UNIVERSIDADE FEDERAL DO CEARÁ

CAMPUS QUIXADÁ

CURSO DE GRADUAÇÃO EM ENGENHARIA DE SOFTWARE

ROBSON DO AMARAL DIÓGENES

UTILIZANDO APRENDIZADO FEDERADO PARA DETECÇÃO DE INTRUSOS EM REDES IOT

QUIXADÁ

2026

ROBSON DO AMARAL DIÓGENES

UTILIZANDO APRENDIZADO FEDERADO PARA DETECÇÃO DE INTRUSOS EM REDES IOT

Projeto de Pesquisa apresentado ao Curso de Graduação em Engenharia de Software do Campus Quixadá da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de bacharel em Engenharia de Software.

Orientador: Prof. Me. Francisco Victor da Silva Pinheiro.
Coorientadora: Ma. Cleitianne Oliveira da Silva.

QUIXADÁ

2026

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

D622u Diógenes, Robson do Amaral.
Utilizando aprendizado federado para detecção de intrusos em redes IOT / Robson do Amaral Diógenes. – 2026.
59 f. : il. color.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Campus de Quixadá, Curso de Engenharia de Software, Quixadá, 2026.
Orientação: Prof. Me. Francisco Victor da Silva Pinheiro.
Coorientação: Profa. Ma. Cleitianne Oliveira da Silva.

1. Aprendizado federado. 2. Segurança em iot. 3. Internet das coisas. 4. Sistemas de detecção de intrusos. 5. Aprendizado de máquina. I. Título.

CDD 005.1

ROBSON DO AMARAL DIÓGENES

UTILIZANDO APRENDIZADO FEDERADO PARA DETECÇÃO DE INTRUSOS EM REDES IOT

Projeto de Pesquisa apresentado ao Curso de Graduação em Engenharia de Software do Campus Quixadá da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de bacharel em Engenharia de Software.

Aprovada em: 16 de Janeiro de 2026

BANCA EXAMINADORA

Prof. Me. Francisco Victor da Silva Pinheiro (Orientador)
Universidade Federal do Ceará (UFC)

Ma. Cleitianne Oliveira da Silva (Coorientadora)
Universidade Federal do Ceará (UFC)

Prof. Dr. Evilasio Costa Junior
Universidade Federal do Ceará (UFC)

Prof. Dr. João Marcelo Uchôa de Alencar
Universidade Federal do Ceará (UFC)

À minha família, por sua capacidade de acreditar em mim e investir em mim. Mãe, seu cuidado e dedicação foi que deram, em alguns momentos, a esperança para seguir. Pai, sua presença significou segurança e certeza de que não estou sozinho nessa caminhada.

AGRADECIMENTOS

Aos meus pais e familiares, pelo amor incondicional, pelo incentivo constante e por todo o suporte emocional que me permitiu focar nos meus estudos. Sem a base de vocês, essa conquista não seria possível.

À Universidade Federal do Ceará, Campus de Quixadá, por oferecer um ambiente de ensino de excelência e a infraestrutura necessária para a minha formação acadêmica e profissional.

Ao meu orientador, Prof. Me. Francisco Victor da Silva Pinheiro, pela orientação segura, pela paciência durante todo o processo de desenvolvimento deste trabalho e por todo o conhecimento compartilhado, fundamentais para a realização desta pesquisa.

À minha coorientadora, Ma. Cleitianne Oliveira da Silva, pelo apoio técnico, pelas discussões enriquecedoras e pela parceria que contribuiu significativamente para a qualidade deste trabalho.

Agradeço imensamente aos membros da banca examinadora, Prof. Dr. João Marcelo Uchôa de Alencar (UFC Quixadá) e Prof. Dr. Evilásio Costa Junior (UFC Sobral), pela disponibilidade, pela leitura atenta do meu trabalho e pelas valiosas contribuições e sugestões apresentadas durante a defesa, que certamente engrandeceram o resultado final.

Aos meus amigos e colegas de curso, pelo companheirismo, pela troca de experiências e pelos momentos de descontração que tornaram essa jornada mais leve.

"Segurança é um processo, não um produto."

(SCHNEIER, 2000, p. 257.)

RESUMO

A ampla adoção da Internet das Coisas (IoT) tem ampliado a superfície de ataque das redes computacionais, tornando a segurança um desafio central desse ecossistema. Sistemas de detecção de intrusos baseados em aprendizado de máquina apresentam bons resultados, porém abordagens centralizadas exigem a transferência de dados sensíveis, gerando preocupações relacionadas à privacidade e à conformidade com legislações como a LGPD. Nesse contexto, o Aprendizado Federado surge como uma alternativa ao permitir o treinamento colaborativo de modelos sem o compartilhamento de dados brutos. Este trabalho avalia a viabilidade e o desempenho de um Sistema de Detecção de Intrusos baseado em Aprendizado Federado aplicado a redes IoT, considerando cenários realistas com dados severamente desbalanceados e não independentes e identicamente distribuídos (Non-IID). Os experimentos utilizam o conjunto de dados CICIoT2023 e uma arquitetura de rede neural profunda otimizada para ambientes federados, implementada com o framework Flower e a estratégia FedProx. Os resultados indicam que o modelo federado alcançou acurácia global de 82,80% e F1-Score macro de 81,20%, demonstrando desempenho competitivo aliado à preservação da privacidade e à viabilidade de implantação em dispositivos de borda com recursos limitados.

Palavras-chave: internet das coisas; sistemas de detecção de intrusos; aprendizado federado; segurança em iot; aprendizado de máquina.

ABSTRACT

The widespread adoption of the Internet of Things (IoT) has expanded the attack surface of computer networks, making security a central challenge in this ecosystem. Machine learning–based intrusion detection systems have shown promising results; however, centralized approaches require the transfer of sensitive data, raising concerns related to privacy and compliance with regulations such as the Brazilian General Data Protection Law (LGPD). In this context, Federated Learning emerges as an alternative by enabling collaborative model training without sharing raw data. This work evaluates the feasibility and performance of a Federated Learning–based Intrusion Detection System applied to IoT networks, considering realistic scenarios with severely imbalanced and non-independent and identically distributed (Non-IID) data. Experiments were conducted using the CICIoT2023 dataset and a deep neural network architecture optimized for federated environments, implemented with the Flower framework and the FedProx aggregation strategy. The results indicate that the federated model achieved a global accuracy of 82.80% and a macro F1-score of 81.20%, demonstrating competitive performance while preserving data privacy and ensuring feasibility for deployment on resource-constrained edge devices.

Keywords: internet of things; intrusion detection systems; federated learning; iot security; machine learning.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE QUADROS

SUMÁRIO

1 INTRODUÇÃO

Nas últimas décadas, a sociedade têm passado por uma transformação impulsionada pelo avanço da tecnologia da informação. No entanto, junto com os benefícios da digitalização, surgem também importantes desafios, especialmente no que diz respeito à segurança da informação, à privacidade e à integridade dos dados. Segundo Khando *et al.* (2021), tanto os cibercriminosos quanto às violações de dados aumentaram drasticamente nos últimos anos. Diante desse cenário, proteger sistemas e dispositivos conectados deixou de ser apenas uma preocupação técnica e passou a ser uma necessidade estratégica para a sociedade como um todo.

A Internet das Coisas (*Internet of Things* - IoT) refere-se a uma rede de objetos físicos dotados de capacidades sensoriais e computacionais, que se comunicam entre si e interagem com o mundo real (Tournier *et al.*, 2021). Impulsionada pelo avanço tecnológico, a IoT surgiu como parte inevitável da vida humana, incluindo aprendizado online, casas, carros, redes, cidades, agricultura e saúde (Shukla *et al.*, 2023). A crescente relevância da IoT vem abrindo discussões sobre a segurança dessas redes. Aproximadamente 30 anos após o surgimento da IoT, a sociedade enfrenta desafios significativos em relação à segurança da IoT, devido à interconectividade e ao uso generalizado de dispositivos IoT. Os ataques cibernéticos têm impactos generalizados em diversas partes interessadas (Schiller *et al.*, 2022).

Dentre as inúmeras dificuldades relacionadas ao ambiente IoT, destaca-se o desafio em encontrar e mitigar ataques cibernéticos, como ataques distribuídos de negação de serviços (DDoS) e acessos não autorizados. Essas falhas de segurança podem comprometer não apenas os dispositivos individuais, mas todo o ecossistema conectado, afetando sua disponibilidade e integridade. Conforme discutido em Kaur *et al.* (2023), diferentes padrões de comunicação e protocolo, padrões de segurança fracos e a dificuldade na distribuição de atualizações agravam as ameaças à segurança cibernética, especialmente em aplicações críticas que empregam o uso da IoT.

Além disso, os dispositivos IoT geralmente possuem baixo poder de processamento, memória restrita e consumo energético reduzido. Essas limitações dificultam a execução de algoritmos complexos diretamente nos dispositivos, o que reforça a necessidade de soluções leves e distribuídas. Essa fragilidade facilita a criação de *botnets*, amplamente utilizadas em ataques de DDoS em larga escala (Saheed *et al.*, 2022).

Uma das técnicas utilizadas para combater as vulnerabilidades encontradas em redes IoT, é a utilização de aprendizado de máquina (*Machine Learning* - ML) para encontrar anomalias

e ameaças em ambientes de Internet das Coisas. Como destacado por Tekin *et al.* (2023), a detecção de intrusão baseada em ML é uma solução promissora para lidar com as preocupações de segurança e privacidade das redes IoT. No entanto, a maioria das soluções baseadas em ML ainda adota um modelo centralizado, no qual os dados coletados pelos dispositivos são enviados para um servidor central que realiza a análise. Embora seja uma solução eficaz, levanta questões relacionadas à privacidade, uma vez que exige a transferência contínua de dados potencialmente sensíveis, o que implica maior exposição a violações e riscos associados.

Neste contexto, o aprendizado federado (*Federal Learning* - FL) surge como uma alternativa para resolver o problema de segurança em redes IoT sem comprometer a privacidade dos dados presentes na rede. O FL é uma técnica segura e distribuída de ML que executa algoritmos de ML de forma cooperativa em múltiplos dispositivos de borda (*Edge Devices*) ou servidores distribuídos, sem que os dados privados saiam do ambiente local. Wen *et al.* (2023). enfatizam que o aprendizado federado permite que modelos de aprendizado sejam treinados de maneira colaborativa, dessa forma preservando a privacidade dos dados, já que as informações sensíveis nunca são compartilhadas para um servidor central. Em vez disso, apenas atualizações do modelo (como gradientes ou parâmetros) são transmitidos.

A segurança em redes IoT e as limitações do aprendizado de máquina centralizado em relação à privacidade são temas importantes e atuais. Diante disso, o uso do aprendizado federado para detectar intrusos em redes IoT se mostra uma abordagem promissora. O FL é uma nova abordagem que permitiu o compartilhamento de conhecimento com manutenção da privacidade e redução de custos (Friha *et al.*, 2022). Embora já existam pesquisas sobre detecção de ataques em redes IoT, poucas dessas pesquisas exploram soluções baseadas em modelos distribuídos de aprendizado. Diante disso, este trabalho propõe um modelo de detecção de intrusos que utiliza técnicas de Aprendizado Federado. A proposta integra uma Rede Neural Profunda (Deep Neural Network - DNN) a algoritmos de agregação robustos, visando demonstrar como essa abordagem distribuída pode oferecer uma alternativa eficiente e segura aos modelos centralizados tradicionais.

Ao mover a detecção de ML para a borda usando FL, surge um desafio fundamental, a natureza estatística dos dados em redes IoT reais, tecnicamente classificada como Non-IID (Non-Independent and Identically Distributed). Isso significa que os dados não seguem uma distribuição uniforme entre os participantes, cada dispositivo observa um subconjunto específico e enviesado do tráfego, o que dificulta a convergência de um modelo global único. Conforme

demonstrado por Li *et al.* (2022), o desempenho de algoritmos FL é drasticamente afetado pelo tipo de desbalanceamento dos dados, seja ele na distribuição dos rótulos ou nas características. A própria análise deste trabalho confirma que o conjunto de dados CICIoT2023 do trabalho Neto *et al.* (2023). sofre de um severo desbalanceamento de classes, que é um dos cenários Non-IID mais desafiadores. Embora estudos como o de Li *et al.* (2022) comparem algoritmos FL em conjuntos de dados genéricos como MNIST e CIFAR-10, falta uma investigação de seu desempenho no domínio específico de detecção de intrusos em larga escala.

Nesse contexto, um trabalho relevante é o sistema MARIA, desenvolvido por Silva *et al.* (2025). Embora essa pesquisa utilize uma base de dados distinta, composta pela fusão de dois conjuntos de dados, ela estabelece um importante referencial para o monitoramento de ataques em redes 5G e IoT. A pesquisa se destaca por usar um modelo de aprendizado supervisionado centralizado (*CatBoost*) para obter alta performance (90,72% de acurácia). Apesar da eficácia, o modelo centralizado gera limitações de privacidade, um problema que o FL se propõe a resolver. O presente trabalho, portanto, busca preencher essa lacuna, investigando se uma abordagem de aprendizado federado, operando em um ambiente Non-IID desafiador, consegue atingir uma eficácia de detecção comparável à forte linha de base centralizada estabelecida pelo MARIA, mas com os benefícios de privacidade inerentes.

Neste trabalho, propõe-se uma investigação experimental focada na comparação de desempenho entre abordagens centralizadas e federadas. Diferente da proposta original do MARIA, esta pesquisa introduz uma arquitetura específica de Aprendizado Federado baseada em DNN e no algoritmo de agregação FedProx para a análise do conjunto de dados CICIoT2023. O objetivo é avaliar a eficácia dessa implementação, simulada através do *framework Flower*, comparando-a com os resultados do modelo centralizado (*CatBoost*) do trabalho MARIA. A proposta busca, assim, quantificar a eficácia de detecção de intrusos (usando métricas como acurácia, *F1-Score*, precisão e *recall*) de soluções FL em um cenário de dados de segurança complexo e desbalanceado, validando sua viabilidade como uma alternativa que preserva a privacidade.

1.1 Objetivos

1.1.1 Objetivo Geral

Avaliar a viabilidade e o desempenho de modelos de aprendizado federado para detecção de intrusos em redes IoT, considerando cenários Non-IID e comparando-os com abordagens centralizadas sob restrições de privacidade.

1.1.2 Objetivos específicos

- Realizar uma análise e requisitos e desafios em redes IoT e aprendizado federado;
- Estudar os principais ataques e vulnerabilidades em redes IoT;
- Captar conjuntos de dados de redes IoT para o treinamento dos algoritmos;
- Realizar treinamento dos modelos de aprendizado de máquina; e
- Realizar análise de desempenho dos modelos;

2 FUNDAMENTAÇÃO TEÓRICA

Neste Capítulo, serão apresentados os conceitos relevantes e necessários para compreender este trabalho. Na Seção 2.1, serão abordados os conceitos fundamentais da Internet das Coisas (IoT), incluindo sua arquitetura, protocolos e os desafios de segurança inerentes a esse ecossistema. Em seguida, a Seção 2.3 explorará os Sistemas de Detecção de Intrusos (IDS), detalhando suas diferentes tipologias e o papel crucial que desempenham na proteção de redes, a Seção 2.4 e a Seção 2.5 fornecerão uma visão sobre as técnicas de inteligência utilizadas, abordando os conceitos de Aprendizado de Máquina e, com destaque, a abordagem de Aprendizado Federado, explicando suas funcionalidades e sua relevância para a segurança em ambientes distribuídos. Por fim a Seção 2.6 abordará os conceitos relacionado ao *framework flower*.

