IP de origem), 'dest_ip' (IP de destino), 'src_port' (porta de origem), 'dest_port' (porta de destino), 'proto' (protocolo utilizado na comunicação), 'hour' (hora), 'minute' (minuto), 'seconds' (segundos), 'severity' (severidade),

'pkts_toserver' (quantidade de pacotes enviados para o servidor), 'pkts_toclient' (quantidade de pacotes enviados para o cliente), 'bytes_toserver' (quantidade de bytes enviados para o servidor), 'bytes_toclient' (quantidade de bytes enviados para o cliente),

‘class’ (tipo de pacote). Após isso, são entregues a um encoder que transforma o tipo dos itens do dataset em um tipo amigável para a rede neural, passam por um processo de padronização, onde os dados são transformados em valores com variância unitária, são divididos em dados de treinamento e dados de teste e finalmente são passados para a rede neural.

A rede neural é composta por duas camadas GRU, com 160 unidades, e intercaladas com camadas de *dropout* e *batch normalization*. Em seguida estão duas camadas dense, com 32 e 16 unidades respectivamente, intercaladas por camadas de *dropout*. Por fim, na saída da rede neural existe uma camada densa com 5 unidades. Foram testadas diferentes funções de ativação e inicializadores de *kernel*, mas foram utilizados as funções de ativação ReLu (*Rectified Linear Unit*) e SoftMax (Função Exponencial Normalizada), já para os inicializadores kernel foi utilizado o *Glorot Uniform*.

3.7. Mirai Controller

O malware Mirai foi um *malware* responsável pela execução de ataques de DDoS massivos entre 2016 e 2020. É originalmente desenvolvido com o objetivo de infectar dispositivos IoT (*Internet of Things*) e torná-los em *bots*, para tanto se valia da falta de segurança que é uma característica conhecida desse tipo de dispositivo para obter o controle do equipamento. A infecção testava credenciais padrões de acesso, sendo que quando o acesso era bem sucedido o dispositivo baixava um arquivo contendo um executável que transformava o dispositivo em um bot, que se conectava ao CNC (*Command and Control Server*) que indicava servidores aos quais os bots deveriam atacar.

3.8. Integração dos Elementos

Os elementos da arquitetura proposta para a detecção dos ataques, se integram por meio de uma *Virtual Local Area Network* (VLAN), criando um ambiente isolado da rede externa para análise segura. No centro da estrutura está o NIDS, o qual contém o Suricata, Apache Kafka, e o Apache Spark, formando um pipeline contínuo de análise dos dados.

4. Testes e Resultados

Este trabalho tem como objetivo realizar uma comparação entre modelos descentralizados, representados pelo *Federated Learning*, e modelos centralizados. Para isso, foram conduzidos testes com ataques de negação de serviço (DoS) em uma rede real, utilizando uma *botnet* real criada a partir do *malware* Mirai, conforme detalhado na seção Metodologia. A análise de identificação utilizando o modelo descentralizado foi realizada em (de Caldas et al, 2023), enquanto o presente artigo faz uma análise comparativa, utilizando os dados obtidos em (de Caldas et al, 2023),

4.1. Execução do Ataque

O Mirai funciona em um modelo *bot* e CNC (Servidor de Comando e Controle). O princípio básico de seu funcionamento é que o dispositivo infectado (*bot*), se comunica

com o servidor de controle e comando, pelos protocolos Telnet ou SSH, e aguarda para receber comandos sobre quais elementos da rede ele deve mirar e que tipo de ataque será executado. O Mirai permite que sejam executados ataques com os mais diversos tipos de protocolos, assim, o dispositivo mal intencionado pode escolher qual tipo de ataque executar e a duração do ataque.

Uma vez ativados, os *bots* enviam um enorme número de pacotes para o alvo, causando um enorme volume de tráfego na rede e, portanto, a indisponibilidade do serviço prestado pelo alvo.

4.2. Modelo Escolhido

O modelo *Gated Recurrent Unit* foi escolhido a fim de possibilitar a comparação dos resultados com aqueles apresentados em (Nandanwar & Katarya, 2024), o qual utilizou de dados de uma rede fictícia, e como uma melhoria proposta ao trabalho (de Caldas et al, 2023), que se utilizou de outros modelos para treinamento

4.3. Treinamento e Predição

Os dados obtidos a partir da rede *botnet* construída em (de Caldas et al, 2023), foram submetidos aos seguintes tratamentos a fim de dar maior acurácia no treinamento do modelo centralizado proposto.

Os dados foram divididos em quinze classificações diferentes (*labels*), conforme explicado anteriormente. Foi realizado tratamento a fim identificar valores inutilizáveis (null, none, entre outros). Os dados são submetidos a uma operação de *encoder*, que adequa os valores das colunas a tipos propícios ao treinamento. Em seguida foi feita padronização utilizando *standard scaler*, calculada como:

$$Z = \frac{x - u}{d} \quad (1)$$

Onde x é a amostra de dado, u é a média dos dados e d é o desvio padrão calculado sobre o conjunto de dados.

