PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ

WELLINGTON PATRICK DE LIMA

TÉCNICA DE BAIXA COMPLEXIDADE PARA IDENTIFICAÇÃO DE DISPOSITIVOS LORA USANDO ASSINATURA DE RADIOFREQUÊNCIA

MESTRADO EM
CIÊNCIA DA COMPUTAÇÃO
PUCPR

CURITIBA
2026

WELLINGTON PATRICK DE LIMA

Técnica de Baixa Complexidade para Identificação de Dispositivos LoRa usando Assinatura de Radiofrequência

apresentada ao Programa de Pós-Graduação em Informática da Pontifícia Universidade Católica do Paraná como requisito parcial para obtenção do título de mestre em Informática.

Pontifícia Universidade Católica do Paraná - PUCPR

Programa de Pós-Graduação em Informática - PPGIa

Orientador: MARCELO EDUARDO PELLENZ

Curitiba - PR, Brasil

2026

Dados da Catalogação na Publicação
Pontifícia Universidade Católica do Paraná
Sistema Integrado de Bibliotecas – SIBI/PUCPR
Biblioteca Central

L732t
2026

Lima, Wellington Patrick de
Técnica de baixa complexidade para identificação de dispositivos LoRa usando assinatura de radiofrequência / Wellington Patrick de Lima : orientador: Marcelo Eduardo Pellenz – 2026.
74 f. : 30 cm

Dissertação (mestrado) – Pontifícia Universidade Católica do Paraná, Curitiba, 2026
Bibliografia: f. 68-72

1.Informática. 2. Comunicação. 3. Radiofrequência. 4. Internet das coisas. I. Pellenz, Marcelo Eduardo. II. Pontifícia Universidade Católica do Paraná. Programa de Pós-Graduação em Teologia. III. Título.
.

CDD 22. ed. – 004

Cesar Almeida Filho – CRB 9/1908

Pontifícia Universidade Católica do Paraná
Escola Politécnica
Programa de Pós-Graduação em Informática

Curitiba, 18 de março de 2026.

14-2026

DECLARAÇÃO

Declaro para os devidos fins, que **WELLINGTON PATRICK DE LIMA** defendeu a dissertação de Mestrado intitulada "**Técnica de Baixa Complexidade para Identificação de Dispositivos LoRa usando Assinatura de Radiofrequência"**, na área de concentração Ciência da Computação no dia 10 de dezembro de 2025, no qual foi aprovado.

Declaro ainda, que foram feitas todas as alterações solicitadas pela Banca Examinadora, cumprindo todas as normas de formatação definidas pelo Programa.

Por ser verdade firmo a presente declaração.

gov.br Documento assinado digitalmente
JEAN PAUL BARDDAL
Data: 01/04/2026 14:11:01-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. Jean Paul Barddal
Coordenador do Programa de Pós-Graduação em Informática

Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Agradeço, em primeiro lugar, a Deus, que me sustentou nessa jornada. Também agradeço à minha família, que me apoiou e incentivou nesse período, assim como à minha namorada, que me deu suporte e inspiração nessa etapa final. Agradeço à PUCPR por essa oportunidade e pelo apoio recebido ao longo do curso. Da mesma forma, agradeço ao meu orientador, que me guiou nos momentos em que estive sem direção, e, por fim, ao departamento do curso, que teve papel importantíssimo na coordenação dos alunos.

Este trabalho é dedicado a Deus, minha família, namorada e amigos que sempre compreenderam meus sonhos e caminhos e me apoiaram nos momentos mais difíceis

"Confia ao Senhor as tuas obras,
e os teus pensamentos
serão estabelecidos".
Provérbios 16:3

Resumo

Os sistemas de Internet das Coisas (IoT) vêm se expandindo rapidamente, integrando milhões de dispositivos conectados em aplicações que vão desde ambientes domésticos até infraestruturas críticas. Nesse cenário, a identificação correta de dispositivos torna-se fundamental para aumentar a segurança, pois permite detectar nós não autorizados e reforçar mecanismos de autenticação na camada física, sendo especialmente relevante em aplicações como cidades inteligentes e sistemas de automação industrial. Este trabalho propõe uma abordagem de baixa complexidade computacional para identificação de dispositivos LoRa em redes IoT, utilizando características temporais extraídas das amostras IQ do sinal recebido em banda base. A proposta busca substituir métodos mais complexos baseados em *deep learning*, oferecendo uma alternativa mais eficiente para implementação em dispositivos embarcados com recursos limitados. A metodologia emprega a biblioteca Catch22, que fornece um conjunto de métricas estatísticas para caracterização de séries temporais. O método proposto utiliza dados extraídos de segmentos curtos do preâmbulo dos quadros LoRa, que após a extração das características são processados por algoritmos de aprendizagem de máquina para construção de modelos de *fingerprinting* de radiofrequência (RF) capazes de distinguir entre diferentes transmissores. Os resultados atingiram acurácias de aproximadamente 87%, com complexidade computacional significativamente menor que as abordagens com aprendizagem profunda, o que torna a solução especialmente adequada para aplicações em dispositivos IoT de baixo consumo.

Palavras-chave: Rádio LoRa, Assinatura de Radiofrequência, Aprendizagem de Máquina, Catch22, Internet das Coisas, Segurança em IoT.