2.1 Internet das Coisas

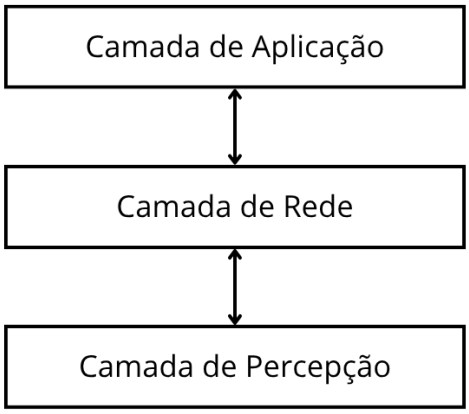
A Internet das Coisas refere-se a uma rede de objetos físicos (coisas) capazes de se comunicar (entre si e com entidades externas) e de sentir e interagir com o mundo real, essas coisas têm diferentes capacidades computacionais e sensoriais, proporcionando interações complexas com seu ambiente ou usuários (Tournier *et al.*, 2021). A IoT visa conectar diferentes coisas para a Internet, formando um ambiente conectado onde a detecção de dados, a computação e as comunicações são realizadas automaticamente sem envolvimento humano (Nguyen *et al.*, 2022).

A IoT está desempenhando um papel significativo na transformação de fábricas tradicionais em fábricas inteligentes na Indústria 4.0, utilizando uma rede de dispositivos, sensores e softwares interconectados para monitorar e otimizar o processo de produção (Soori *et al.*, 2023).

De acordo com Domínguez-Bolaño *et al.* (2022), atualmente ainda não existe um padrão ou tecnologia predominantemente utilizado por plataformas e dispositivos IoT, ao invés disso, diferentes dispositivos e plataformas consideram diversos padrões e tecnologias. Alguns padrões arquiteturais para sistemas IoT foram propostos na literatura para sanar esse problema, uma das mais simples é o modelo de três camadas reproduzido na Figura 1. Esse modelo é composto por três camadas principais: (i) camada de aplicação, que disponibiliza serviços e interfaces aos usuários finais, como automação residencial, monitoramento de saúde e gestão de

recursos urbanos; (ii) a camada de rede, que viabiliza a transmissão das informações através de diferentes protocolos de comunicação; e (iii) a camada de percepção, responsável pela coleta de dados por meio de sensores e atuadores.

Figura 1 – Arquitetura de três camadas



Fonte: Adaptada de Domínguez-Bolaño *et al.* (2022).

Apesar do crescimento desta tecnologia, a mesma ainda enfrenta alguns desafios em sua implementação como tecnologia global. Segundo Sobin (2020), há muitas questões em aberto que precisam ser abordadas para uma implementação eficiente da IoT, algumas delas são: falta de arquiteturas e protocolos padronizados, segurança e privacidade. A grande adoção juntamente com a falta de padronização de arquiteturas e protocolos, resultam em vulnerabilidades no âmbito da segurança e privacidade em redes IoT.

2.2 Ameaças e Desafios de Segurança em Redes IoT

Considerando a taxa de aceitação da IoT, o número de dispositivos conectados à IoT aumenta a cada dia, no entanto, existem diversos desafios de segurança enfrentados nas camadas da IoT (Litoussi *et al.*, 2020). Segundo Shaukat *et al.* (2021) um ataque de DDoS em 2016 foi realizado por uma botnet que consistia em muitos dispositivos IoT, como câmeras IP, impressoras e monitores de bebê.

Segundo Booij *et al.* (2022), ataques em redes IoT são altamente diversos e incluem varredura de portas (*scanning*), injeção de comandos, ataques DDoS e DoS, *ransomware*, *backdoors*, *cross-site scripting* (XSS), roubo de senhas e ataques do tipo *man-in-the-middle* (MITM). Muitas dessas ameaças exploram falhas comuns nos dispositivos, como a execução de software desatualizado e a ausência de criptografia nos dados trafegados, uma prática que, segundo um relatório de 2020, afeta 98% de todo o tráfego IoT (Booij *et al.*, 2022).

Ataques em redes IoT, especialmente em ambientes com dispositivos de baixa potência, causam impactos significativos na eficiência e confiabilidade da comunicação. Segundo Sharma e Verma (2021), ataques a redes IoT afetam diretamente métricas cruciais, como eficiência energética, escalabilidade da rede e a construção do grafo acíclico orientado ao destino, essencial para o roteamento. Esses impactos não apenas degradam o desempenho da rede, mas também comprometem a confiabilidade de aplicações críticas que dependem de comunicação em tempo real, como saúde, transporte e monitoramento ambiental.

2.3 Sistemas de Detecção de Intrusões (*Intrusion Detection Systems* - IDS)

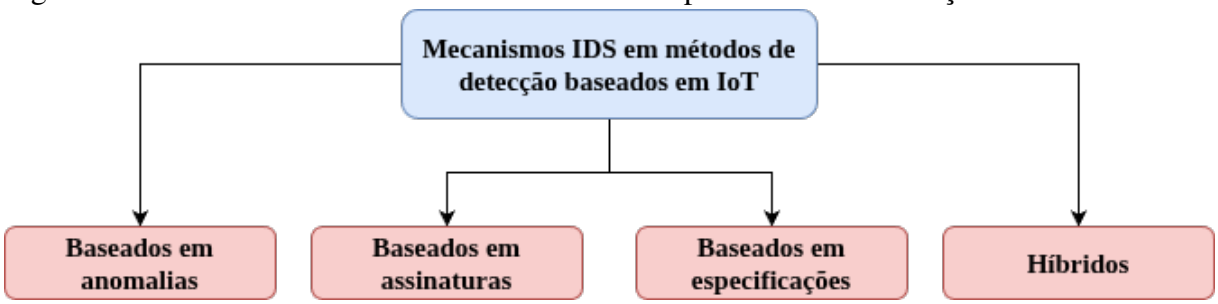
Um IDS é um sistema de software ou hardware que identifica ações maliciosas em sistemas de computador para permitir que a segurança do sistema seja mantida (Liao *et al.*, 2013). De acordo com Elrawy *et al.* (2018), o funcionamento de um IDS pode ser dividido em três estágios principais: o de monitoramento, que utiliza sensores baseados em rede ou host para coletar dados; o de análise, responsável por aplicar métodos de extração de características ou identificação de padrões; e o estágio de detecção, no qual são aplicadas técnicas de detecção de anomalias ou de uso indevido.

A IoT possui diferentes tipos de IDSs, como híbridos, baseados em anomalias, baseados em assinaturas e baseados em especificações (Heidari; Jamali, 2022), conforme ilustrado na Figura 2. Os IDSs baseados em assinaturas detectam ameaças por meio da comparação de padrões conhecidos de ataques com o tráfego atual da rede. Os IDSs baseados em anomalias monitoram o comportamento da rede e disparam alertas quando identificam desvios significativos do padrão considerado normal. Enquanto que os IDSs baseados em especificações definem regras de comportamento aceitável para os dispositivos IoT e alertam sempre que essas regras são violadas. Por fim, os IDSs híbridos combinam duas ou mais dessas abordagens para obter maior precisão na detecção.

Embora os IDS tradicionais desempenhem um papel importante na segurança de redes computacionais, sua aplicação direta em ambientes IoT enfrenta limitações significativas. Isso ocorre devido à natureza dinâmica, heterogênea e frequentemente virtualizada desses ambientes. Segundo Heidari e Jamali (2022), os sistemas IDS precisam ser adaptáveis e dinâmicos para lidar com a constante adição e remoção de nós em redes IoT, além de serem capazes de classificar ataques com base na integridade, disponibilidade e confidencialidade dos dados.

A implementação de IDSs em redes IoT impõe desafios em comparação com redes

Figura 2 – A taxonomia dos mecanismos IDS na IoT por método de detecção



Fonte: Adaptada de Heidari e Jamali (2022).

tradicionais. As limitações de hardware dos dispositivos dificultam o uso de técnicas de detecção mais elaboradas, exigindo soluções leves e eficientes. Além disso, a dinâmica da IoT, com dispositivos sendo frequentemente adicionados ou removidos, compromete a eficácia de regras fixas de detecção. Também é necessário considerar a privacidade dos dados monitorados, especialmente em aplicações sensíveis como saúde ou automação residencial. Essas particularidades reforçam a necessidade de novas abordagens de detecção, como as baseadas em aprendizado de máquina distribuído, tema abordado nas próximas seções deste trabalho.

2.4 Aprendizado de Máquina

A inteligência artificial (IA), e em particular o aprendizado de máquina, experimentaram um crescimento nos últimos anos, impulsionando a análise de dados e o desenvolvimento de aplicações inteligentes (Sarker *et al.*, 2021a). Essencialmente, o aprendizado de máquina concede aos sistemas a capacidade de aprender e aprimorar seu desempenho a partir da experiência, de forma autônoma e sem a necessidade de programação explícita para cada tarefa (Sarker *et al.*, 2021b). Essa versatilidade tem permitido a aplicação bem-sucedida de suas técnicas nos mais diversos campos, que vão desde o reconhecimento de padrões, visão computacional e engenharia espacial, até finanças, entretenimento e aplicações nas áreas biomédica e médica (Naqa; Murphy, 2015).

2.4.1 Aprendizado Supervisionado

A aprendizagem supervisionada (*Supervised Learning* - SL) é uma abordagem fundamental da IA, na qual um modelo é treinado a partir de um conjunto de dados previamente rotulados para realizar previsões sobre dados novos e não vistos. Entre os algoritmos mais comuns desta categoria estão as árvores de decisão, máquinas de vetores de suporte (SVM),

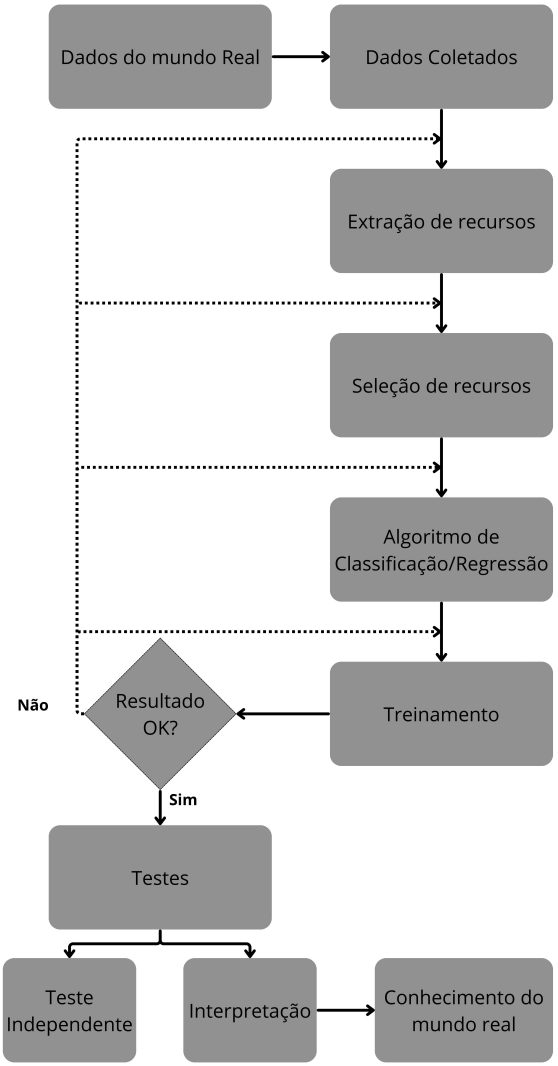
regressão linear e logística, e o k-vizinho mais próximo (k-NN) (Pantanowitz *et al.*, 2025).

O objetivo matemático das técnicas de SL consiste em aprender uma função capaz de mapear um conjunto de variáveis de entrada para uma variável de saída. Formalmente, busca-se prever uma variável aleatória alvo $Y \in \mathscr{Y}$, com base em um conjunto de variáveis aleatórias explicativas (ou recursos) denotadas por $X \in \mathscr{X}$. Os espaços $\mathscr{Y}$ e $\mathscr{X}$ dependem do problema, mas frequentemente podem ser representados por $\mathbb{R}$ e $\mathbb{R}^d$, respectivamente (Crisci *et al.*, 2012). A tarefa é denominada classificação quando a variável alvo é categórica (por exemplo, "sim"ou "não", "doente"ou "saudável") e regressão quando seu valor é contínuo e real (como prever um preço, temperatura ou idade) (Fabris *et al.*, 2017).

O processo para desenvolver uma solução de aprendizado supervisionado segue um fluxo de trabalho estruturado, conforme ilustrado na Figura 3. O ciclo se inicia com a coleta de dados do mundo real, que são então organizados em um conjunto de dados coletados para o projeto. A etapa seguinte, de pré-processamento, envolve a extração e formatação de recursos, onde informações relevantes são geradas, e a Seleção de Recursos, na qual as variáveis mais preditivas são escolhidas para compor o modelo.

Com os dados devidamente preparados, um Algoritmo de Classificação/Regressão é selecionado e passa pelo Treinamento, fase em que aprende os padrões existentes nos dados. O desempenho do modelo é então validado. Se o resultado não for satisfatório (o caminho "Não"do fluxograma), o processo retorna a etapas anteriores para ajustes. Uma vez que um resultado aceitável é alcançado (caminho "Sim"), o modelo é submetido aos testes finais com um conjunto de dados independente para uma avaliação imparcial de sua capacidade de generalização, No contexto deste trabalho, o critério para seguir o caminho 'Sim' é a convergência do modelo, definida como o momento em que a curva de acurácia estabiliza e deixa de apresentar crescimento significativo. Uma vez alcançado esse estado, o modelo é submetido aos testes finais com um conjunto de dados independente para uma avaliação imparcial de sua capacidade de generalização. Por fim, a interpretação desses resultados permite gerar conhecimento do mundo real, cumprindo o objetivo final do processo.

Figura 3 – Visão geral do processo de aprendizagem supervisionada.



Fonte: Adaptada de Kuncheva (2004).

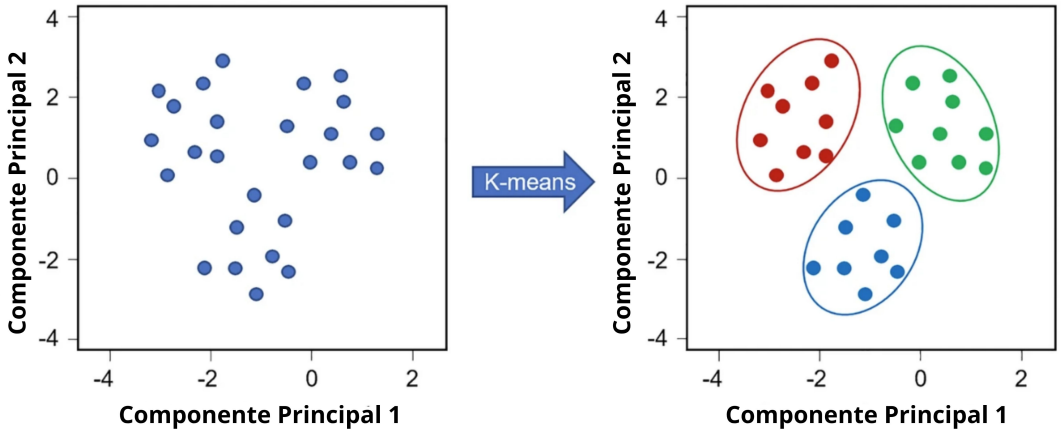
2.4.2 Aprendizado Não Supervisionado

A aprendizagem não supervisionada utiliza conjuntos de dados não rotulados para descobrir padrões e estruturas ocultas, muitas vezes de forma não intuitiva. Exemplos dessa abordagem são algoritmos de agrupamento (*clustering*), redução de dimensionalidade e detecção de anomalias (Pantanowitz *et al.*, 2025). A principal diferença em relação à aprendizagem supervisionada é a ausência de uma variável-alvo (rótulo) para guiar o treinamento.