Finalmente, os dados são divididos em sub-conjuntos de treinamento e teste, com o uso da função *train test split*. O tamanho do conjunto de teste foi de 25\% do tamanho total do conjunto. Para acelerar o treinamento foi escolhido o otimizador Adam e taxa de aprendizagem (*learning rate*) de 0,0001.

A classificação dos dados ocorreu entre 5 classes diferentes: *Botnet*, *Bruteforce*, *DoS*, *Normal* e *Scan*, na qual *Botnet* seria o tráfego que ocorre entre os bots e o CNC, *Bruteforce* seria o tráfego relacionado com a tentativa de descoberta de credenciais dos dispositivos, *DoS* o tráfego relacionado com os ataques de negação de serviço, *Normal* se refere a tráfego que não está relacionado com ataques ou tentativas de infecção e *Scan* o tráfego ligado à tentativa de descobrir dispositivos vulneráveis na rede.

4.4. Resultados Obtidos

Os testes que serão descritos objetivam comparar o modelo centralizado de aprendizado de máquina presente neste experimento com o modelo de *federated learning* disposto em (de Caldas et al, 2023). Ademais, será feito comparativos com modelos centralizados a fim de estabelecer parâmetros para diagnósticos no presente experimento.

Partindo para a análise entre a captura de dados do NIDS feita em (de Caldas et al, 2023) e a classificação do ataque por meio do modelo centralizado de aprendizado de máquina. A Figura 3 a seguir descreve a classificação do ataque ocorrido no modelo centralizado relatado na subseção 3.1.

É possível classificar os dados obtidos em três (3) tipos. O primeiro parâmetro sendo o de acurácia determina o quão preciso foi a modelagem na detecção do tipo de anomalia, que foi descrito na subseção 3.8.

A Figura 4 a seguir, demonstra um gráfico da acurácia sobre as épocas (*epochs*).

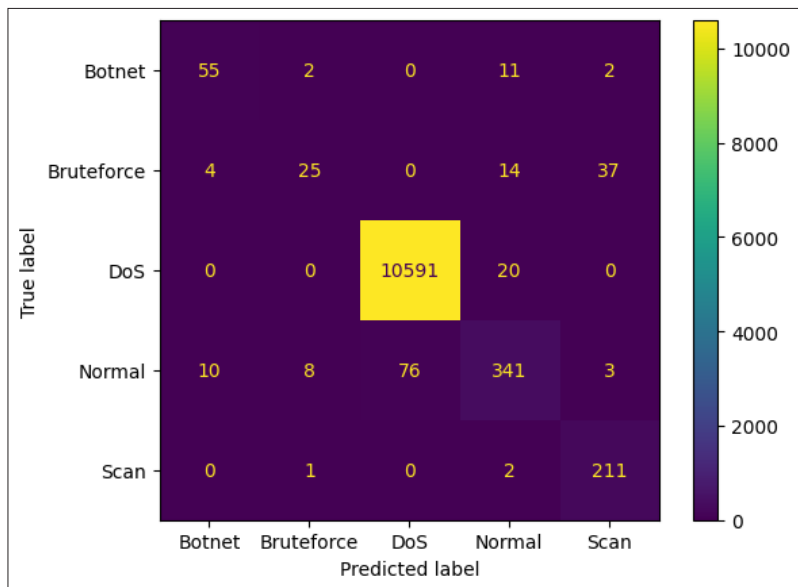


Figura 3 – Resultado da classificação.

Além disso, há o segundo parâmetro *cross-entropy*, que estipula quão difícil foi identificar um dado evento.

Por fim, foi determinado o último parâmetro que é o de perda do modelo proposto (verifica a performance do modelo), que foi explicado na subseção 3.1.