Abstract

The Internet of Things (IoT) ecosystem has been rapidly expanding, integrating millions of connected devices across applications ranging from domestic environments to critical infrastructure systems. In this scenario, the correct identification of devices becomes fundamental for enhancing security, as it enables the detection of unauthorized nodes and strengthens physical-layer authentication mechanisms, which is particularly relevant in applications such as smart cities and industrial automation systems. This work proposes a low–computational complexity approach for identifying LoRa devices in IoT networks, using temporal features extracted from IQ samples of the received baseband signal. The proposed method seeks to replace more complex *deep learning*-based solutions by offering a more efficient alternative suitable for implementation on resource-constrained embedded devices. The methodology employs the Catch22 library, which provides a set of statistical metrics for characterizing time series. Data extracted from short segments of the LoRa preamble are processed by machine learning algorithms to build radio-frequency (*RF*) fingerprinting models capable of distinguishing different transmitters. The results achieved accuracies of approximately 87%, with computational complexity significantly lower than deep learning approaches, making the solution particularly suitable for low-power IoT applications.

Keywords: LoRa Radio, Radio-Frequency Fingerprinting, Machine Learning, Catch22, Internet of Things, IoT Security.

Lista de ilustrações

Lista de tabelas

Lista de abreviaturas e siglas

IoT	Internet das Coisas, do inglês *Internet of Things*
LoRa	Modulação de rádio de longo alcance, do inglês *Long Range*
IQ	Amostras em quadratura, do inglês *In-phase and Quadrature*
Catch22	Conjunto de 22 características canônicas de séries temporais
RF	Radiofrequência
ISM	Banda Industrial, Científica e Médica, do inglês *Industrial, Scientific and Medical*
MHz	Megahertz
LoRaWAN	Rede de longo alcance de baixa potência, do inglês *Long Range Wide-Area Network*
RFFI	Identificação por Impressão Digital de RF, do inglês *Radio Frequency Fingerprinting Identification*
CNN	Rede Neural Convolucional, do inglês *Convolutional Neural Network*
MIMO	Múltiplas Entradas e Múltiplas Saídas, do inglês *Multiple Input Multiple Output*
Wi-Fi	Tecnologia de redes sem fio baseada no padrão IEEE 802.11
ZigBee	Protocolo de comunicação baseado em IEEE 802.15.4
SigFox	Tecnologia LPWAN proprietária para IoT
LPWAN	Rede de Longo Alcance e Baixa Potência, do inglês *Low-Power Wide-Area Network*
DoS	Ataque de Negação de Serviço, do inglês *Denial of Service*
ML	Aprendizagem de Máquina, do inglês *Machine Learning*
RFID	Identificação por Radiofrequência, do inglês *Radio Frequency Identification*
WSN	Rede de Sensores Sem Fio, do inglês *Wireless Sensor Network*
GPS	Sistema de Posicionamento Global, do inglês *Global Positioning System*

3G	Terceira Geração de redes móveis
4G	Quarta Geração de redes móveis
5G	Quinta Geração de redes móveis
6G	Sexta Geração de redes móveis
HSTRN	Rede Híbrida Satélite–Terrestre, do inglês *Hybrid Satellite-Terrestrial Network*
NOMA	Acesso Múltiplo Não Ortogonal, do inglês *Non-Orthogonal Multiple Access*
RSMA	Acesso Múltiplo por Divisão de Taxa, do inglês *Rate-Splitting Multiple Access*
RIS	Superfície Inteligente Reconfigurável, do inglês *Reconfigurable Intelligent Surface*
VANT	Veículo Aéreo Não Tripulado
CoAP	Protocolo de Aplicação Constrained, do inglês *Constrained Application Protocol*
MQTT	Protocolo de Telemetria de Fila de Mensagens, do inglês *Message Queue Telemetry Transport*
AMQP	Protocolo Avançado de Enfileiramento de Mensagens, do inglês *Advanced Message Queuing Protocol*
HTTP	Protocolo de Transferência de Hipertexto, do inglês *HyperText Transfer Protocol*
BLE	Bluetooth de Baixa Energia, do inglês *Bluetooth Low Energy*
IEEE	Instituto de Engenheiros Eletricistas e Eletrônicos, do inglês *Institute of Electrical and Electronics Engineers*
NB-IoT	Internet das Coisas Banda Estreita, do inglês *NarrowBand IoT*
Wi-SUN	Rede Sem Fio Inteligente, do inglês *Wireless Smart Utility Network*
Z-Wave	Protocolo de automação residencial de baixa potência
IPv4	Protocolo de Internet versão 4
IPv6	Protocolo de Internet versão 6

PLC	Comunicação por Linha de Energia, do inglês *Power Line Communication*
NFC	Comunicação por Campo de Proximidade, do inglês *Near-Field Communication*
UWB	Banda Ultra Larga, do inglês *Ultra-Wideband*
MiWi	Protocolo sem fio baseado em IEEE 802.15.4
Li-Fi	Comunicação por Luz Visível, do inglês *Light Fidelity*
SDN	Rede Definida por Software, do inglês *Software Defined Network*
DDS	Serviço de Distribuição de Dados, do inglês *Data Distribution Service*
UDP	Protocolo de Datagrama de Usuário, do inglês *User Datagram Protocol*
XMPP	Protocolo Extensível de Mensagens e Presença, do inglês *Extensible Messaging and Presence Protocol*
QoS	Qualidade de Serviço, do inglês *Quality of Service*
GHz	Gigahertz
GSM	Sistema Global para Comunicações Móveis, do inglês *Global System for Mobile Communications*
LTE	Long Term Evolution
LTE-M	LTE para IoT Máquina-a-Máquina
3GPP	Parceria de Padronização de 3ª Geração, do inglês *3rd Generation Partnership Project*
SIG	Grupo de Interesse Especial, do inglês *Special Interest Group*
ISO	Organização Internacional de Padronização, do inglês *International Organization for Standardization*
IEC	Comissão Eletrotécnica Internacional, do inglês *International Electrotechnical Commission*
WPAN	Rede Pessoal Sem Fio, do inglês *Wireless Personal Area Network*
LR-WPAN	Rede Pessoal Sem Fio de Baixa Taxa, do inglês *Low-Rate Wireless Personal Area Network*
6LoWPAN	IPv6 sobre WPAN de baixa taxa