Uma aplicação comum e poderosa de técnicas não supervisionadas é a combinação da redução de dimensionalidade com o agrupamento para a exploração de dados. A Figura 4 ilustra exatamente esse processo. Primeiramente, os dados são projetados em um espaço de menor dimensão através da Análise de Componentes Principais (PCA), onde os eixos representam os

"Componentes Principais". Em seguida, um algoritmo de agrupamento, neste caso o K-means, é aplicado para organizar os dados em grupos distintos com base em suas características de similaridade, revelando assim uma estrutura inerente que não era óbvia inicialmente.

Figura 4 – Ilustração do algoritmo de agrupamento K-means.



Fonte: Adaptada de Eckhardt *et al.* (2023).

Além da exploração de dados, a aprendizagem não supervisionada também é frequentemente usada em conjunto com a aprendizagem supervisionada em ambientes de aprendizagem semissupervisionada. Nesses cenários, ela serve para pré-processar os dados, auxiliando na criação de uma boa representação de recursos (*features*) e na descoberta de padrões em dados não rotulados, o que pode, por fim, melhorar o desempenho de um modelo supervisionado (Usama *et al.*, 2019).

2.4.3 Métricas de Avaliação para Modelos de Classificação

A avaliação de modelos de ML é uma etapa fundamental para descartar métodos com baixo desempenho e otimizar aqueles que se mostram promissores (Rainio *et al.*, 2024). Para tarefas de classificação, a performance é frequentemente avaliada por meio de métricas derivadas da matriz de confusão, que compara as classes previstas pelo modelo com as classes reais (Naidu *et al.*, 2023; Rainio *et al.*, 2024).

2.4.3.1 A Matriz de Confusão

Para uma tarefa de classificação binária, cada predição pode ser categorizada em uma de quatro designações possíveis (Rainio *et al.*, 2024). Essas categorias são a base para a maioria das métricas de avaliação e são organizadas na matriz de confusão, conforme visto no

Quadro 1.

Quadro 1 – Definições dos Componentes da Matriz de Confusão

Termo	Definição
Verdadeiro Positivo (TP)	Uma instância positiva que foi corretamente classificada como positiva (Naidu *et al.*, 2023; Rainio *et al.*, 2024).
Verdadeiro Negativo (TN)	Uma instância negativa que foi corretamente classificada como negativa (Naidu *et al.*, 2023; Rainio *et al.*, 2024).
Falso Positivo (FP)	Uma instância negativa que foi incorretamente classificada como positiva (Rainio *et al.*, 2024). Este cenário também é conhecido como Erro Tipo I (Naidu *et al.*, 2023).
Falso Negativo (FN)	Uma instância positiva que foi incorretamente classificada como negativa (Rainio *et al.*, 2024). Este cenário é conhecido como Erro Tipo II (Naidu *et al.*, 2023).

Fonte: Elaborado pelo autor.

A soma de TP e TN representa as instâncias corretamente classificadas, enquanto a soma de FP e FN constitui as instâncias classificadas incorretamente (Naidu *et al.*, 2023).

2.4.3.2 Acurácia (Accuracy)

A Acurácia é uma das métricas mais comuns e representa a porcentagem de instâncias classificadas corretamente em relação ao total de instâncias (Rainio *et al.*, 2024). Embora seja intuitiva, pode ser uma métrica enganosa em conjuntos de dados desbalanceados, onde o número de instâncias negativas e positivas é muito diferente (Rainio *et al.*, 2024). Sua fórmula é dada por:

$$\text{Acurácia} = \frac{TP + TN}{TP + TN + FP + FN} \tag{2.1}$$

O valor ideal para a Acurácia é 1.0, enquanto o pior é 0.0 (Naidu *et al.*, 2023).

2.4.3.3 Precisão (Precision)

A Precisão, também conhecida como Valor Preditivo Positivo, foca nas predições positivas do modelo (Rainio *et al.*, 2024). Ela mede a proporção de predições positivas que estavam de fato corretas, sendo calculada como o número de verdadeiros positivos dividido pelo número total de predições positivas (TP + FP) (Naidu *et al.*, 2023). Sua fórmula é:

$$\text{Precisão} = \frac{TP}{TP + FP} \tag{2.2}$$

Assim como a acurácia, seu melhor valor é 1.0 e o pior é 0.0 (Naidu *et al.*, 2023).

2.4.3.4 Revocação (Recall ou Sensibilidade)

A Revocação, comumente chamada de Sensibilidade (*Sensitivity*) ou Taxa de Verdadeiros Positivos (*True Positive Rate*), mede a capacidade do modelo de identificar corretamente todas as instâncias que são realmente positivas (Rainio *et al.*, 2024). É calculada como o número de verdadeiros positivos dividido pelo número total de instâncias que pertencem à classe positiva (TP + FN) (Naidu *et al.*, 2023). A fórmula é definida como:

$$\text{Revocação} = \frac{TP}{TP + FN} \tag{2.3}$$

O valor ideal para a Revocação também é 1.0, e o pior é 0.0 (Naidu *et al.*, 2023).

2.4.3.5 F1-Score (F-Measure)

O *F1-Score*, ou *F-Measure*, é definido como a média harmônica da Precisão e da Revocação (Rainio *et al.*, 2024). Essa métrica é particularmente útil porque busca um equilíbrio entre as duas métricas anteriores, sendo uma medida de acurácia geral do teste (Naidu *et al.*, 2023). É calculado pela seguinte fórmula:

$$\textit{F1-Score} = 2 \times \frac{\text{Precisão} \times \text{Revocação}}{\text{Precisão} + \text{Revocação}} \tag{2.4}$$

O valor do *F1-Score* também varia de 0.0 (pior) a 1.0 (melhor) (Naidu *et al.*, 2023). O uso de uma média harmônica, em vez de uma média simples, penaliza mais severamente os valores extremos, o que torna o *F1-Score* uma métrica robusta para avaliar o desempenho do classificador.

2.4.4 O Papel do Aprendizado de Máquina em Cibersegurança

O Aprendizado de Máquina tornou-se uma tecnologia de cibersegurança importante, capaz de aprender continuamente ao analisar dados para identificar padrões, detectar malware em tráfego criptografado, encontrar ameaças internas e proteger dados na nuvem ao descobrir atividades suspeitas (Sarker, 2021). Uma vez treinado, o modelo pode ser utilizado em tempo real para monitorar novas atividades e indicar possíveis ataques. A principal vantagem dos sistemas baseados em ML reside na sua capacidade de identificar padrões complexos de comportamento e detectar anomalias sem a necessidade de regras fixas previamente definidas, o que os torna mais adaptáveis frente à evolução constante das ameaças.

No entanto, a aplicação de ML em ecossistemas de IoT apresenta desafios particulares. De acordo com Farooq *et al.* (2022), as infraestruturas de IoT são heterogêneas, incertas

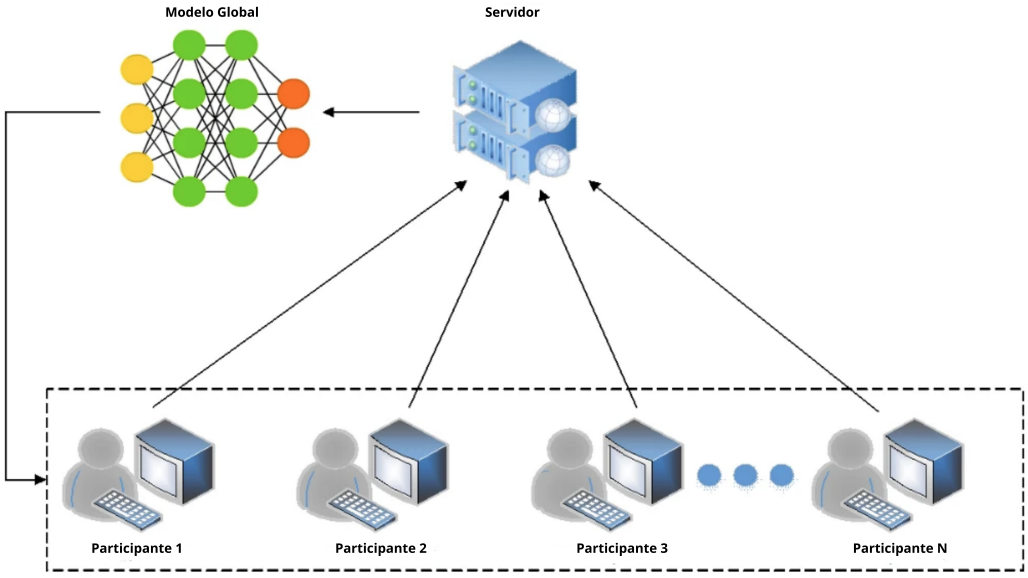
e produzem dados em massa. Portanto, é necessário considerar as limitações de recursos dos dispositivos, a sensibilidade a atrasos de certas aplicações e as necessidades da análise de dados. Outro obstáculo significativo está na preservação da privacidade, pois o envio de dados sensíveis para um servidor central pode representar riscos de segurança e de conformidade com regulamentações.

2.5 Aprendizado Federado

O Aprendizado Federado é uma abordagem de aprendizado de máquina distribuído e criptografado na qual múltiplos participantes podem construir um modelo de forma colaborativa sem a necessidade de divulgar seus dados subjacentes. Dessa forma, a informação sensível de cada participante não sai do seu ambiente local, garantindo a privacidade (Zhang *et al.*, 2021).

A arquitetura básica desse sistema, ilustrada na Figura 5, é composta por dois tipos de elementos: um servidor central, responsável por coordenar o treinamento, e um conjunto de participantes (ou clientes), que são os detentores dos dados. O processo ocorre de forma iterativa: o servidor distribui um modelo global inicial para os participantes. Cada participante treina o modelo com seus dados locais e envia apenas as atualizações (por exemplo, os gradientes ou pesos da rede neural) de volta ao servidor. Por fim, o servidor agrega as atualizações recebidas para aprimorar o modelo global, que é então enviado novamente aos clientes para a próxima rodada de treinamento.

Figura 5 – Estrutura básica da arquitetura de Aprendizado Federado.



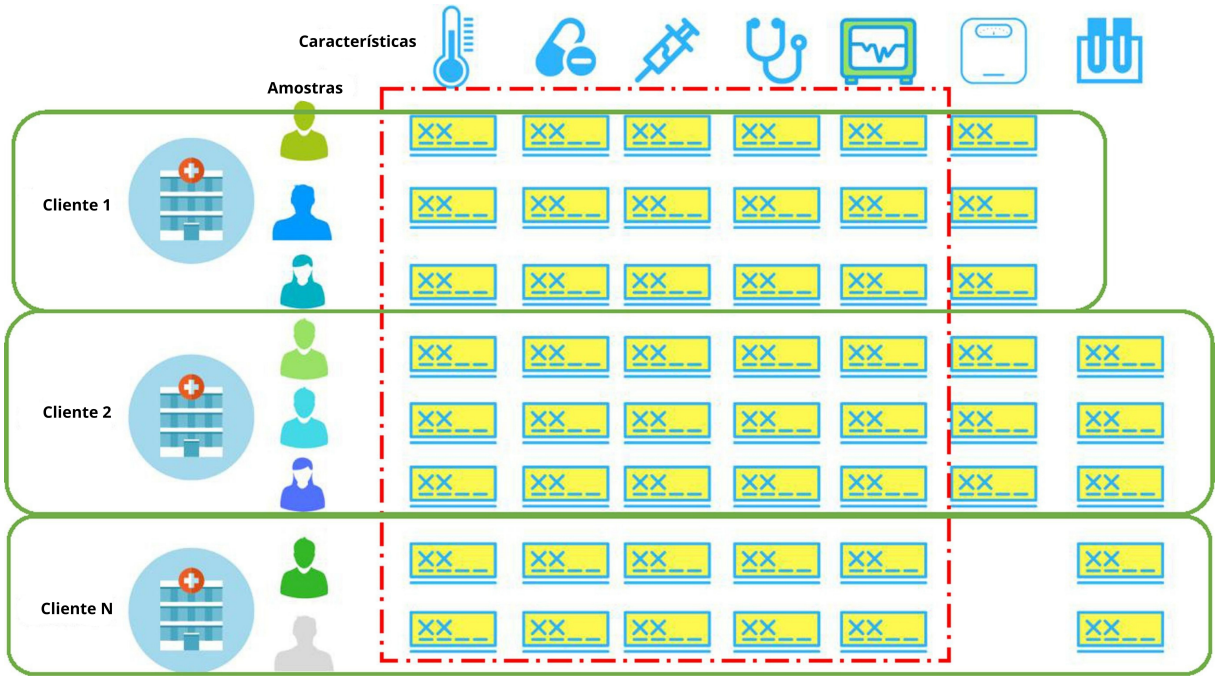
Fonte: Adaptada de Wen *et al.* (2023).

2.5.1 Aprendizado Federado Horizontal (Horizontal Federated Learning - HFL)

O HFL é aplicado em cenários nos quais os dados distribuídos entre diferentes participantes compartilham o mesmo espaço de características (*features*), mas possuem amostras de dados distintas (Li *et al.*, 2020). Um exemplo prático e comum de aplicação do HFL ocorre com dispositivos inteligentes, como em soluções para atualização de celulares Android. Nesses casos, os dados de cada usuário podem ser muito diferentes (espaço amostral distinto), mas as características coletadas pelo sistema são as mesmas para todos (espaço de características semelhante).

A Figura 6 ilustra essa arquitetura no contexto da saúde. Nela, diferentes hospitais (Clientes 1, 2, ..., N) possuem seus próprios conjuntos de pacientes (Amostras), mas todos registram o mesmo conjunto de informações e exames (Características). O HFL permite que esses hospitais treinem um modelo colaborativo sem compartilhar os dados sensíveis dos pacientes, aproveitando a diversidade de amostras para construir um modelo mais robusto e generalizável.

Figura 6 – Ilustração do Aprendizado Federado Horizontal.



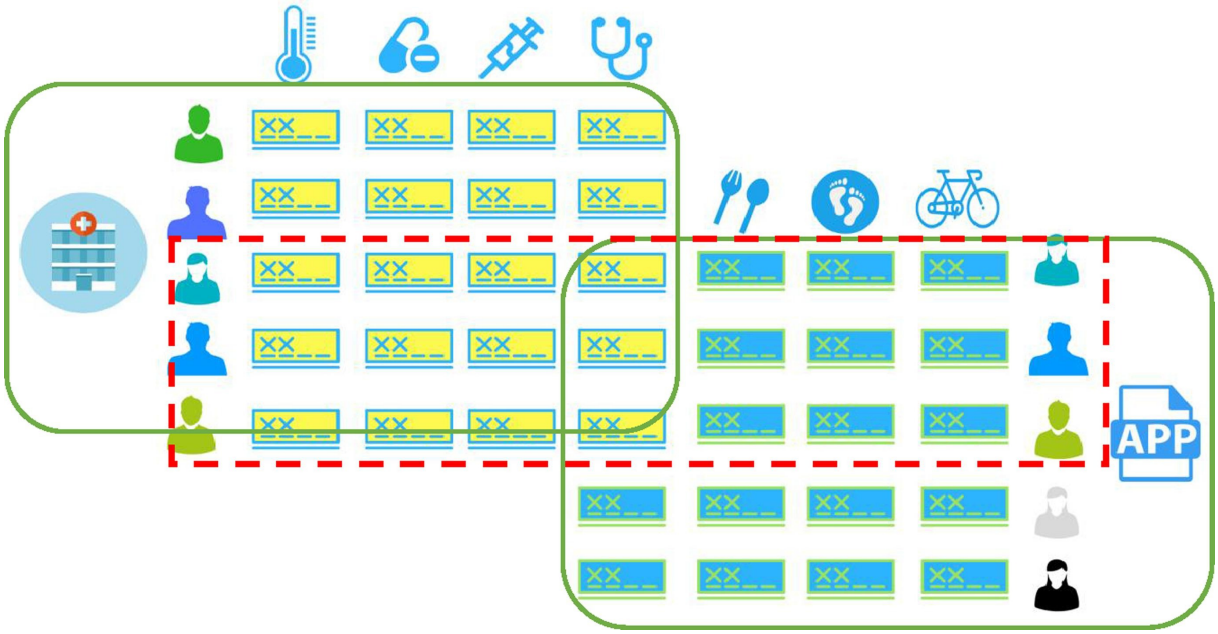
Fonte: Adaptada de Li *et al.* (2020).