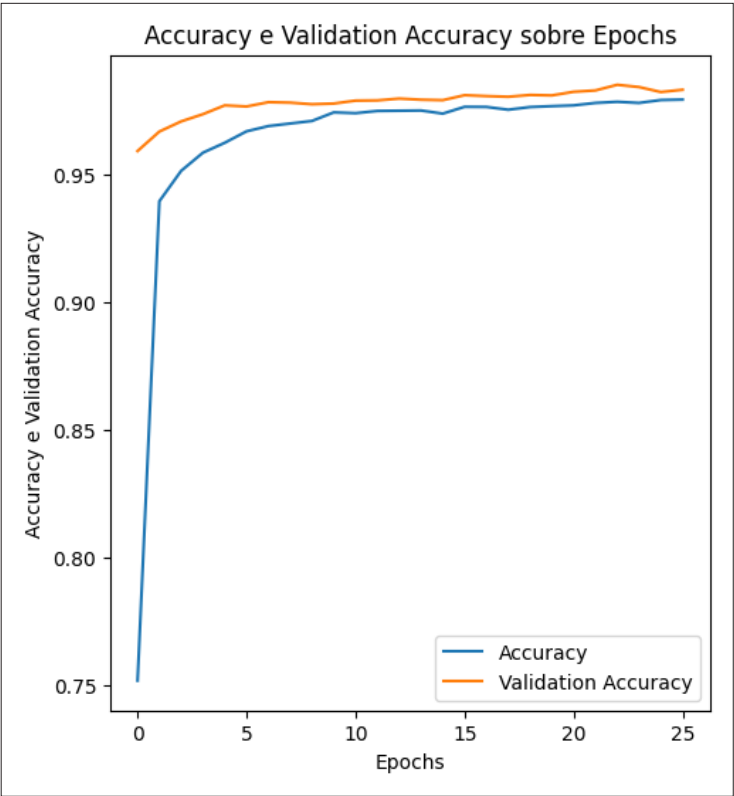


Figura 4 – Acurácia em épocas.

Todos os cenários obtidos resultam nos itens demonstrados na Tabela 1 a seguir.

Acurácia	Perda	Cross-entropy
0.9834	0.0333	0.0333

Tabela 1 – Tabela dos resultados obtidos.

O modelo federado em (de Caldas et al, 2023), modelo no qual será feito o comparativo, teve os dados dos testes demonstrados na Tabela 2 a seguir.

Acurácia	Perda
0.89753	0.26705

Tabela 2 – Tabela dos resultados *federated learning*.

Nota-se que o modelo centralizado, descrito na subseção 3.1, obteve menos perdas, logo um desempenho de performance quantitativamente maior, além da acurácia ser

relativamente maior, ou seja, uma maior precisão. O *benchmarking* com outros modelos centralizados também foi realizado vide (Gugueoth et al, 2023) em que os dados estão descritos abaixo na Tabela 3.

Acurácia
0.9746

Tabela 3 – Tabela dos resultados modelo centralizado.

Percebe-se que modelos centralizados possuem acurácias semelhantes devido ao funcionamento do modelo, sendo possível certificar que o modelo desenvolvido não destoou do proposto.

5. Conclusões

Este trabalho teve foco na comparação do modelo centralizado desenvolvido com o modelo federado proporcionado em (de Caldas et al, 2023). A fim de criar ambientes de detecção de anomalias distintas para verificar as diferenças entre parâmetros de acurácia e perdas.

Utilizando a abordagem descrita na seção 3, foi possível detectar e identificar ataques a rede, ademais foi integrado o modelo NIDS para geração de logs e o envio de dados em tempo real possibilitou converter os dados capturados na rede por arquivos pcap, para dados que podem ser processados pela GRU.

O modelo proposto se mostrou robusto em detectar ataques de negação de serviço com alta taxa de precisão. Como resultado, obteve-se acurácia de 98,34%.

Contudo, no aprendizado centralizado constatou-se uma barreira na elevada volumetria de dados em um único dispositivo central, fato que pode representar uma barreira na escalabilidade da solução.

Como objetivo futuro, buscamos acrescentar neste modelo a SDN (*Software Defined Network*), que vai proporcionar a mitigação dos ataques no dispositivo. E também, buscamos expandir o experimento com orquestradores em nuvem, facilitando o gerenciamento desses ambientes.

Referências

- Ali, M., Ahmed, S., & Younis, M. (2023). Ddos attacks in IoT networks: A comprehensive systematic literature review. *IEEE Access*, 11, 32,954–32,978.
- Ali, S., Li, Q., & Yousafzai, A. (2024). Blockchain and federated learning-based intrusion detection approaches for edge-enabled industrial IoT networks: A survey. *Ad Hoc Networks*, 152, 103320. Disponível online em 10 de outubro de 2023.
- Bhale, P., Roy Chowdhury, D., Biswas, S., & Nandi, S. (2023). Optimist: Lightweight and transparent IDS with optimum placement strategy to mitigate mixed-rate DDoS attacks in IoT networks. *IEEE Internet of Things Journal*, 10(10), 8357.