kb	Quilobit
km	Quilômetro
FSK	Modulação por Chaveamento de Frequência, do inglês *Frequency-Shift Keying*
TP	Potência de Transmissão, do inglês *Transmission Power*
SF	Fator de Espalhamento, do inglês *Spreading Factor*
BW	Largura de Banda, do inglês *Bandwidth*
CR	Taxa de Codificação, do inglês *Coding Rate*
CF	Frequência Portadora, do inglês *Carrier Frequency*
kHz	Quilohertz
ANATEL	Agência Nacional de Telecomunicações
DR	Taxa de Dados, do inglês *Data Rate*
API	Interface de Programação de Aplicações, do inglês *Application Programming Interface*
MITM	Ataque Man-in-the-Middle
ADS-B	Sistema Automático de Vigilância Dependente, do inglês *Automatic Dependent Surveillance–Broadcast*
RNA	Rede Neural Artificial
DT	Árvore de Decisão, do inglês *Decision Tree*
RF (modelo)	Floresta Aleatória, do inglês *Random Forest*
GBM	Máquina de Gradiente Boosting, do inglês *Gradient Boosting Machine*
CART	Árvores de Classificação e Regressão, do inglês *Classification and Regression Trees*
GOSS	Amostragem por Gradiente Útil, do inglês *Gradient-Based One-Side Sampling*
EFB	Estrutura de Características Exclusivas, do inglês *Exclusive Feature Bundling*
FFT	Transformada Rápida de Fourier, do inglês *Fast Fourier Transform*

CFO	Desvio de Frequência de Portadora, do inglês *Carrier Frequency Offset*
RiFyFi_VDG	Biblioteca Virtual de Fingerprinting de RF
SISO	Sistema de Entrada Única e Saída Única, do inglês *Single Input Single Output*
CR-FP	*Channel-Resilient Fingerprinting*
RSSI	Indicador de Intensidade do Sinal Recebido, do inglês *Received Signal Strength Indicator*
SNR	Relação Sinal-Ruído, do inglês *Signal-to-Noise Ratio*
SDR	Rádio Definido por Software, do inglês *Software-Defined Radio*
USRP B210	Hardware SDR da Ettus Research
MS/s	Megamostras por segundo
SigMF	Formato de Metadados de Sinal, do inglês *Signal Metadata Format*
HCTSA	Catálogo de Séries Temporais Altamente Comparáveis, do inglês *Highly Comparative Time-Series Analysis*

Sumário

1 Introdução

A Internet das Coisas (Internet of Things - IoT) surgiu com o conceito de aumentar a conectividade e integração de dispositivos em redes, aprimorando a relação entre a conectividade dos mundos físico e virtual. Suas aplicações abrangem diversos domínios, incluindo cidades inteligentes (ZANELLA et al., 2014), sistemas industriais (MUMTAZ et al., 2017), agricultura de precisão (ARIF et al., 2024) e automação residencial (DOTSON; CENEK; MICHAELSON, 2019). Ao otimizar processos, aumentar a eficiência e reforçar a segurança, a IoT busca agilizar operações cotidianas, ao mesmo tempo em que reduz custos. No contexto de cidades inteligentes (ZANELLA et al., 2014), a IoT possibilita o uso eficiente de recursos públicos, eleva a qualidade dos serviços aos cidadãos e diminui os custos operacionais das administrações públicas, por meio de uma infraestrutura de comunicação urbana dedicada que oferece acesso unificado, econômico e escalável a uma ampla gama de serviços públicos.

Tecnologias de comunicação sem fio desempenham papel crucial na integração de sistemas e na adaptação a requisitos operacionais diversos. Entretanto, muitas aplicações enfrentam desafios relativos ao alcance do sinal, à capacidade de processamento, à estabilidade de conexão e à segurança da rede (LOUNIS; ZULKERNINE, 2020). Para superar essas limitações, sistemas embarcados aliam microprocessadores e microcontroladores, permitindo o desenvolvimento de soluções cada vez mais sofisticadas e eficientes. Como resultado, a Internet das Coisas permite que objetos físicos monitorem, processem e atuem, viabilizando comunicação contínua, troca de informações e tomada de decisão de forma autônoma (AL-FUQAHA et al., 2015).

Em redes IoT, os dispositivos normalmente se comunicam via Wi-Fi ou por outras tecnologias sem fio, como, por exemplo, Zigbee, LoRa, SigFox e Bluetooth, entre outras (DING et al., 2020). Muitas aplicações de IoT dependem de redes sem fio de baixo consumo e ampla cobertura, comumente denominadas *Low-Power Wide Area Networks* (LPWANs) (IEEE, 2020). Essas tecnologias têm sido cada vez mais adotadas para conectividade em larga escala e a baixo custo, permitindo comunicação eficiente entre dispositivos distribuídos. Suas principais vantagens incluem consumo reduzido de energia e alcance estendido, tornando-as ideais para aplicações em que eficiência energética e cobertura são mais importantes do que altas taxas de dados e baixa latência.