2.5.2 Aprendizado Federado Vertical (Vertical Federated Learning - VFL)

VFL, também conhecido como FL baseado em recursos, é normalmente usado em configurações onde dois (ou mais) conjuntos de dados de clientes compartilham um espaço de ID de amostra semelhante, mas têm diferentes espaços de recursos de entrada (Banabilah *et al.*, 2022). Em outras palavras, diferentes organizações possuem informações diferentes sobre o mesmo grupo de indivíduos.

A Figura 7 exemplifica um caso de uso clássico para o VFL. Nela, um hospital e um aplicativo de estilo de vida atendem ao mesmo conjunto de usuários (amostras). O hospital detém as características médicas de cada indivíduo (temperatura, exames, etc.), enquanto o aplicativo possui suas características de atividade e dieta (passos, calorias, tipo de exercício). Utilizando o VFL, essas duas entidades podem colaborar para treinar um modelo mais completo, por exemplo, para avaliar riscos de saúde, sem que nenhuma delas precise revelar suas informações privativas à outra, alinhando os dados de forma segura através dos identificadores comuns dos usuários.

Figura 7 – Ilustração do Aprendizado Federado Vertical



Fonte: Retirada de Li *et al.* (2020).

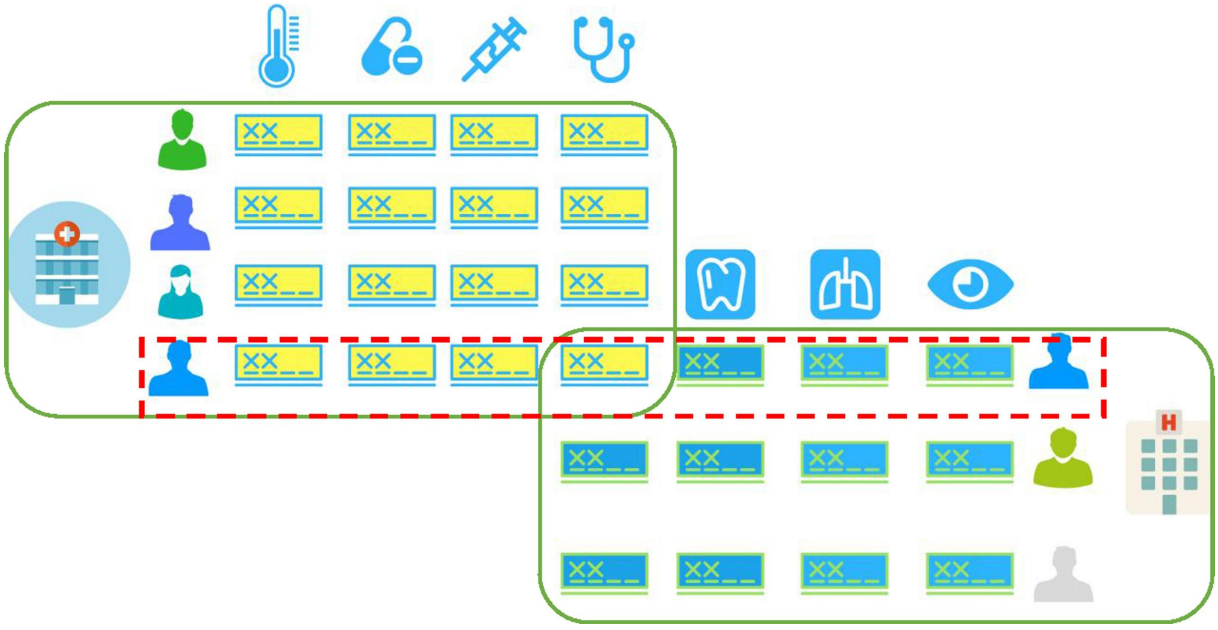
2.5.3 Aprendizagem de Transferência Federada (Federated Transfer Learning - FTL)

No caso de os usuários e os recursos dos usuário dos dois conjuntos de dados raramente se sobreporem, não segmentamos os dados, mas podemos utilizar o aprendizado de

transferência para superar a falta de dados ou tags. Esse método é chamado de FTL (Liu *et al.*, 2020). Essa técnica emprega o aprendizado por transferência para superar a falta de dados ou de rótulos em um dos domínios, permitindo que o conhecimento de um modelo seja aproveitado em outro.

Nela, duas instituições distintas, como um hospital geral (à esquerda) e uma clínica especializada (à direita), atendem a públicos diferentes e coletam dados distintos. O FTL possibilita que o conhecimento aprendido no domínio de origem (o hospital geral, com mais dados) seja transferido para auxiliar no treinamento de um modelo mais preciso no domínio de destino (a clínica especializada), que pode possuir dados mais escassos. Isso é feito de forma federada, preservando a privacidade de ambas as partes. A Figura 8 ilustra essa arquitetura.

Figura 8 – Ilustração da Aprendizagem de Transferência Federada.



Fonte: Retirada de Li *et al.* (2020).

2.6 Framework Flower

Para a implementação e orquestração do ambiente de Aprendizado Federado, este trabalho utiliza o *Flower* (*flwr*), um *framework* unificado projetado para viabilizar tanto a pesquisa experimental em simulações quanto a implantação de sistemas FL em larga escala em dispositivos reais de borda (Beutel *et al.*, 2020).

Diferente de outras ferramentas que focam exclusivamente em simulações em máquina única ou exigem *hardware* homogêneo, o *Flower* foi arquitetado para suportar a hetero-

geneidade de sistemas (diferentes capacidades de computação, memória e largura de banda) e escalar para grandes coortes de clientes (Beutel *et al.*, 2020). Esta característica é fundamental para este estudo, pois permite a simulação realista de múltiplos nós IoT com recursos limitados em um ambiente controlado.

No que tange à arquitetura, o lado do servidor é responsável por orquestrar o processo de aprendizado, gerenciando a seleção de clientes através do *ClientManager* e executando o laço de FL (*FL Loop*). A lógica de agregação e configuração é abstraída na classe *Strategy*, permitindo a implementação flexível de algoritmos como *FedAvg* e *FedProx* sem alterar o núcleo do sistema. Por sua vez, no lado do cliente, o *Flower* oferece uma implementação agnóstica de *Machine Learning* onde, através da abstração *Client*, é possível integrar o *framework* com diversas bibliotecas de treinamento, como *PyTorch* e *TensorFlow*, permitindo que os dispositivos locais utilizem seus próprios *pipelines* de treinamento e avaliação (Beutel *et al.*, 2020).

Uma inovação do *Flower* para a pesquisa acadêmica é o *Virtual Client Engine* (VCE). Este motor permite a virtualização de clientes para maximizar a utilização do *hardware* disponível, instanciando clientes sob demanda e gerenciando recursos de memória de forma transparente (Beutel *et al.*, 2020). Isso possibilita a execução de experimentos com centenas ou milhares de clientes simulados em uma única máquina física, viabilizando a validação de cenários *Non-IID* complexos sem a necessidade de um *cluster* de alto desempenho.

O *framework* fornece implementações de referência para algoritmos de agregação do estado da arte, facilitando a comparação de desempenho. Conforme detalhado por Beutel *et al.* (2020), a abstração de *Strategy* permite configurar a seleção de clientes disponíveis para as etapas de treinamento e avaliação, bem como a definição dos hiperparâmetros locais, tais como número de épocas e taxa de aprendizado, que são enviados aos clientes. Além disso, essa abstração gerencia a agregação dos pesos ou gradientes retornados, utilizando métodos consolidados como a Média Ponderada no algoritmo *FedAvg*.

3 TRABALHOS RELACIONADOS

A aplicação de ML e FL na segurança de redes IoT enfrenta desafios significativos relacionados à heterogeneidade dos dados e à necessidade de privacidade. Para fundamentar a investigação experimental deste estudo, foram selecionados cinco trabalhos que abordam diferentes facetas do problema: a avaliação experimental de dados *Non-IID*, a robustez dos algoritmos em cenários distribuídos, a aplicação prática de FL em segurança IoT e o desempenho de linhas de base centralizadas.

3.1 *Federated Learning on Non-IID Data Silos: An Experimental Study*

O trabalho de Li *et al.* (2022) apresenta uma investigação experimental aprofundada sobre a heterogeneidade dos dados distribuídos. Os autores desenvolveram um *benchmark* denominado *NIID-Bench* para sistematizar a avaliação de algoritmos de FL em cenários onde a distribuição dos dados varia drasticamente entre os clientes.

Os experimentos avaliaram quatro algoritmos fundamentais — *FedAvg*, *FedProx*, *SCAFFOLD* e *FedNova* — em diversos conjuntos de dados públicos. As principais conclusões indicam que o cenário de desbalanceamento de rótulos (*label distribution skew*) é o desafio mais crítico para o FL horizontal, causando a maior queda na acurácia dos modelos federados em comparação com o treinamento centralizado. Este trabalho é fundamental para o presente estudo, pois fornece a metodologia experimental e as estratégias de particionamento de dados necessárias para simular um ambiente realista de IoT utilizando o conjunto de dados *CICIoT2023*.

3.2 *Robust and Communication-Efficient Federated Learning From Non-i.i.d. Data*

O trabalho de Sattler *et al.* (2019) aborda a robustez do aprendizado federado em ambientes onde os dados não são independentes e identicamente distribuídos. Os autores destacam que a distribuição heterogênea dos dados degrada significativamente a convergência de algoritmos padrões como o *FedAvg*.

Embora os autores proponham um protocolo de compressão, a contribuição crucial de Sattler *et al.* (2019) para esta pesquisa é a comprovação experimental de que o *FedAvg* diverge em cenários *Non-IID*. Essa constatação fundamenta diretamente a decisão deste trabalho de adotar o algoritmo *FedProx*. Ao mitigar a instabilidade apontada pelos autores através de regularização, a abordagem proposta assegura a convergência do modelo mesmo diante dos

padrões de tráfego heterogêneos inerentes aos dispositivos IoT.

3.3 *Federated-Learning-Based Anomaly Detection for IoT Security Attacks*

No trabalho de Mothukuri *et al.* (2022), propõe-se uma abordagem prática de detecção de anomalias para redes IoT utilizando FL para superar as limitações de privacidade dos sistemas centralizados. Os autores implementaram um sistema descentralizado utilizando modelos baseados em *Gated Recurrent Units* (GRU) para identificar ataques em dados de tráfego IoT industrial (protocolo *Modbus*).

Os resultados experimentais evidenciam que a abordagem federada não apenas garantiu a privacidade dos dados, mas, em seus testes específicos, chegou a superar a versão centralizada em acurácia (99,25% contra 86,13%). Este trabalho serve como um importante precedente para esta pesquisa, demonstrando que é possível aplicar FL para detecção de intrusos em IoT com alta eficácia. No entanto, a validação de Mothukuri *et al.* (2022) foi conduzida em um ambiente simulado e com um conjunto de dados específico de IoT industrial, deixando em aberto a validação em cenários de tráfego de rede massivo e heterogêneo como o proposto pelo conjunto de dados *CICIoT2023*.

3.4 MARIA: Monitoramento e Análise para Resposta Imediata a Ataques à Rede 5G no Contexto da IoT

Desenvolvido por Silva *et al.* (2025), o sistema MARIA é uma solução projetada para enfrentar os desafios de segurança decorrentes da integração entre redes 5G e ambientes IoT. O trabalho destaca-se pela validação prática em um ambiente de *testbed* controlado.

O núcleo de detecção do MARIA utiliza um modelo de aprendizado de máquina supervisionado para identificar pacotes maliciosos. Após uma avaliação rigorosa de seis algoritmos, os autores selecionaram o algoritmo *CatBoost* treinado de forma centralizada, que apresentou o melhor desempenho com uma acurácia de 90,72% e um *F1-score* de 90,24%.

O diferencial do MARIA reside na sua eficácia comprovada. Contudo, o sistema opera sob um paradigma centralizado, exigindo que os dados de tráfego sejam agregados em um ponto central, o que levanta preocupações de privacidade. No contexto deste TCC, o sistema MARIA serve como o *baseline* de desempenho ideal. O objetivo desta pesquisa é investigar se uma arquitetura federada, operando sob as restrições de dados *Non-IID* e preservando a

privacidade, consegue atingir métricas de detecção de intrusos comparáveis a essa linha de base estabelecida.

3.5 *A Physics-Based Hyper Parameter Optimized Federated Multi-Layered Deep Learning Model for Intrusion Detection*

O trabalho recente de Chandnani *et al.* (2025) propõe o modelo *Fed-MLDL*, uma abordagem de aprendizado federado baseada em *Multilayer Perceptrons* (MLP) para detecção de intrusão em redes IoT. O estudo utiliza o mesmo conjunto de dados adotado nesta pesquisa, o *CICIoT2023*, além de outros *benchmarks*.

A principal contribuição dos autores é a utilização do algoritmo de otimização baseado em física *FedRIME* para o ajuste dinâmico de hiperparâmetros (como taxa de aprendizado e momento) em cada cliente. Os resultados reportados são expressivos, alcançando acurácias superiores a 99% em cenários distribuídos.

No entanto, uma distinção metodológica crítica deve ser notada: para lidar com o desbalanceamento de classes, os autores aplicaram a técnica SMOTE (*Synthetic Minority Over-sampling Technique*). Isso implica que o treinamento foi realizado com uma quantidade significativa de dados sintéticos gerados artificialmente para inflar as classes minoritárias. Embora eficaz para métricas laboratoriais, essa abordagem pode não refletir a realidade de dispositivos de borda em IoT, que raramente possuem capacidade computacional para gerar dados sintéticos em tempo real antes do treinamento. Este trabalho serve como um contraponto importante para a presente pesquisa, que busca avaliar o desempenho utilizando apenas dados reais e técnicas de subamostragem.

3.6 Comparação entre os Trabalhos

O Quadro 2 apresenta a comparação entre os trabalhos analisados e a proposta deste estudo. Os critérios destacam as diferenças metodológicas e os objetivos experimentais:

- (1) Abordagem Federada: Indica se o trabalho utiliza treinamento distribuído, ou *Federated Learning* (FL), em vez de centralizado.
- (2) Preservação da Privacidade: Refere-se à capacidade arquitetural de manter os dados sensíveis nos dispositivos de origem.
- (3) Foco em Segurança/*IoT*: Indica se o trabalho é aplicado especificamente ao domínio

de detecção de intrusos em redes *IoT*.