- Zhang, Z., Wang, X., & Li, Y. (2024). Federated deep learning for cyber security in the Internet of Things: Concepts, applications, and experimental analysis. *Journal of Cyber Security and Privacy*, 5(2), 145–162.
- Xu, L., Chen, H., & Liu, Q. (2022). SDNTruth: Innovative DDoS detection scheme for software-defined networks (SDN). *Computers & Security*, 113, 102418.
- Barrett, S., Boswell, B., & Dorai, G. (2023). Exploring the vulnerabilities of IoT devices: A comprehensive analysis of Mirai and Bashlite attack vectors. Augusta, USA: School of Computer and Cyber Sciences, Augusta University.
- Motlagh, H. (2012). Near field communication (NFC): A technical overview. University of Vaasa, Tech. Rep.
- Zhang, T., He, C., Ma, T., Gao, L., Ma, M., & Avestimehr, S. (2021). Federated learning for Internet of Things. *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems*, 413–419.
- de Câmara, X. S., Flores, J. L., Arellano, C., Urbieto, A., & Zurutuza, U. (2023). Clustered federated learning architecture for network anomaly detection in large scale heterogeneous IoT networks. *Computers Security*, 131, 103299.
- Hafizur Rahman, M. I. H. (2020). A comprehensive survey on semantic interoperability for Internet of Things: State-of-the-art and research challenges.
- Bellavista, P. et al. (2019). A survey on fog computing for the Internet of Things. *Pervasive and Mobile Computing*, 52, 71–99.
- Anawar, M. R., Wang, S., Zia, M. A., Jadoon, A. K., Akram, U., & Raza, S. (2018). Fog computing: An overview of big IoT data analytics. *Wireless Communications and Mobile Computing*, 2018, 7157192. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1155/2018/7157192>
- de Caldas Filho, F. L. et al. (2023). Botnet detection and mitigation model for IoT networks using federated learning. *Sensors*, 23(14), 6305.
- Gugueoth, V., Safavat, S., & Shetty, S. (2023). Security of Internet of Things (IoT) using federated learning and deep learning — Recent advancements, issues and prospects. *ICT Express*, 9(5), 941–960.
- Mustafa, Z., Amin, R., Aldabbas, H., & Ahmed, N. (2024). Intrusion detection systems for software-defined networks: A comprehensive study on machine learning-based techniques. *Cluster Computing*.
- Zhang, T., He, C., Ma, T.-S., Ma, M., & Avestimehr, S. (2021). Federated learning for Internet of Things. *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems*. Disponível em: <https://api.semanticscholar.org/CorpusID:235435775>
- Radjaa, B., Nabila, L., & Salameh, H. B. (2023). Federated deep learning-based intrusion detection approach for enhancing privacy in fog-IoT networks, 156–160.

- Liu, Z., Guo, C., Liu, D., & Yin, X. (2023). An asynchronous federated learning arbitration model for low-rate DDoS attack detection. *IEEE Access*, 11, 18,448–18,460.
- Luo, Y., Chen, X., Sun, H., Li, X., Ge, N., Feng, W., & Lu, J. (2024). Securing 5G/6G IoT using transformer and personalized federated learning: An access-side distributed malicious traffic detection framework. *IEEE Open Journal of the Communications Society*, 5, 1325–1339.
- Nandanwar, H., & Katarya, R. (2024). Deep learning enabled intrusion detection system for industrial IoT environment. *Expert Systems with Applications*, 249, 123808.
- Wang, Z., Li, J., Yang, S., Luo, X., Li, D., & Mahmoodi, S. (2024). A lightweight IoT intrusion detection model based on improved BERT-of-Theseus. *Expert Systems with Applications*, 238, 122045.

Critérios Editoriais

A RISTI (Revista Ibérica de Sistemas e Tecnologias de Informação) é um periódico científico, que foca a investigação e a aplicação prática inovadora no domínio dos sistemas e tecnologias de informação.

O Conselho Editorial da RISTI incentiva potenciais autores a submeterem artigos originais e inovadores para avaliação pelo Conselho Científico.

A submissão de artigos para publicação na RISTI deve realizar-se de acordo com as chamadas de artigos e as instruções e normas disponibilizadas no sítio Web da revista (<http://www.risti.xyz>).

Todos os artigos submetidos são avaliados por um conjunto de membros do Conselho Científico, não inferior a três elementos.

Em cada número da revista são publicados entre cinco a oito dos melhores artigos submetidos.

Criterios Editoriales

La RISTI (Revista Ibérica de Sistemas y Tecnologías de la Información) es un periódico científico, centrado en la investigación y en la aplicación práctica innovadora en el dominio de los sistemas y tecnologías de la información.

El Consejo Editorial de la RISTI incentiva autores potenciales a enviar sus artículos originales e innovadores para evaluación por el Consejo Científico.

Lo envío de artículos para publicación en la RISTI debe hacerse de conformidad con las llamadas de los artículos y las instrucciones y normas establecidas en el sitio Web de la revista (<http://www.risti.xyz>).

Todos los trabajos enviados son evaluados por un número de miembros del Consejo Científico de no menos de tres elementos.

En cada número de la revista se publican cinco a ocho de los mejores artículos enviados.



Revista Ibérica de Sistemas e Tecnologias de Informação
Iberian Journal of Information Systems and Technologies

©ITMA 2026 <http://www.risti.xyz>