A tecnologia de rádio *Long Range* (LoRa) (LORA® ALLIANCE, 2015) consolidou-se como um elemento-chave nas implementações de LPWAN, oferecendo comunicação de longo alcance robusta por meio de modulação por espalhamento espectral usando a técnica *chirp*. Essa tecnologia suporta a transmissão de dados por distâncias superiores a

10 km em áreas rurais e alcance significativo em ambientes urbanos, mesmo na presença de interferências. Com boa eficiência energética, devido ao baixo consumo de potência, o rádio LoRa pode operar por bateria durante longos períodos de tempo, sendo ideal para aplicações de IoT como monitoramento remoto, cidades inteligentes, agricultura de precisão e rastreamento de ativos (SILVA et al., 2017). Além disso, integra-se de forma ideal ao protocolo LoRaWAN, que define a arquitetura de rede, gerencia dispositivos conectados e assegura comunicação eficiente em implantações de larga escala.

Este trabalho está organizado da seguinte forma. O Capítulo 2 apresenta a fundamentação teórica necessária para a compreensão dos conceitos de IoT, protocolo LoRa e as bases do *Radio Frequency Fingerprinting*. O Capítulo 3 detalha o estado da arte e os trabalhos relacionados, com ênfase no rastreamento e monitoramento de dispositivos usando RFFI e técnicas de aprendizado de máquina. Já o Capítulo 4 descreve a metodologia proposta, incluindo o detalhamento do conjunto de dados utilizado, o processo de extração de características com a biblioteca *Catch22* e a configuração dos modelos de classificação. No Capítulo 5, são apresentados e discutidos os resultados experimentais obtidos, incluindo a análise de complexidade computacional. Por fim, o Capítulo 6 apresenta as considerações finais e sugestões para trabalhos futuros.

1.1 Motivação

Com a rápida expansão da IoT, garantir a segurança dos dispositivos conectados tornou-se um desafio crítico. As redes IoT são suscetíveis a diversas ameaças e ataques, incluindo configurações incorretas, intrusões, perda de sinal, negação de serviço (DoS), ataques *man-in-the-middle*, interrupção e interceptação de dados (LIU et al., 2016). Para mitigar esses riscos, desenvolveram-se métodos de identificação única de dispositivos. Uma abordagem promissora é a identificação de dispositivos por assinatura de radiofrequência (Radio Frequency Fingerprint Identification - RFFI), que tenta classificar de forma única dispositivos sem fio ao analisar distorções no sinal recebido causadas por imperfeições inerentes ao hardware de cada dispositivo (SHEN et al., 2021).

Por meio de métodos de Aprendizado de Máquina (Machine Learning - ML), é possível integrar a arquitetura de rede LoRa a algoritmos de *fingerprinting* capazes de identificar dispositivos conhecidos, elevando significativamente a segurança e o controle dos serviços de transmissão. Diversas técnicas de RFFI baseadas em ML foram propostas em (ELMAGHBUB; HAMDAOUI, 2021), (GASKIN et al., 2023), (BASHA et al., 2023) e (SHEN et al., 2021). Contudo, essas abordagens requerem longas sequências de amostras do sinal em banda base, nas componentes em fase e em quadratura (In-phase and Quadrature – IQ), para realização da análise. Isso acarreta alta complexidade computacional e limita sua aplicação prática a níveis mais robustos da rede, como concentradores ou *gateways*.

1.2 Objetivos

1.2.1 Objetivo Geral

Neste trabalho, propomos um método de baixa complexidade para identificação de dispositivos LoRa, baseado exclusivamente na análise das amostras IQ do preâmbulo do quadro (frame) recebido. O método emprega extração de características de séries temporais, utilizando especificamente o conjunto de características *CAnonical Time-series CHaracteristics* (*Catch22*) (LUBBA et al., 2019), em combinação com técnicas de ML, para realizar *fingerprinting* de RF. Para avaliar e validar a abordagem, utilizamos o conjunto de dados disponibilizado pela Oregon University, conforme descrito em (ELMAGHBUB; HAMDAOUI, 2021).

1.2.2 Questões de Pesquisa

Este trabalho busca responder às seguintes questões de pesquisa:

- **Q1**: Como identificar dispositivos LoRa de forma eficiente e precisa, considerando as restrições computacionais em sistemas IoT?
- **Q2**: De que maneira métricas analíticas, como as extraídas pelo *Catch22*, podem constituir alternativa viável às abordagens baseadas em CNN, que demandam elevado uso de recursos computacionais?
- **Q3**: Quais melhorias de segurança e monitoramento podem ser alcançadas com a abordagem proposta para identificação de dispositivos em redes IoT?

1.2.3 Objetivos Específicos

Para atingir o objetivo geral proposto, este trabalho estabelece os seguintes objetivos específicos:

- Avaliar a eficiência das métricas do *Catch22* no processo de extração de características para aplicação de *fingerprints*, as quais foram extraídas diretamente de amostras IQ provenientes de dispositivos LoRa.
- Investigar a influência de diferentes configurações de parâmetros de análise, como o tamanho da janela de amostragem e o número de dias combinados, sobre a acurácia e estabilidade da identificação de dispositivos.
- Comparar o desempenho de diferentes algoritmos de aprendizado de máquina (*Decision Tree*, *Gradient Boosting*, *LightGBM*, *Random Forest* e *XGBoost*) aplicados às

métricas do *Catch22*, de modo a identificar o modelo mais eficiente em termos de acurácia e custo computacional.