– (4) Análise de Dados *Non-IID*: Avalia se o estudo investiga explicitamente o impacto da distribuição heterogênea dos dados no desempenho.

– (5) *Benchmark* de Algoritmos: Indica a comparação experimental entre múltiplos algoritmos para validar o desempenho.

Quadro 2 – Comparação entre os trabalhos analisados e a proposta atual

Critério	**Li (2022)**	**Sattler (2020)**	**Mothukuri (2022)**	**MARIA (2025)**	**Chandnani (2025)**	**Proposta**
1	✓	✓	✓	✗	✓	✓
2	✓	✓	✓	✗	✓	✓
3	✗	✗	✓	✓	✓	✓
4	✓	✓	✗	✗	✗	✓
5	✓	✓	✓	✓	✗	✓

Fonte: Elaborado pelo autor.

4 METODOLOGIA

Neste Capítulo, são apresentados os procedimentos metodológicos adotados para o desenvolvimento deste trabalho. As etapas descritas detalham a abordagem investigativa utilizada para responder às questões de pesquisa, com ênfase na eficácia comparativa entre modelos centralizados e federados, no tratamento da heterogeneidade dos dados e na viabilidade arquitetural da proposta para redes IoT.

O Quadro 3 apresenta as questões norteadoras e seus respectivos objetivos que guiaram os experimentos.

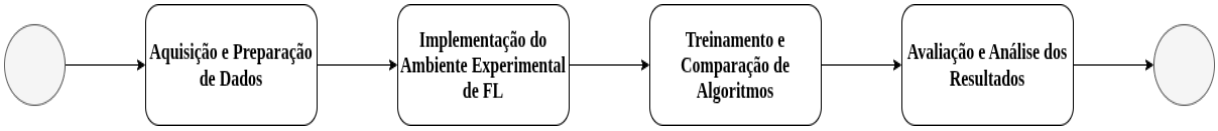
Quadro 3 – Questões de Pesquisa e seus respectivos Objetivos

ID	Questão de Pesquisa	Objetivo
QP1	Como o desempenho de detecção de intrusos do modelo federado se compara ao *baseline* de aprendizado centralizado (MARIA)?	Avaliar se o modelo federado atinge métricas de acurácia e F1-Score competitivas em relação à solução centralizada, validando sua eficácia técnica.
QP2	Qual o impacto da distribuição de dados Non-IID e do desbalanceamento de classes no treinamento do modelo?	Verificar a robustez dos algoritmos federados frente ao desbalanceamento severo de classes e propor estratégias de mitigação como o *Hybrid Undersampling*.
QP3	A arquitetura federada proposta é viável para preservar a privacidade dos dados em um cenário realista de IoT?	Validar se a distribuição do treinamento impede a exposição de dados brutos, mantendo a utilidade do modelo global.

Fonte: Elaborado pelo autor.

Para atingir os objetivos propostos, o estudo foi executado seguindo cinco etapas principais: (i) Aquisição e Preparação de Dados; (ii) Análise Exploratória e Engenharia de Atributos; (iii) Implementação do Ambiente Experimental de FL; (iv) Treinamento e Otimização de Hiperparâmetros; e (v) Avaliação e Análise dos Resultados. A Figura 9 ilustra o fluxo metodológico aplicado.

Figura 9 – Fluxo dos procedimentos metodológicos realizados



Fonte: Elaborado pelo autor.

4.1 Aquisição e Preparação de Dados

A primeira etapa consistiu na seleção e preparação do conjunto de dados. Foi utilizado o **CICIoT2023**, um *benchmark* moderno que contém tráfego realista de redes IoT e abrange 33 tipos de ataques, além de tráfego benigno.

Devido ao volume massivo e ao desbalanceamento extremo original do conjunto de dados (onde classes de DDoS representam a vasta maioria das amostras), foi desenvolvido um *pipeline* de engenharia de dados composto por três fases críticas:

4.1.1 Agrupamento Semântico (Label Grouping)

Para reduzir a dispersão das classes e focar o modelo nos comportamentos macroscópicos da rede, os 34 tipos de tráfegos originais foram remapeados para 8 famílias principais. Esta redução de dimensionalidade no espaço de saída facilita a convergência do modelo federado sem perda significativa de granularidade para a defesa de rede. As famílias definidas foram: *DDoS, DoS, Mirai, Recon, Spoofing, Web-based, BruteForce* e *Benign*.

4.1.2 Subamostragem Híbrida (Hybrid Undersampling)

Diferente de abordagens que recorrem a dados sintéticos (como *SMOTE*) para inflar artificialmente classes minoritárias, este trabalho optou por preservar a integridade dos dados reais mediante a aplicação de um limitador dinâmico de amostras durante o carregamento nos clientes.

Especificamente, para as Classes Volumétricas (DDoS/DoS), definiu-se um teto expandido de até 100.000 registros por cliente, permitindo que o modelo capture a variabilidade necessária dessas classes para maximizar a Acurácia Global. Simultaneamente, as Classes Raras (como *Web* e *BruteForce*) foram preservadas integralmente até um limite de 5.000 amostras, estratégia que impede o enviesamento total do modelo em direção às classes majoritárias e assegura a preservação do *F1-Score*.

4.1.3 Normalização e Particionamento

Foi aplicada a padronização *StandardScaler* aos atributos numéricos, ajustando os dados para média zero e desvio padrão unitário. O particionamento dos dados entre os clientes simulou um cenário Non-IID, onde cada cliente possui acesso apenas a fragmentos específicos

dos dados, refletindo a realidade de dispositivos IoT isolados (silos de dados).

4.2 Implementação do Ambiente Experimental de FL

Nesta etapa, foi construído o ambiente de simulação denominado *Federated Learning Intrusion Detection System* (FL-IDS). O ambiente foi desenvolvido na linguagem *Python*, utilizando a biblioteca *Flower* (*flwr*) para a orquestração do aprendizado federado, permitindo a comunicação assíncrona entre múltiplos clientes simulados e o servidor de agregação. A implementação foi estruturada de forma modular, facilitando sua reutilização e adaptação para futuros experimentos com diferentes conjuntos de dados ou arquiteturas de rede.

4.2.1 Arquitetura do Modelo Neural

Para a detecção de intrusos, foi desenvolvida uma arquitetura baseada em DNN, otimizada para ambientes federados. Diferente de *Multilayer Perceptrons* (MLPs) rasas tradicionais, a arquitetura proposta incorporou mecanismos específicos para lidar com a instabilidade do treinamento distribuído. Adotou-se uma estrutura sequencial profunda com camadas ocultas expandidas, contendo 128 e 256 neurônios, projetadas para capturar relações não-lineares complexas nos dados de tráfego. Em substituição ao *Batch Normalization*, optou-se pelo uso de *Layer Normalization*, uma vez que, em cenários federados *Non-IID*, as estatísticas de lote (*batch statistics*) variam drasticamente entre clientes, o que prejudica a convergência do *Batch Normalization*. O *Layer Normalization* normaliza as ativações por amostra, tornando o modelo mais robusto à heterogeneidade dos dados dos clientes. Além disso, utilizou-se a aplicação de camadas de *Dropout* de 30% após as ativações como medida de regularização para prevenir o *overfitting* nos dados locais dos clientes.

4.2.2 Ambiente Computacional e Reprodutibilidade

Todos os experimentos foram conduzidos em um ambiente computacional de uso geral (*commodity hardware*), intencionalmente restrito para validar a viabilidade da execução do treinamento federado em dispositivos com recursos limitados, aproximando-se da realidade de servidores de borda (*Edge Servers*) ou *Gateways* IoT.

As especificações do equipamento utilizado são:

– **Processador:** Intel Core i5 (9ª Geração);

- **Memória RAM:** 24 GB;
- **Sistema Operacional:** Linux (Ubuntu);
- **Linguagem e Frameworks:** Python 3.10, utilizando *PyTorch* para as redes neurais e *Flower* (Flwr) para a orquestração federada.

Vale ressaltar que, diferentemente de trabalhos correlatos que utilizam *clusters* de alto desempenho com múltiplas GPUs industriais (como NVIDIA A100) para processar grandes volumes de dados sintéticos, este trabalho obteve convergência estável utilizando apenas a CPU e a memória disponível neste equipamento.

4.3 Treinamento e Comparação de Algoritmos

As rodadas de treinamento federado foram conduzidas iterativamente. Em cada rodada, um subconjunto de clientes foi selecionado para treinar localmente com seus dados privados, enviando posteriormente as atualizações de pesos para o servidor global.

4.3.1 Estratégia de Agregação e Otimização

Para a agregação dos modelos, utilizou-se a estratégia *FedProx*. Este algoritmo introduz um termo proximal na função de perda local, limitando a divergência dos modelos clientes em relação ao modelo global, o que se mostrou crucial para manter a estabilidade em cenários *Non-IID*. O processo de treinamento envolveu um ajuste fino, ou *fine-tuning*, rigoroso dos hiperparâmetros. Em relação à taxa de aprendizado, ou *learning rate*, utilizou-se um escalonamento manual, iniciando com valores mais altos para uma exploração rápida do espaço de busca e reduzindo progressivamente para refinar a convergência. O número de épocas locais de treinamento foi aumentado para 10, permitindo que os clientes extraíssem características mais profundas de seus dados antes da etapa de agregação. Além disso, a participação dos clientes foi ajustada com uma fração de treinamento de 50%, o que corresponde a 65 clientes por rodada, garantindo que a agregação global fosse estatisticamente representativa e estável.

4.4 Avaliação e Análise dos Resultados

A avaliação quantitativa foi realizada utilizando um conjunto de teste global, composto por mais de dois milhões de amostras não vistas durante o treinamento, garantindo a validação em um cenário realista. Entre as métricas adotadas, utilizou-se a acurácia para medir

o percentual global de acertos, sendo esta fortemente influenciada pela capacidade de detectar a classe majoritária. O *F1-Score* (no modo Macro) foi definido como a métrica principal para avaliar o desempenho em classes desbalanceadas, garantindo que o modelo mantenha a capacidade de detectar ataques raros e não apenas prioritize as classes volumétricas. Também foram analisadas a precisão e o *recall* para verificar a taxa de falsos positivos e a sensibilidade do modelo a ameaças reais. Adicionalmente, utilizou-se o mecanismo de *checkpointing* para salvar o estado do modelo global a cada rodada, permitindo a recuperação e a análise contínua do processo de convergência.

5 RESULTADOS E DISCUSSÃO

Neste capítulo, são apresentados e discutidos os resultados obtidos com a implementação do sistema de detecção de intrusos baseado em Aprendizado Federado. A exposição inicia-se com a caracterização do conjunto de dados e as etapas de pré-processamento, fundamentais para a compreensão dos desafios enfrentados. Em seguida, são detalhados os experimentos realizados, comparando diferentes configurações de hiperparâmetros e estratégias de balanceamento, culminando na análise de desempenho do modelo final proposto e sua contextualização frente aos trabalhos relacionados.

5.1 Conjunto de Dados e Caracterização

Para a realização deste estudo, foi utilizado o conjunto de dados **CICIoT2023**, retirado da pesquisa do Neto *et al.* (2023), um *benchmark* de referência para segurança em *IoT*. A escolha justifica-se pela sua escala massiva e diversidade de ataques, compreendendo 33 tipos distintos organizados em 7 categorias principais, totalizando mais de 45 milhões de registros.

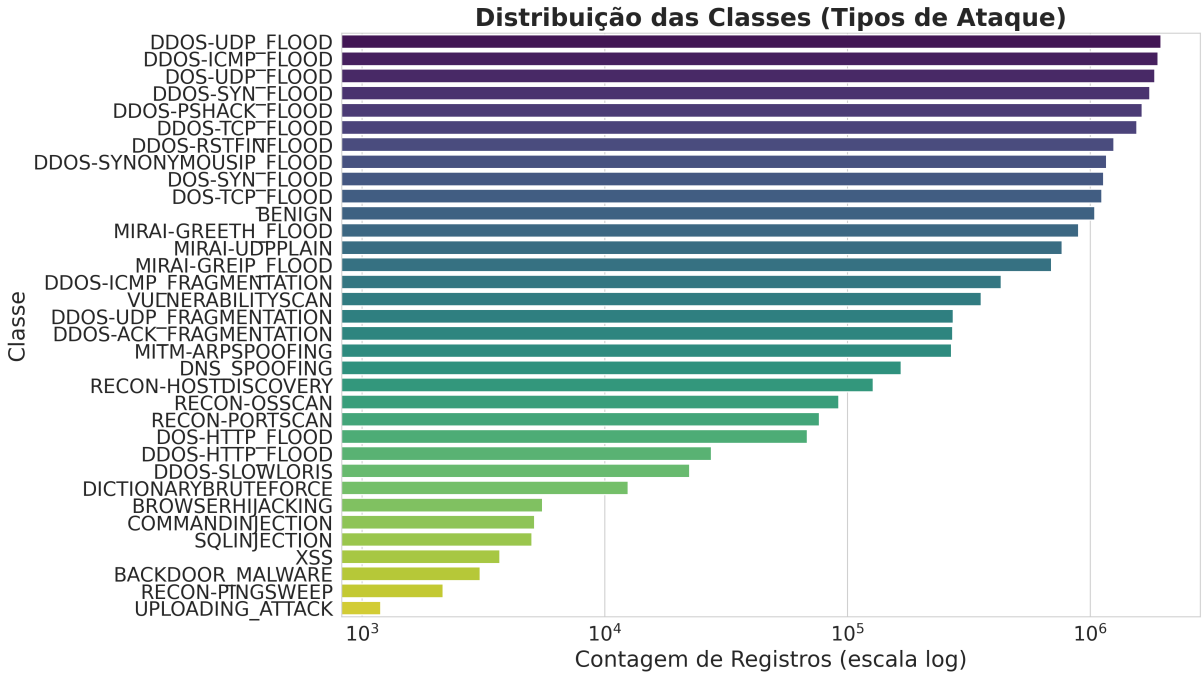
A Análise Exploratória dos Dados (AED), realizada preliminarmente, revelou um desbalanceamento severo entre as classes. Como ilustrado na Figura 10, ataques volumétricos do tipo *DDoS* dominam o conjunto de dados, enquanto ataques críticos como *SQL Injection* e *XSS* representam uma fração ínfima das amostras.

A visualização da estrutura dos dados via PCA (Figura 11) confirmou a complexidade da tarefa: embora existam agrupamentos distintos para alguns ataques, há uma sobreposição significativa na região central do espaço de características, sugerindo que modelos lineares simples teriam dificuldade na separação das classes.

5.2 Evolução dos Experimentos e Ajuste de Hiperparâmetros

O desenvolvimento do modelo seguiu uma abordagem iterativa, buscando o equilíbrio entre a capacidade de generalização (Acurácia Global) e a capacidade de detecção de classes minoritárias (*F1-Score*). Os experimentos foram conduzidos em um ambiente federado simulado com 130 clientes, utilizando a estratégia de agregação *FedProx*.

Figura 10 – Distribuição das classes de ataque no conjunto de dados CICIoT2023, evidenciando o desbalanceamento extremo.



Fonte: Elaborado pelo autor.

5.2.1 Fase 1: Balanceamento Estrito e o Limite de Aprendizado

Inicialmente, aplicou-se uma subamostragem estrita, limitando todas as classes a 5.000 amostras por cliente. Embora essa abordagem tenha garantido um *F1-Score* equilibrado, a acurácia global estagnou em torno de 76%. Isso ocorreu porque o modelo, ao ver poucos exemplos da classe majoritária (*DDoS*), não conseguiu capturar a variabilidade total desse tipo de ataque, sendo penalizado no teste global.