- Propor um método de identificação de dispositivos de baixa complexidade, baseado na análise estatística do sinal no domínio do tempo, que mantenha desempenho competitivo frente a abordagens baseadas em Redes Neurais Convolucionais (Convolutional Neural Network - CNN).

1.3 Publicações

Os resultados parciais desta pesquisa foram publicados em:

- LIMA, Wellington P. de; PELLENZ, Marcelo E.; TEIXEIRA, Marco A. S.; ALBERTI, Antonio M. **LoRa Device Identification: A Lightweight Alternative to CNN-Based Methods**. In: *International Conference on Artificial Intelligence and Soft Computing (ICAISC 2025)*, Zakopane, Poland, 2025. Lecture Notes in Artificial Intelligence (LNAI), Springer, 2025.

2 Fundamentação Teórica

Antes de explorar os aspectos específicos deste trabalho, é fundamental compreender os princípios que sustentam a proposta. Assim, este capítulo apresenta os conceitos básicos relacionados à Internet das Coisas, abordando os principais protocolos de comunicação, as tecnologias utilizadas e os desafios de segurança envolvidos. Também são discutidos os fundamentos do processo de *fingerprinting*, que possibilita a identificação única de dispositivos com base em suas características físicas e de transmissão, bem como as principais técnicas de aprendizado de máquina empregadas neste estudo para aumentar a precisão e a confiabilidade dos mecanismos de autenticação e detecção em sistemas IoT.

2.1 Internet das Coisas

De forma geral, é notável como a tecnologia e a internet estão cada vez mais presentes no cotidiano das pessoas, tornando suas rotinas cada vez mais dependentes desse tipo de acesso e comunicação. Nesse contexto, a Internet das Coisas tem impulsionado o desenvolvimento de soluções inovadoras em diversas áreas, como *smart homes, smart cities, smart grids*, agricultura de precisão e Indústria 5.0, entre outras. Sua principal característica é a capacidade de integrar diferentes tecnologias em um ecossistema conectado. A IoT permite que objetos físicos coletem dados, monitorem eventos e executem ações, compartilhando informações e contribuindo para processos autônomos de tomada de decisão (AL-FUQAHA et al., 2015).

Com o objetivo de ampliar o acesso à internet e, consequentemente, melhorar a qualidade de vida das pessoas, as *smart homes* têm incorporado tecnologias que vão desde sistemas de monitoramento e segurança até soluções de iluminação inteligente e automação residencial. No contexto das smart cities, um dos principais desafios da IoT está relacionado à conectividade contínua, ao meio de transmissão e ao alcance dos sinais. Apesar desses desafios, trata-se de um cenário altamente promissor, pois pode trazer diversos benefícios para a gestão e otimização de serviços públicos, como transporte, estacionamento, iluminação, vigilância, manutenção de áreas públicas, preservação do patrimônio cultural, coleta de resíduos, entre outros. Além disso, a coleta e a disponibilidade de diferentes tipos de dados permitem ampliar a transparência das ações governamentais, aproximar a administração pública dos cidadãos e estimular o desenvolvimento de novos serviços e ferramentas baseados em IoT (ZANELLA et al., 2014).

Mansour et al. (2023) descreve em seu trabalho as principais tecnologias hoje presentes em sistemas IoT, as quais podem ser analisadas através da distribuição de cinco camadas principais: física, enlace de dados, rede, transporte e aplicação. A camada física,

também denominada “camada de percepção” ou “camada de reconhecimento” no contexto da IoT, tem como função detectar as características físicas dos objetos no ambiente. Ela baseia-se em tecnologias de sensoriamento, como Identificação por Radiofrequência (RFID), Redes de Sensores Sem Fio (WSN) e Sistema de Posicionamento Global (GPS). Além disso, é responsável por converter as informações coletadas em sinais digitais, facilitando sua transmissão pela rede.

A camada de enlace de dados é responsável por delimitar pacotes, realizar a sincronização de quadros, gerenciar endereços de origem e destino, detectar erros no canal físico e prevenir colisões (OLIVEIRA et al., 2019). Cada protocolo associado a essa camada apresenta características específicas de projeto e implementação, incluindo mecanismos de controle de acesso ao meio, taxas de transmissão, topologia de comunicação, alcance de cobertura e consumo energético.

Já a camada de rede fornece os canais de roteamento necessários para o envio e recebimento de dados em formato de pacotes por toda a infraestrutura de comunicação (FAROOQ et al., 2015). Ela engloba equipamentos como *switches* e roteadores, além de tecnologias e protocolos de comunicação de diferentes gerações, como 3G, 4G, 5G, Wi-Fi, ZigBee e infravermelho, que garantem a conectividade e a interoperabilidade entre dispositivos.

A camada de transporte, por sua vez, atua em conjunto com a camada de aplicação para assegurar a entrega correta dos dados, oferecendo funcionalidades como controle de ordem de pacotes, prevenção de congestionamento, multiplexação, integridade e confiabilidade na comunicação (VASHI et al., 2017).

Por fim, a camada de aplicação representa o ponto de interação entre a arquitetura IoT e os usuários finais, sendo responsável por materializar o potencial da tecnologia. Ela oferece plataformas, interfaces e ferramentas que permitem o desenvolvimento de aplicações voltadas a diferentes domínios, como mencionado anteriormente cidades inteligentes, automação residencial, transporte, saúde e indústria conectada. (MANSOUR et al., 2023). A Figura 1 destaca essa visão por camadas.

Quanto ao futuro da IoT podemos dizer que a evolução das comunicações sem fio rumo ao padrão 6G deve viabilizar a cobertura global de redes IoT, integrando tecnologias emergentes como comunicação via satélite, Redes Híbridas de Satélite-Terrestre (Hybrid Satellite–Terrestrial Networks - HSTN), e plataformas aéreas autônomas. Segundo Mansour et al. (2023), a próxima geração de redes buscará garantir conectividade onipresente, latência mínima e baixo consumo energético, apoiando-se em técnicas avançadas de acesso múltiplo, como o Acesso Múltiplo Não Ortogonal (Non-Orthogonal Multiple Access - NOMA) e o Acesso Múltiplo por Divisão de Taxa (Rate-Splitting Multiple Access - RSMA), capazes de lidar com o aumento exponencial de dispositivos conectados. Além disso, o uso de Superfícies Inteligentes Reconfiguráveis (Reconfigurable Intelligent Surfaces

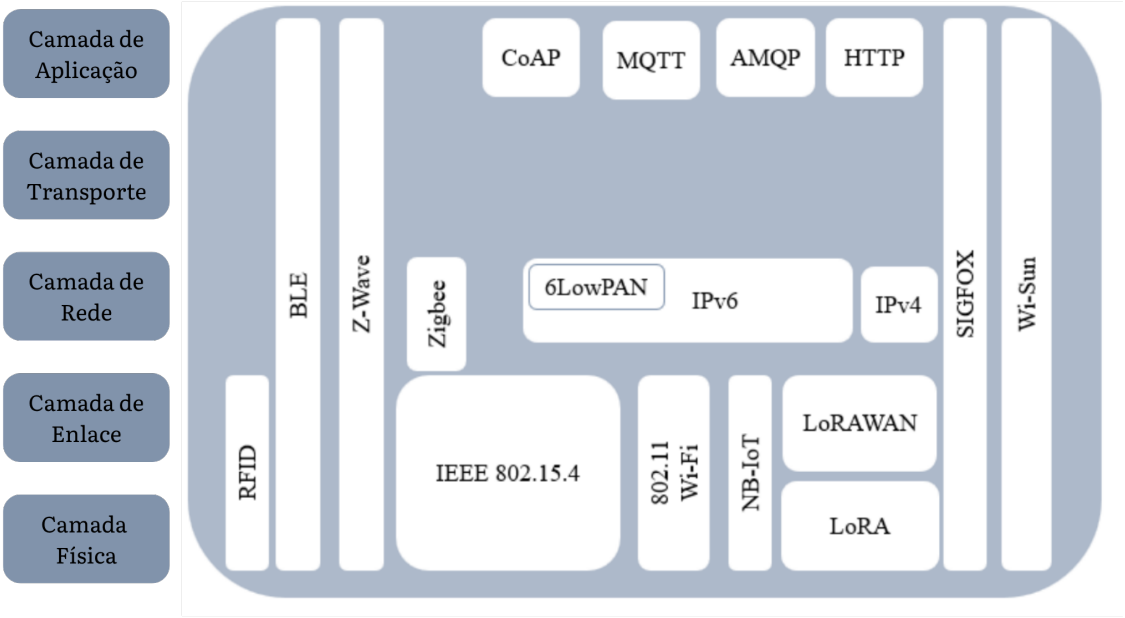


Figura 1 – Tecnologias IoT e suas camadas.

Fonte: adaptado de (MANSOUR et al., 2023).

- RIS) e Veículos Aéreos Não Tripulados (VANTs) é apontado como solução promissora para ampliar a cobertura e otimizar a propagação de sinais em ambientes desafiadores. Essa combinação de tecnologias tem como objetivo superar as limitações estruturais das redes terrestres tradicionais, proporcionando conectividade contínua e escalável para aplicações IoT em cenários urbanos, industriais e remotos.

2.2 Tecnologias aplicadas

A escolha do meio de comunicação depende das restrições de energia e dos requisitos de alcance, taxa de dados, latência, escalabilidade e segurança de cada aplicação. Entre as tecnologias com fio e sem fio empregadas estão a *Power-Line Communication* (PLC) e o X10, que utilizam a rede elétrica para transmissão de dados e controle de dispositivos em ambientes domésticos e industriais. Em curtas distâncias, tecnologias como *Near-Field Communication* (NFC) e *Ultra-Wide Band* (UWB) permitem trocas rápidas e seguras, sendo amplamente usadas em pagamentos inteligentes e rastreamento de ativos. No espectro de redes locais, o Wi-Fi e o padrão IEEE 802.15.4 (base para ZigBee, *Thread*,

WirelessHART e MiWi) são amplamente aplicados em automação residencial e industrial. Outras opções de curto alcance incluem *Bluetooth Low Energy* (BLE) e ANT+, voltadas para dispositivos de baixo consumo, como vestíveis e sensores biomédicos. (HASSAN, 2018).