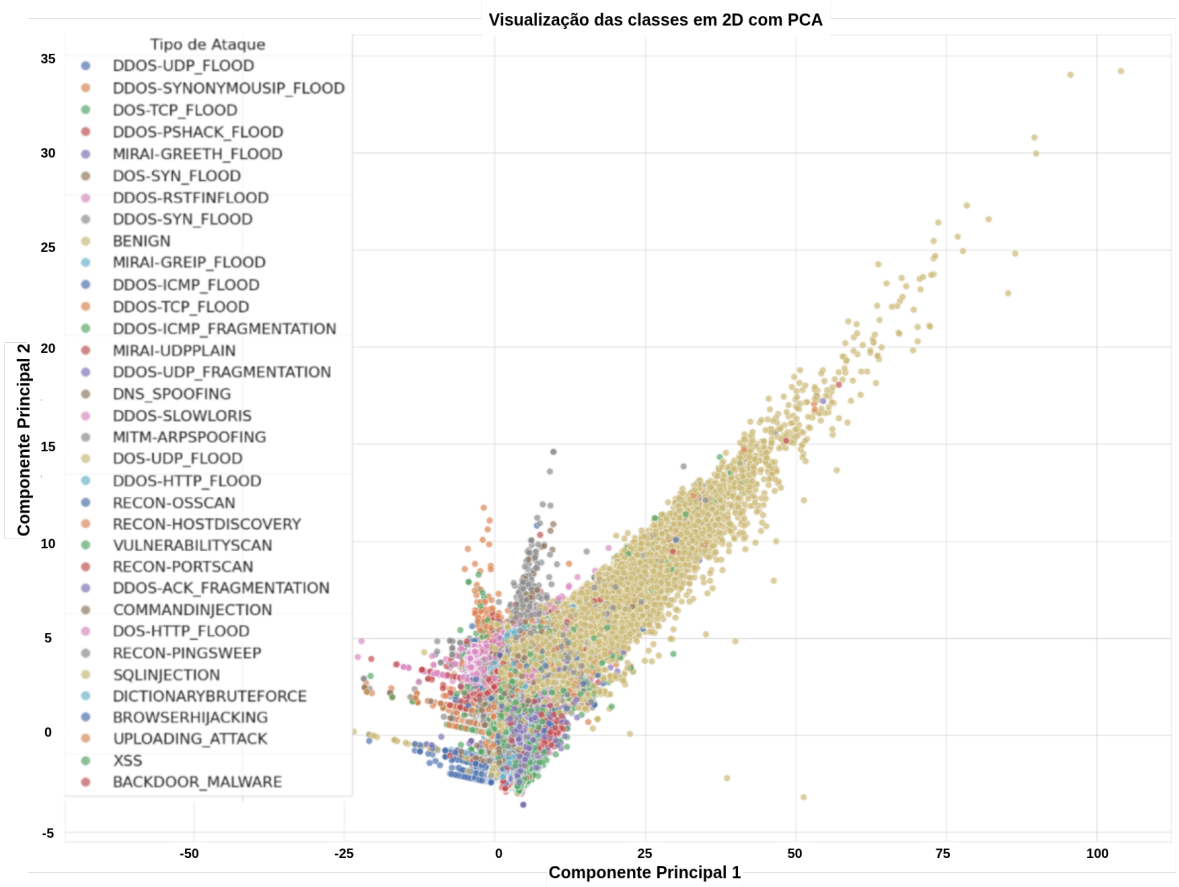
5.2.2 Fase 2: Estratégia Híbrida e Convergência

Para superar o gargalo identificado, implementou-se a estratégia de **Subamostragem Híbrida**, permitindo que as classes volumétricas (*DDoS/DoS*) tivessem até 100.000 amostras, enquanto as classes raras mantiveram o teto de 5.000.

A Tabela 1 resume a evolução do desempenho do modelo final (*Net v5* com *LayerNorm*) ao longo das rodadas de comunicação, utilizando a configuração otimizada de taxa de aprendizado ($\eta = 0.001$) e épocas locais ($E = 8$).

Observa-se um crescimento consistente das métricas. O salto de desempenho entre as rodadas 10 e 20 demonstra a eficácia do aumento da participação dos clientes (*fraction_train*

Figura 11 – Visualização 2D (PCA) demonstrando a sobreposição de classes no espaço de características.



Fonte: Elaborado pelo autor.

Tabela 1 – Evolução das Métricas de Desempenho do Modelo Global (Teste Centralizado)

Rodada Global	Acurácia	F1-Score	Precisão	Recall
10	74.57%	75.97%	82.80%	74.57%
20	79.89%	76.52%	79.48%	79.89%
30	81.19%	79.00%	80.34%	81.19%
40 (Final)	**82.80%**	**81.20%**	**82.10%**	**82.80%**

Fonte: Elaborado pelo autor.

= 0.5), que estabilizou a agregação dos pesos.

5.3 Análise do Modelo Final

O modelo final atingiu uma **Acurácia Global de 82,80%**. Mais importante que a acurácia bruta é o ***F1-Score* de 81,20%**, que indica que o modelo não se tornou enviesado apenas para a classe majoritária.

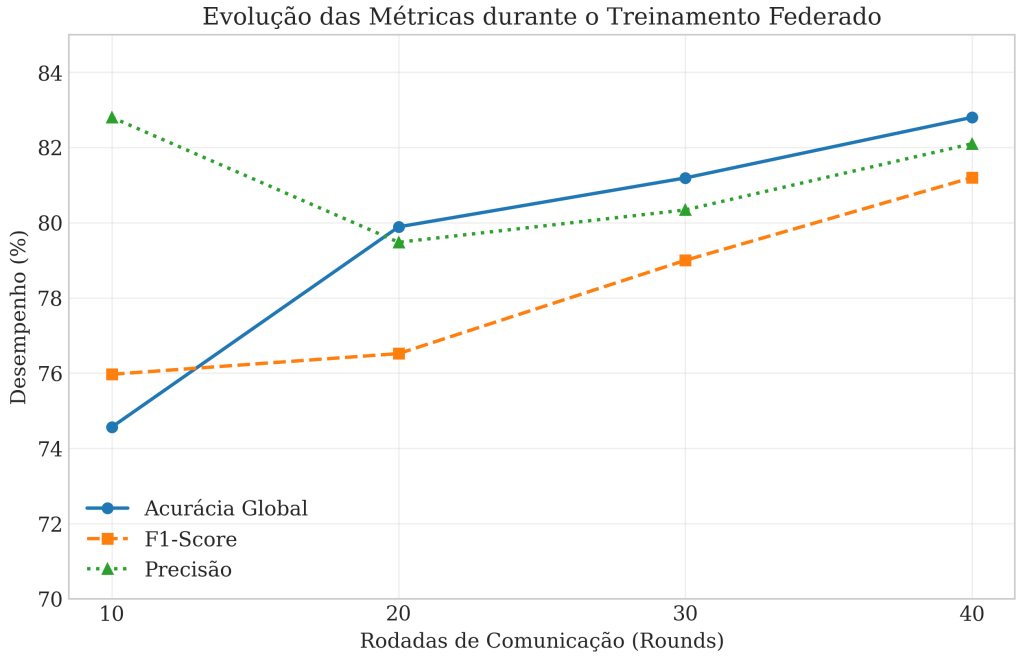
A Alta Precisão (82,10%) é um indicador crítico para sistemas de detecção de

intrusos, pois reflete uma baixa taxa de falsos positivos. Isso significa que, quando o sistema alerta sobre um ataque, há uma alta probabilidade de ser uma ameaça real, o que é essencial para evitar a fadiga de alertas em administradores de rede.

5.3.1 Análise Visual da Convergência

A Figura 12 ilustra a trajetória de aprendizado do modelo global ao longo das 40 rodadas de comunicação. Observa-se que, após a introdução da estratégia de aumento de participação dos clientes e ajuste fino da taxa de aprendizado, houve um ganho significativo de estabilidade. O *F1-Score* acompanhou o crescimento da acurácia, demonstrando que o modelo manteve sua capacidade de generalização para classes minoritárias.

Figura 12 – Evolução das métricas de desempenho (Acurácia, F1-Score e Precisão) ao longo das rodadas de treinamento federado.



Fonte: Elaborado pelo autor.

5.3.2 Desempenho Detalhado por Família de Ataque

Para compreender as nuances de detecção do modelo federado, a Tabela 2 apresenta as métricas de desempenho desagregadas por família de ataque, obtidas no conjunto de teste global consolidado (partições 130 a 144, totalizando 2.175.215 amostras).

A análise detalhada dos resultados revela pontos de alta eficácia, bem como limita-

Tabela 2 – Métricas de desempenho por classe de ataque (Modelo Federado Final).

Classe	Precisão	Recall	F1-Score	Suporte (Amostras)
DDoS	0,8277	0,9518	0,8854	1.273.468
DoS	0,7476	0,4206	0,5383	433.078
Mirai	0,9998	0,9924	**0,9961**	243.711
Recon	0,6678	0,7745	0,7172	67.982
Spoofing	0,7780	0,6240	0,6925	45.064
Web-based	0,0000	0,0000	0,0000	2.120
Brute Force	0,8693	0,1949	0,3184	1.365
Benign	0,7856	0,7968	0,7912	108.427

Fonte: Elaborado pelo autor.

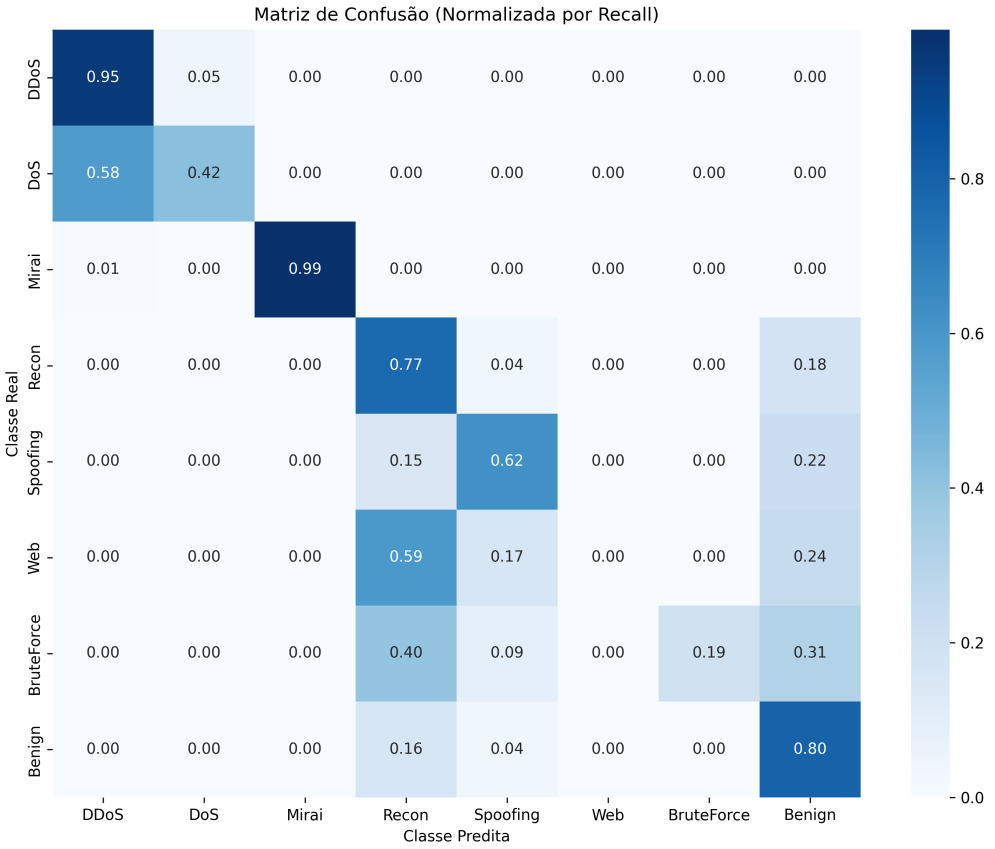
ções inerentes à abordagem adotada. Destaca-se, primeiramente, a robustez do modelo contra *Botnets*, obtendo um desempenho quase perfeito na detecção de ataques Mirai, com um *F1-Score* de 0,9961. Considerando que a *Mirai* é a *botnet* mais prevalente em dispositivos IoT, este resultado valida a utilidade prática do sistema proposto para a segurança de borda. Simultaneamente, a classe majoritária DDoS apresentou uma excelente recuperação (*Recall* de 95%), indicando que o sistema é eficaz em bloquear inundações massivas de rede.

Em contrapartida, observa-se o fenômeno das classes raras, onde categorias como *Web-based* e *Brute Force* apresentaram desempenho crítico, com a classe *Web* não sendo detectada (0,00). Este comportamento deve-se ao extremo desbalanceamento dos dados, visto que a classe *Web* representa menos de 0,1% do conjunto de teste. Diferentemente de trabalhos correlatos que utilizam dados sintéticos (*SMOTE*) para inflar artificialmente essas classes, este trabalho optou pela fidelidade aos dados reais. O resultado evidencia que, em um ambiente federado estrito e sem a geração de dados falsos, ataques extremamente raros tendem a ser suprimidos pelos gradientes das classes volumétricas durante o processo de agregação com FedProx.

A Figura 13 detalha as confusões críticas observadas durante a inferência. Nota-se que a maior taxa de confusão ocorre na classe *DoS* (com *Recall* de apenas 42,06%), que é frequentemente classificada incorretamente como *DDoS*. Esse comportamento é justificado pela alta similaridade espectral entre as duas categorias, diferenciadas muitas vezes apenas pelo volume e origem dos pacotes, uma nuance difícil de capturar em um ambiente federado com agregação de pesos.

Adicionalmente, a matriz confirma a supressão da classe *Web-based*, cujas amostras foram dispersas majoritariamente entre as classes *Benign* e *DDoS*. Isso demonstra que, embora o modelo seja robusto para ataques volumétricos massivos, ele conserva uma margem de erro para ataques de baixa assinatura ou que compartilham características estruturais com a classe

Figura 13 – Matriz de Confusão Normalizada do modelo final.



Fonte: Elaborado pelo autor.

majoritária.

5.3.3 Análise de Custo Computacional e de Comunicação

A viabilidade de implantação em dispositivos de borda depende estritamente do consumo de recursos. O modelo proposto (DNN Profunda) possui apenas 73.096 parâmetros treináveis, resultando em um arquivo de apenas **292 KB**.

Considerando o cenário experimental com 40 rodadas de comunicação, destaca-se a eficiência de tráfego obtida. Cada cliente precisou transmitir e receber apenas o arquivo de pesos a cada rodada, resultando em um consumo total de dados por dispositivo inferior a **25 MB** ao final do treinamento. Em contrapartida, uma abordagem centralizada tradicional exigiria a transmissão de gigabytes de dados brutos (*PCAP/CSV*) de todos os dispositivos para a nuvem. Esses números comprovam que a abordagem federada proposta reduz drasticamente o uso de largura de banda, apresentando-se como uma solução ideal para redes restritas, como 4G ou *LoRaWAN*.

5.4 Trade-off: Privacidade, Desempenho e Viabilidade

A síntese dos resultados estabelece um claro compromisso (*trade-off*) tripartido enfrentado neste estudo:

1. **Privacidade (Alta):** Diferente do baseline centralizado (MARIA), a abordagem federada garante, criptograficamente e arquiteturalmente, que os dados sensíveis saiam das redes IoT.
2. **Desempenho (Moderado):** Houve uma redução de acurácia global ($\approx$7%) em comparação ao estado da arte. A análise experimental sugere que essa diferença é majoritariamente fruto da arquitetura leve escolhida, e não apenas do processo federado.
3. **Viabilidade (Alta):** Ao utilizar dados reais (sem *SMOTE*) e uma rede neural compacta (<300 KB), o sistema apresenta fortes indícios para ser executável em *gateways* IoT reais, ao contrário de soluções que dependem de geração de dados sintéticos ou modelos de *ensemble* pesados.