Para comunicações de longo alcance e baixo consumo, destacam-se as tecnologias LoRaWAN, SigFox, Weightless, NB-IoT e redes celulares (3G, 4G, 5G), que suportam milhões de nós em áreas extensas. Tecnologias emergentes como *Light Fidelity* (Li-Fi), *Software Defined Networking* (SDN) e Superfícies Inteligentes Reconfiguráveis (RIS) também são investigadas para ampliar eficiência, escalabilidade e cobertura. (HASSAN, 2018).

Na camada de aplicação, os nós utilizam protocolos específicos para troca de dados. O *Constrained Application Protocol* (CoAP) segue o modelo de requisição e resposta (similar ao HTTP), sendo adequado a dispositivos restritos por usar UDP. Já os protocolos MQTT e AMQP utilizam o modelo de publicação/assinatura (*publish/subscribe*), que reduz o tráfego e o consumo energético em redes com muitos nós. Alternativamente, o XMPP e o *Data Distribution Service* (DDS) oferecem suporte a comunicação em tempo real, sendo o último amplamente adotado em sistemas industriais pela alta confiabilidade e múltiplas políticas de Qualidade de Serviço (QoS). (HASSAN, 2018).

Protocolos como MQTT e CoAP destacam-se por sua leveza e simplicidade, enquanto DDS e AMQP oferecem robustez para sistemas distribuídos críticos. Assim, a integração dessas camadas e protocolos permite construir arquiteturas IoT escaláveis, eficientes e seguras. (HASSAN, 2018). A tabela abaixo mostra um comparativo entre cada uma dessas tecnologias.

Tabela 1 – Tecnologias de comunicação em IoT e suas características

Nome	**Frequência**	**Alcance**	**Norma**
Bluetooth	2.4 GHz	1-100 m	IEEE 802.15.1
BLE	2.4 GHz	> 100 m	Bluetooth SIG
GSM, LTE, LTE-M	EU: 900 MHz e 1.8 GHz; EUA: 850 MHz e 1.9 GHz	Variável conforme topologia	3GPP
6LoWPAN	2.4 GHz	10-30 m	IPv6 sobre IEEE 802.15.4
LoRa	< 1 GHz (ISM)	2-5 km (urbano); 15 km (suburbano); até 45 km (rural)	LoRa Alliance (LoRaWAN)
NB-IoT	700-900 MHz	10-15 km (rural)	3GPP LTE
NFC	13,56 MHz	< 0,2 m	ISO/IEC 18092
RFID	120-150 kHz (LF); 13.56 MHz (HF); 2.45-5.8 GHz (UHF); 3.1-10 GHz (SHF)	10 cm-200 m	ISO 18000
SigFox	900 MHz	3-10 km (urbano); 30-50 km (rural)	SigFox Alliance
Wi-Fi	2.4 GHz, 3.6 GHz, 4.9-5 GHz	> 100 m	IEEE 802.11
ZigBee	2.4 GHz; 784 MHz (China); 868 MHz (Europa); 915 MHz (EUA e Austrália)	10-20 m	IEEE 802.15.4

Fonte: adaptado de Hassan (2018).

2.3 LoRa e LoRaWAN

A padronização das tecnologias de comunicação na Internet das Coisas (IoT) segue lógica semelhante à da Internet tradicional, exigindo protocolos e normas específicas para interoperabilidade. Nesse contexto, o protocolo IEEE 802.15 (IEEE, 2020) define os princípios fundamentais das comunicações *wireless* de curto alcance, abrangendo tecnologias amplamente utilizadas, como Bluetooth, ZigBee, 6LoWPAN e outras soluções de baixo consumo energético, incluindo sistemas LPWAN como LoRa e SigFox.

O padrão IEEE 802.15 foi inicialmente concebido para redes pessoais sem fio, conhecidas como *Wireless Personal Area Networks* (WPANs), voltadas a aplicações de curta distância, baixo custo e reduzido consumo de energia. Com o avanço das aplicações IoT, surgiu uma ramificação voltada às *Low-Rate WPANs* (LR-WPANs), das quais derivam

tecnologias como ZigBee e 6LoWPAN. Mais recentemente, a especificação IEEE 802.15.4w foi introduzida para contemplar as *Low Power Wide Area Networks* (LPWANs), destinadas a comunicações de longo alcance e baixa potência, representadas por sistemas como LoRa e SigFox (IEEE, 2020).

As LPWANs destacam-se pela capacidade de fornecer conectividade econômica a dispositivos distribuídos em grandes áreas, priorizando autonomia energética e ampla cobertura em detrimento de altas taxas de dados e baixa latência. Essas redes geralmente operam com margens de conexão entre 155 dB e 160 dB, permitindo comunicações em distâncias de 10 a 15 km em ambientes rurais, com potência de transmissão em torno de 14 dBm, mantendo baixo consumo de energia (IEEE, 2020).

Entre as tecnologias LPWAN, o LoRa (Long Range) representa uma solução de modulação de rádio voltada à IoT, desenvolvida originalmente pela empresa francesa Cycleo e posteriormente adquirida pela Semtech Corporation em 2012. É importante distinguir entre LoRa e LoRaWAN, frequentemente tratados como sinônimos, mas que operam em camadas distintas do modelo de rede. LoRa refere-se à camada física, responsável pela modulação *Chirp Spread Spectrum* (CSS), em que os dados são transmitidos por meio de sinais do tipo "*chirp*", cuja frequência aumenta ou diminui ao longo do tempo. Já o LoRaWAN atua na subcamada MAC (Medium Access Control), pertencente à camada de enlace, e define o protocolo de rede responsável pela comunicação bidirecional, segurança ponta a ponta, mobilidade e serviços de localização entre os nós finais e os gateways (IDRIS; KARUNATHILAKE; FöRSTER, 2022).