5.5 Comparação com Trabalhos Relacionados

Para validar a eficácia da solução proposta e contextualizar os resultados obtidos, realizou-se um comparativo direto com trabalhos representativos do estado da arte, conforme apresentado na Tabela 3 e visualizado na Figura 14.

Tabela 3 – Comparativo de Desempenho e Características Arquiteturais.

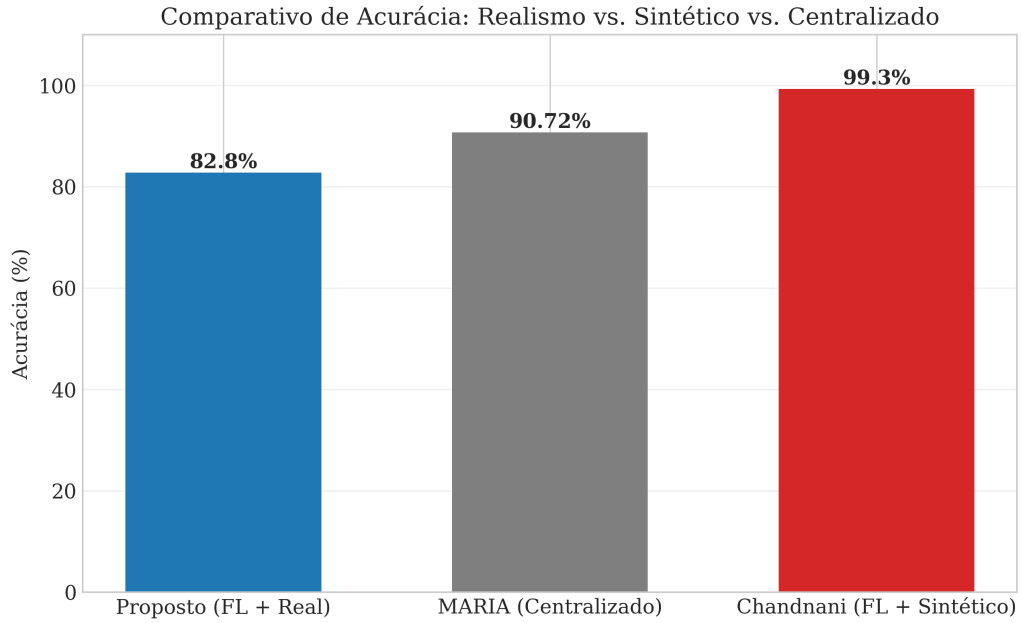
Trabalho	Abordagem	Dados	Acurácia	Privacidade
MARIA (Silva *et al.*, 2025)	Centralizado (CatBoost)	Reais	90,72%	Baixa (Dados centralizados)
Chandnani et al. (Chandnani *et al.*, 2025)	Federado (MLP)	Sintéticos (SMOTE)	99,30%	Alta
Proposto (FL-IDS)	**Federado (DNN Profunda)**	**Reais (Híbrido)**	**82,80%**	**Alta**

Fonte: Elaborado pelo autor.

5.5.1 Discussão e Validação Teórica

A análise comparativa e os fenômenos observados durante os experimentos permitem estabelecer conexões importantes com a literatura fundamental da área. Quando confrontada com a abordagem centralizada do sistema MARIA, observa-se que este modelo atinge cerca de 90% de acurácia, enquanto a solução federada proposta alcança aproximadamente 83%, representando uma redução de sete pontos percentuais. Esta diferença caracteriza o custo da privacidade. Em contrapartida a essa variação de desempenho, a solução federada assegura que

Figura 14 – Comparativo de acurácia entre a abordagem proposta (dados reais) e abordagens baseadas em dados sintéticos ou centralizados.



Fonte: Elaborado pelo autor.

os dados de tráfego sensíveis deixem os dispositivos IoT, atendendo a requisitos regulatórios de privacidade.

Em relação à abordagem com dados sintéticos, o trabalho de Chandnani *et al.* (2025) reporta uma acurácia de 99,3% utilizando a técnica SMOTE para gerar dados sintéticos . Embora impressionante, a geração de dados sintéticos em dispositivos de borda é computacionalmente proibitiva. A abordagem deste trabalho, ao utilizar dados reais com *Hybrid Undersampling*, oferece uma estimativa de desempenho mais honesta e robusta para implantação em produção.

No que tange à robustez e heterogeneidade, a instabilidade observada nas fases iniciais do treinamento corrobora os achados de Li *et al.* (2022), que identificam o *label skew* como o cenário mais desafiador para algoritmos federados. Além disso, conforme argumentado por Sattler *et al.* (2019), a divergência estatística entre clientes tende a degradar a convergência do *FedAvg*. Os resultados deste trabalho demonstram que a adoção do *FedProx* combinada com a normalização robusta via *LayerNorm* foi eficaz para mitigar essa divergência, permitindo que o modelo superasse a barreira dos 80%.

Por fim, ao contextualizar o desempenho frente ao trabalho de Mothukuri *et al.* (2022), embora tenham sido reportadas acurácias superiores a 99% com *Federated Learning* , o estudo focou em protocolos industriais como o *Modbus*, que possuem padrões mais determinísticos. Ao enfrentar o conjunto de dados CICIoT2023, que simula uma rede IoT aberta e massiva,

a acurácia de 82,80% obtida neste trabalho representa um resultado robusto dada a fronteira de decisão mais difusa e complexa deste cenário. Em suma, os resultados obtidos representam um equilíbrio ótimo para sistemas reais, onde a fidelidade aos dados de tráfego e a privacidade do usuário são prioritárias em relação à maximização artificial de métricas.

5.5.2 Análise do Limite Arquitetural (Centralizado vs. Federado)

Para isolar o impacto do paradigma federado no desempenho, realizou-se um experimento de controle treinando a mesma arquitetura DNN de forma centralizada. Observou-se que o modelo centralizado atingiu acurácia de ≈76,5% nas primeiras épocas, com uma curva de convergência gradual similar à do cenário federado.

Isso indica que a diferença de desempenho em relação ao estado da arte (MARIA, 90,7%) não deve ser atribuída exclusivamente às dificuldades do aprendizado federado (Non-IID), mas sim à **capacidade da arquitetura**. Enquanto o MARIA utiliza *ensembles* robustos (*CatBoost*) de alta complexidade computacional, a proposta deste trabalho utiliza uma DNN leve (≈292 KB). Portanto, o teto de performance de ≈83% reflete uma escolha de projeto consciente: priorizar um modelo leve e exportável para dispositivos IoT em detrimento da maximização absoluta da acurácia via modelos massivos.

5.6 Síntese e Resposta às Questões de Pesquisa

Os experimentos realizados e discutidos ao longo deste capítulo forneceram as evidências empíricas necessárias para validar o objetivo central do trabalho e responder objetivamente às questões de pesquisa (QP).

Em resposta à **QP1**, que investigou o desempenho comparativo do modelo federado frente a abordagens centralizadas, os resultados demonstraram que o sistema atingiu uma Acurácia Global de 82,80%. Embora este valor apresente uma redução em relação aos 90,72% do estado da arte centralizado (MARIA), a análise de controle (onde a mesma arquitetura DNN atingiu ≈76,5% em modo centralizado) evidencia que o desempenho é limitado primariamente pela escolha de uma arquitetura leve, e não apenas pelo processo federado. O sistema, portanto, oferece um nível de detecção competitivo para ameaças críticas, pagando um custo aceitável em troca da privacidade.

Quanto à **QP2**, referente ao impacto da heterogeneidade dos dados e classes des-

balanceadas, observou-se um comportamento dual. A estratégia de *Hybrid Undersampling* combinada com o algoritmo *FedProx* garantiu robustez excepcional contra classes volumétricas e prevalentes (*Mirai* com 99,6% de F1-Score e *DDoS* com 88,5%). Por outro lado, a fidelidade aos dados reais expôs a dificuldade intrínseca de modelos federados em aprender classes com representatividade ínfima sem o uso de dados sintéticos, como evidenciado pela não detecção da classe *Web-based*.

Por fim, atendendo à **QP3** sobre a viabilidade técnica e operacional em dispositivos de borda, a análise de custos comprovou a eficiência da solução proposta. Com um modelo final de apenas 292 KB e um consumo total de tráfego inferior a 25 MB por cliente ao longo de 40 rodadas, o sistema validou-se como tecnicamente viável para *gateways* IoT com restrições severas de largura de banda e processamento, superando as limitações de transferência de dados brutos exigidas por sistemas centralizados.

Esses achados consolidam o IDS federado proposto como uma solução de equilíbrio ideal entre a soberania dos dados e a segurança da rede, fundamentando as conclusões finais e as sugestões de trabalhos futuros apresentadas no capítulo a seguir.

6 CONCLUSÕES E TRABALHOS FUTUROS

Este trabalho investigou a viabilidade e a eficácia da aplicação de FL para a detecção de intrusos em redes *IoT*, um cenário caracterizado por volumes massivos de dados, desbalanceamento extremo de classes e restrições severas de privacidade. Ao longo do desenvolvimento, partiu-se de uma arquitetura básica de *Multilayer Perceptron*, que se mostrou insuficiente diante da complexidade do conjunto de dados *CICIoT2023*, evoluindo para uma arquitetura baseada em Redes Neurais Profundas com conexões residuais, otimizada especificamente para ambientes distribuídos.

A investigação foi guiada por três questões de pesquisa fundamentais. Em resposta à primeira questão, sobre a comparação de desempenho com linhas de base centralizadas, os resultados demonstraram que o modelo federado atingiu uma acurácia global de 82,89% e um *F1-Score* de 81,30%. Embora inferior aos 90,72% reportados pelo sistema centralizado MARIA, este resultado estabelece um *trade-off* claro: aceita-se uma redução na métrica de detecção absoluta em troca da garantia criptográfica de que os dados sensíveis de tráfego em regra, não deixam o dispositivo de origem.

Quanto ao impacto da heterogeneidade dos dados, abordado na segunda questão de pesquisa, observou-se que a distribuição desigual das classes de ataque entre os clientes causou instabilidade significativa nas fases iniciais do treinamento. A solução encontrada exigiu uma engenharia robusta, combinando a estratégia de agregação *FedProx*, a substituição de *Batch Normalization* por *Layer Normalization* e, a implementação de uma estratégia de *Hybrid Undersampling*. Esta última permitiu que o modelo aprendesse a distinguir ameaças raras sem ser completamente enviesado pela classe majoritária de ataques volumétricos, como *DDoS*.

Por fim, em relação à viabilidade técnica, este estudo provou que é possível treinar modelos robustos em *hardware* de consumo (processador intermediário e sem aceleração de GPU dedicada para o servidor), simulando um cenário realista de servidores de borda (*Edge Computing*). A decisão metodológica de utilizar apenas dados reais, ao contrário de trabalhos correlatos que alcançam métricas superiores a 99% através da geração de dados sintéticos (*SMOTE*), posiciona esta pesquisa como um referencial de desempenho mais honesto e aplicável a dispositivos *IoT* reais, que tipicamente não possuem capacidade computacional para gerar dados sintéticos em tempo real.

Em suma, o sistema proposto demonstrou que a privacidade em redes *IoT* não precisa custar a segurança da rede. Com as otimizações arquiteturais adequadas, o aprendizado federado

oferece um caminho viável para a construção de sistemas de defesa colaborativos e resilientes.

6.1 Trabalhos Futuros

Apesar dos avanços alcançados, o campo de segurança em *IoT* com aprendizado federado apresenta vastas oportunidades para exploração. Uma direção promissora consiste na investigação de técnicas de *model pruning* e quantização de pesos, visando reduzir o tamanho dos pacotes transmitidos entre clientes e servidor, o que se torna crítico para redes *IoT* operando com largura de banda restrita, como *LoRaWAN* ou *NB-IoT*. Simultaneamente, é fundamental avaliar a robustez do modelo contra ataques adversariais, especificamente os de envenenamento, conhecidos como *data poisoning* ou *model poisoning*, onde clientes maliciosos tentam corromper o modelo global, exigindo a implementação de mecanismos de agregação bizantina robusta para aumentar a segurança do sistema.

No âmbito da privacidade, sugere-se a adição de camadas de ruído estatístico através de *Differential Privacy* aos gradientes enviados pelos clientes, enfrentando o desafio de encontrar o ponto de equilíbrio onde a privacidade é maximizada sem degradar a acurácia do modelo abaixo de níveis operacionais aceitáveis. Além disso, o desenvolvimento de mecanismos de aprendizado contínuo, permitiria ao modelo federado aprender novos tipos de ataques, como os de dia zero, sem incorrer no esquecimento do conhecimento adquirido sobre ataques antigos, fenômeno conhecido como *catastrophic forgetting*, sendo esta uma característica essencial para a natureza dinâmica das ameaças cibernéticas. Por fim, recomenda-se a exploração de outras arquiteturas, testando *Transformers* leves ou redes temporais, adaptadas para o contexto federado, com o objetivo de verificar se estas conseguem capturar dependências temporais nos fluxos de pacotes que a abordagem atual baseada em estatísticas estáticas pode ter perdido.

REFERÊNCIAS

BANABILAH, S.; ALOQAILY, M.; ALSAYED, E.; MALIK, N.; JARARWEH, Y. Federated learning review: Fundamentals, enabling technologies, and future applications. **Information Processing Management**, v. 59, n. 6, p. 103061, 2022. ISSN 0306-4573. Disponível em: https://www.sciencedirect.com/science/article/pii/S0306457322001649. Acesso em: 10 jun. 2025.

BEUTEL, D. J.; TOPAL, T.; MATHUR, A.; QIU, X.; FERNANDEZ-MARQUES, J.; GAO, Y.; SANI, L.; LI, K. H.; PARCOLLET, T.; GUSMÃO, P. P. B. D. *et al.* Flower: A friendly federated learning research framework. **arXiv preprint arXiv:2007.14390**, 2020. Disponível em: https://doi.org/10.48550/arXiv.2007.14390. Acesso em: 05 jun. 2025.

BOOIJ, T. M.; CHISCOP, I.; MEEUWISSEN, E.; MOUSTAFA, N.; HARTOG, F. T. H. d. ToN_IoT: The role of heterogeneity and the need for standardization of features and attack types in IoT network intrusion data sets. **IEEE Internet of Things Journal**, v. 9, n. 1, p. 485–496, 2022. Disponível em: https://doi.org/10.1109/JIOT.2021.3085194. Acesso em: 20 ago. 2025.

CHANDNANI, C. J.; AGARWAL, V.; KULKARNI, S. C.; AREN, A.; AMALI, D. G. B.; SRINIVASAN, K. A physics-based hyper parameter optimized federated multi-layered deep learning model for intrusion detection in iot networks. **IEEE Access**, v. 13, p. 21992–22010, 2025. Disponível em: https://doi.org/10.1109/ACCESS.2025.3535952. Acesso em: 01 ago. 2025.

CRISCI, C.; GHATTAS, B.; PERERA, G. A review of supervised machine learning algorithms and their applications to ecological data. **Ecological Modelling**, v. 240, p. 113–122, 2012. ISSN 0304-3800. Disponível em: https://www.sciencedirect.com/science/article/pii/S0304380012001081. Acesso em: 10 jun. 2025.

DOMíNGUEZ-BOLAñO, T.; CAMPOS, O.; BARRAL, V.; ESCUDERO, C. J.; GARCíA-NAYA, J. A. An overview of iot architectures, technologies, and existing open-source projects. **Internet of Things**, v. 20, p. 100626, 2022. ISSN 2542-6605. Disponível em: https://www.sciencedirect.com/science/article/pii/S254266052200107X. Acesso em: 05 ago. 2025.

ECKHARDT, C. M.; MADJAROVA, S. J.; WILLIAMS, R. J.; OLLIVIER, M.; KARLSSON, J.; PAREEK, A.; NWACHUKWU, B. U. Unsupervised machine learning methods and emerging applications in healthcare. **Knee Surgery, Sports Traumatology, Arthroscopy**, v. 31, n. 2, p. 376–381, fev. 2023. ISSN 1433-7347. Disponível em: https://doi.org/10.1007/s00167-022-07233-7. Acesso em: 10 jun. 2025.

ELRAWY, M. F.; AWAD, A. I.; HAMED, H. Intrusion detection systems for iot-based smart environments: a survey. **Journal of Cloud Computing Advances Systems and Applications**, v. 7, dez. 2018. Disponível em: https://doi.org/10.1186/s13677-018-0123-6. Acesso em: 07 jun. 2025.

FABRIS, F.; MAGALHãES, J. P. de; FREITAS, A. A. A review of supervised machine learning applied to ageing research. **Biogerontology**, v. 18, n. 2, p. 171–188, abr. 2017. ISSN 1573-6768. Disponível em: https://doi.org/10.1007/s10522-017-9683-y. Acesso em: 10 jun. 2025.

FAROOQ, U.; TARIQ, N.; ASIM, M.; BAKER, T.; AL-SHAMMA'A, A. Machine learning and the internet of things security: Solutions and open challenges. **Journal of Parallel**

and Distributed Computing, v. 162, p. 89–104, 2022. ISSN 0743-7315. Disponível em: https://www.sciencedirect.com/science/article/pii/S0743731522000235. Acesso em: 07 jun. 2025.

FRIHA, O.; FERRAG, M. A.; SHU, L.; MAGLARAS, L.; CHOO, K.-K. R.; NAFAA, M. Felids: Federated learning-based intrusion detection system for agricultural internet of things. **Journal of Parallel and Distributed Computing**, v. 165, p. 17–31, 2022. ISSN 0743-7315. Disponível em: https://www.sciencedirect.com/science/article/pii/S0743731522000570. Acesso em: 20 jun. 2025.

HEIDARI, A.; JAMALI, M. J. Internet of things intrusion detection systems: a comprehensive review and future directions. **Cluster Computing**, v. 26, p. 3753–3780, out. 2022. Disponível em: https://doi.org/10.1007/s10586-022-03776-z. Acesso em: 07 jun. 2025.

KAUR, B.; DADKHAH, S.; SHOELEH, F.; NETO, E. C. P.; XIONG, P.; IQBAL, S.; LAMONTAGNE, P.; RAY, S.; GHORBANI, A. A. Internet of things (iot) security dataset evolution: Challenges and future directions. **Internet of Things**, v. 22, p. 100780, 2023. ISSN 2542-6605. Disponível em: https://www.sciencedirect.com/science/article/pii/S2542660523001038. Acesso em: 10 ago. 2025.

KHANDO, K.; GAO, S.; ISLAM, S. M.; SALMAN, A. Enhancing employees information security awareness in private and public organisations: A systematic literature review. **Computers Security**, v. 106, p. 102267, 2021. ISSN 0167-4048. Disponível em: https://www.sciencedirect.com/science/article/pii/S0167404821000912. Acesso em: 04 ago. 2025.

KUNCHEVA, L. I. **Combining Pattern Classifiers**: Methods and algorithms. Hoboken, NJ: John Wiley & Sons, 2004. ISBN 9780471210798. Disponível em: https://doi.org/10.1002/0471660264. Acesso em: 10 jun. 2025.

LI, L.; FAN, Y.; TSE, M.; LIN, K.-Y. A review of applications in federated learning. **Computers Industrial Engineering**, v. 149, p. 106854, 2020. ISSN 0360-8352. Disponível em: https://www.sciencedirect.com/science/article/pii/S0360835220305532. Acesso em: 10 jun. 2025.

LI, Q.; DIAO, Y.; CHEN, Q.; HE, B. Federated learning on non-iid data silos: An experimental study. In: INTERNATIONAL CONFERENCE ON DATA ENGINEERING (ICDE), 38., 2022. **Proceedings [...]**. [S. l.]: IEEE, 2022. p. 965–978. Disponível em: https://doi.org/10.1109/ICDE53745.2022.00098. Acesso em: 07 jun. 2025.

LIAO, H.-J.; Richard Lin, C.-H.; LIN, Y.-C.; TUNG, K.-Y. Intrusion detection system: A comprehensive review. **Journal of Network and Computer Applications**, v. 36, n. 1, p. 16–24, 2013. ISSN 1084-8045. Disponível em: https://www.sciencedirect.com/science/article/pii/S1084804512001944. Acesso em: 20 ago. 2025.

LITOUSSI, M.; KANNOUF, N.; El Makkaoui, K.; EZZATI, A.; FARTITCHOU, M. Iot security: challenges and countermeasures. **Procedia Computer Science**, v. 177, p. 503–508, 2020. ISSN 1877-0509. The 11th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN 2020) / The 10th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare (ICTH 2020) / Affiliated Workshops. Disponível em: https://www.sciencedirect.com/science/article/pii/S1877050920323395. Acesso em: 20 ago. 2025.

LIU, Y.; KANG, Y.; XING, C.; CHEN, T.; YANG, Q. A secure federated transfer learning framework. **IEEE Intelligent Systems**, v. 35, n. 4, p. 70–82, 2020. Disponível em: https://doi.org/10.1109/MIS.2020.2988525. Acesso em: 11 jun. 2025.

MOTHUKURI, V.; KHARE, P.; PARIZI, R. M.; POURIYEH, S.; DEHGHANTANHA, A.; SRIVASTAVA, G. Federated-learning-based anomaly detection for iot security attacks. **IEEE Internet of Things Journal**, v. 9, n. 4, p. 2545–2554, 2022. Disponível em: https://doi.org/10.1109/JIOT.2021.3077803. Acesso em: 11 jun. 2025.

NAIDU, G.; ZUVA, T.; SIBANDA, E. M. A review of evaluation metrics in machine learning algorithms. In: SILHAVY, R.; SILHAVY, P. (Ed.). **Artificial Intelligence Application in Networks and Systems**. Cham: Springer International Publishing, 2023. p. 15–25. ISBN 978-3-031-35314-7. Disponível em: https://doi.org/10.1007/978-3-031-35314-7_2. Acesso em: 08 jun. 2025.

NAQA, I. E.; MURPHY, M. J. What is machine learning? In: NAQA, I. E.; LI, R.; MURPHY, M. J. (Ed.). **Machine Learning in Radiation Oncology: Theory and Applications**. Cham: Springer International Publishing, 2015. p. 3–11. ISBN 978-3-319-18305-3. Disponível em: https://doi.org/10.1007/978-3-319-18305-3_1. Acesso em: 10 jun. 2025.

NETO, E. C. P.; DADKHAH, S.; FERREIRA, R.; ZOHOURIAN, A.; LU, R.; GHORBANI, A. A. CICIoT2023: A real-time dataset and benchmark for large-scale attacks in iot environment. **Sensors**, v. 23, n. 7, p. 3538, 2023. Disponível em: https://doi.org/10.3390/s23073538. Acesso em: 07 jun. 2025.

NGUYEN, D. C.; DING, M.; PATHIRANA, P. N.; SENEVIRATNE, A.; LI, J.; NIYATO, D.; DOBRE, O.; POOR, H. V. 6g internet of things: A comprehensive survey. **IEEE Internet of Things Journal**, v. 9, n. 1, p. 359–383, 2022. Disponível em: https://doi.org/10.1109/JIOT.2021.3103320. Acesso em: 05 ago. 2025.

PANTANOWITZ, L.; PEARCE, T.; ABUKHIRAN, I.; HANNA, M.; WHEELER, S.; SOONG, T. R.; TAFTI, A. P.; PANTANOWITZ, J.; LU, M. Y.; MAHMOOD, F.; GU, Q.; RASHIDI, H. H. Nongenerative artificial intelligence in medicine: Advancements and applications in supervised and unsupervised machine learning. **Modern Pathology**, v. 38, n. 3, p. 100680, 2025. ISSN 0893-3952. Disponível em: https://www.sciencedirect.com/science/article/pii/S0893395224002606. Acesso em: 07 jun. 2025.

RAINIO, O.; TEUHO, J.; KLéN, R. Evaluation metrics and statistical tests for machine learning. **Scientific Reports**, v. 14, n. 1, p. 6086, 2024. ISSN 2045-2322. Disponível em: https://doi.org/10.1038/s41598-024-56706-x. Acesso em: 10 jun. 2025.

SAHEED, Y. K.; ABIODUN, A. I.; MISRA, S.; HOLONE, M. K.; COLOMO-PALACIOS, R. A machine learning-based intrusion detection for detecting internet of things network attacks. **Alexandria Engineering Journal**, v. 61, n. 12, p. 9395–9409, 2022. ISSN 1110-0168. Disponível em: https://www.sciencedirect.com/science/article/pii/S1110016822001570. Acesso em: 20 jun. 2025.

SARKER, I. H. Machine learning: Algorithms, real-world applications and research directions. **SN Computer Science**, v. 2, n. 3, p. 160, mar. 2021. ISSN 2661-8907. Disponível em: https://doi.org/10.1007/s42979-021-00592-x. Acesso em: 07 jun. 2025.

SARKER, I. H.; FURHAD, M. H.; NOWROZY, R. Ai-driven cybersecurity: An overview, security intelligence modeling and research directions. **SN Computer Science**, v. 2, n. 3, p. 173, mar. 2021. ISSN 2661-8907. Disponível em: https://doi.org/10.1007/s42979-021-00557-0. Acesso em: 07 jun. 2025.

SARKER, I. H.; HOQUE, M. M.; UDDIN, M. K.; ALSANOOSY, T. Mobile data science and intelligent apps: Concepts, ai-based modeling and research directions. **Mobile Networks and Applications**, v. 26, n. 1, p. 285–303, 2021. ISSN 1572-8153. Disponível em: https://doi.org/10.1007/s11036-020-01650-z. Acesso em: 07 jun. 2025.

SATTLER, F.; WIEDEMANN, S.; MÜLLER, K.-R.; SAMEK, W. Robust and communication-efficient federated learning from non-iid data. **IEEE transactions on neural networks and learning systems**, IEEE, v. 31, n. 9, p. 3400–3413, 2019. Disponível em: https://doi.org/10.1109/TNNLS.2019.2944481. Acesso em: 05 jun. 2025.

SCHILLER, E.; AIDOO, A.; FUHRER, J.; STAHL, J.; ZIÖRJEN, M.; STILLER, B. Landscape of iot security. **Computer Science Review**, Elsevier, v. 44, p. 100467, 2022. Disponível em: https://doi.org/10.1016/j.cosrev.2022.100467. Acesso em: 10 ago. 2025.

SHARMA, S.; VERMA, V. K. Security explorations for routing attacks in low power networks on internet of things. **The Journal of Supercomputing**, Springer, v. 77, p. 4778–4812, 2021. Disponível em: https://doi.org/10.1007/s11227-020-03471-z. Acesso em: 20 ago. 2025.

SHAUKAT, K.; ALAM, T. M.; HAMEED, I. A.; KHAN, W. A.; ABBAS, N.; LUO, S. A review on security challenges in internet of things (iot). In: INTERNATIONAL CONFERENCE ON AUTOMATION AND COMPUTING (ICAC), 26., 2021. **Proceedings [...]**. [S. l.]: [S. n.], 2021. p. 1–6. Disponível em: https://doi.org/10.23919/ICAC50006.2021.9594183. Acesso em: 20 ago. 2025.

SHUKLA, P.; CHALLA, R.; PATIL, N. Iot traffic-based ddos attacks detection mechanisms: A comprehensive review. **The Journal of Supercomputing**, v. 80, p. 1–58, dez. 2023. Disponível em: https://doi.org/10.1007/s11227-023-05843-7. Acesso em: 10 ago. 2025.

SILVA, C.; OLIVEIRA, C.; ANDRADE, R. Maria: Monitoramento e análise para resposta imediata a ataques à rede 5g no contexto da iot. In: SIMPÓSIO BRASILEIRO DE REDES DE COMPUTADORES E SISTEMAS DISTRIBUÍDOS (SBRC), 43., 2025, Natal. **Anais [...]**. Porto Alegre: SBC, 2025. p. 364–377. Disponível em: https://sol.sbc.org.br/index.php/sbrc/article/view/35144. Acesso em: 01 ago. 2025.

SOBIN, C. C. A survey on architecture, protocols and challenges in iot. **Wirel. Pers. Commun.**, Kluwer Academic Publishers, USA, v. 112, n. 3, p. 1383–1429, jun. 2020. ISSN 0929-6212. Disponível em: https://doi.org/10.1007/s11277-020-07108-5. Acesso em: 05 jun. 2025.

SOORI, M.; AREZOO, B.; DASTRES, R. Internet of things for smart factories in industry 4.0, a review. **Internet of Things and Cyber-Physical Systems**, v. 3, p. 192–204, 2023. ISSN 2667-3452. Disponível em: https://www.sciencedirect.com/science/article/pii/S2667345223000275. Acesso em: 02 ago. 2025.

TEKIN, N.; ACAR, A.; ARIS, A.; ULUAGAC, A. S.; GUNGOR, V. C. Energy consumption of on-device machine learning models for iot intrusion detection. **Internet of Things**, v. 21, p. 100670, 2023. ISSN 2542-6605. Disponível em: https://www.sciencedirect.com/science/article/pii/S2542660522001512. Acesso em: 20 ago. 2025.

TOURNIER, J.; LESUEUR, F.; MOUËL, F. L.; GUYON, L.; BEN-HASSINE, H. A survey of iot protocols and their security issues through the lens of a generic iot stack. **Internet of Things**, v. 16, p. 100264, 2021. ISSN 2542-6605. Disponível em: https://www.sciencedirect.com/science/article/pii/S2542660520300986. Acesso em: 01 ago. 2025.

USAMA, M.; QADIR, J.; RAZA, A.; ARIF, H.; YAU, K.-l. A.; ELKHATIB, Y.; HUSSAIN, A.; AL-FUQAHA, A. Unsupervised machine learning for networking: Techniques, applications and research challenges. **IEEE Access**, v. 7, p. 65579–65615, 2019. Disponível em: https://doi.org/10.1109/ACCESS.2019.2916648. Acesso em: 10 jun. 2025.

WEN, J.; ZHANG, Z.; LAN, Y.; CUI, Z.; CAI, J.; ZHANG, W. A survey on federated learning: challenges and applications. **International Journal of Machine Learning and Cybernetics**, Springer, v. 14, n. 2, p. 513–535, 2023. Disponível em: https://link.springer.com/article/10.1007/S13042-022-01647-Y. Acesso em: 20 ago. 2025.

ZHANG, C.; XIE, Y.; BAI, H.; YU, B.; LI, W.; GAO, Y. A survey on federated learning. **Knowledge-Based Systems**, v. 216, p. 106775, 2021. ISSN 0950-7051. Disponível em: https://www.sciencedirect.com/science/article/pii/S0950705121000381. Acesso em: 10 jun. 2025.