O LoRa opera em faixas sub-GHz ISM (Industrial, Scientific and Medical) não licenciadas, variando conforme a regulamentação de cada país. No Brasil, a faixa de operação é de 915–928 MHz, regulamentada pela Anatel, destinada a aplicações de IoT e comunicações de longo alcance com baixo consumo de energia (Agência Nacional de Telecomunicações (ANATEL), 2017).

Atualmente, dispositivos baseados em LoRa e LoRaWAN são amplamente empregados em aplicações de IoT voltadas a desafios globais, como já anteriormente citados eles agregam em aplicações como cidades inteligentes, transporte, gestão de energia, monitoramento ambiental, agricultura de precisão e saúde conectada. Uma rede LoRaWAN típica é composta por quatro elementos: nós finais, *gateways* LoRa, servidor de rede e servidores de aplicação. Os nós finais são dispositivos de baixa potência, equipados com sensores e microcontroladores, que se comunicam com os *gateways* em uma topologia em estrela, favorecendo maior eficiência energética. Os *gateways*, por sua vez, agregam as mensagens recebidas de múltiplos nós e as encaminham ao servidor de rede, que processa, remove duplicatas e direciona os dados aos servidores de aplicação correspondentes (IDRIS; KARUNATHILAKE; FöRSTER, 2022).

Essa comunicação é bidirecional, permitindo tanto o envio de dados dos nós aos

servidores quanto o envio de comandos na direção oposta. O LoRaWAN, projetado pela LoRa Alliance, é mantido como um protocolo aberto e em constante evolução, buscando atender aos requisitos fundamentais da IoT, como escalabilidade, segurança e interoperabilidade. (LORA® ALLIANCE, 2015). Esses aspectos podem ser observados nos trabalhos de Farhad e Pyun (2023), Haxhibeqiri et al. (2018) e Jabbar et al. (2022).

Do ponto de vista técnico, o LoRa apresenta operação de baixa potência (autonomia de até 10 anos por bateria), baixa taxa de dados (até 27 kb/s com fator de espalhamento 7 em canal de 500 kHz ou 50 kb/s com FSK) e longo alcance (2–5 km em áreas urbanas, 15 km em áreas suburbanas e até 45 km em áreas rurais). Quando implementado dentro da arquitetura LoRaWAN, o LoRa torna-se parte de uma rede LPWAN que oferece comunicação bidirecional, segurança de ponta a ponta e suporte à mobilidade e geolocalização (LORA® ALLIANCE, 2015; ADELANTADO et al., 2017).

2.4 Parâmetros de Configuração de Dispositivos LoRa

Idris, Karunathilake e Förster (2022) descrevem cinco parâmetros principais de configuração utilizados em dispositivos LoRa, destacando suas relações e impactos sobre o desempenho da rede. Esses parâmetros são: Potência de Transmissão (TP), Fator de Espalhamento (SF), Largura de Banda (BW), Taxa de Codificação (CR) e Frequência da Portadora (CF).

Potência de Transmissão (TP)

A Potência de Transmissão define a intensidade do sinal enviado pelo transmissor. Em rádios LoRa, a TP varia normalmente de -4 a 20 dBm, com incrementos de 1 dB. Devido a limitações de hardware, essa faixa costuma restringir-se entre 2 e 20 dBm. Valores menores de TP aumentam a vida útil da bateria, mas reduzem o alcance de comunicação. Por outro lado, potências mais elevadas ampliam a cobertura, porém com maior consumo energético.

Fator de Espalhamento (SF)

O Fator de Espalhamento (SF) representa o número de *chirps* gerados por símbolo, com valores típicos entre 7 e 12. Um SF $= 8$, por exemplo, indica que cada símbolo é formado por $2^8 = 256$ *chips*. Valores maiores de SF aumentam o ganho de processamento e a robustez contra ruído, porém reduzem a taxa de dados e elevam o consumo de energia.

A relação entre SF, largura de banda (BW) e duração do *chirp* (T_s) é dada por:

$$2^{SF} = BW \cdot T_s \tag{2.1}$$

A taxa de símbolos (R_s) e a taxa de bits de modulação (R_b) podem ser expressas como:

$$R_s = \frac{1}{T_s} = \frac{BW}{2^{SF}} \quad \text{[símbolos/s]} \tag{2.2}$$

$$R_b = SF \cdot \frac{BW}{2^{SF}} \quad \text{[bits/s]} \tag{2.3}$$

Valores mais altos de SF aumentam a sensibilidade do receptor e o alcance, mas reduzem significativamente a taxa de bits. Em contrapartida, valores menores de SF aumentam a capacidade da rede e reduzem o tempo de transmissão (ToA).

Taxa de Codificação (CR)

A Taxa de Codificação (CR) refere-se ao grau de correção de erros do modem LoRa, fornecendo maior resiliência a interferências. A CR é definida pela razão:

$$CR = \frac{4}{4+n}, \quad n \in \{1, 2, 3, 4\} \tag{2.4}$$

Assim, são possíveis as taxas de 4/5, 4/6, 4/7 e 4/8. Uma CR mais alta (por exemplo, 4/8) oferece maior proteção contra erros, mas aumenta o tempo de transmissão e o consumo de energia.

A taxa de bits efetiva do sinal de dados também pode ser expressa como:

$$R_b = SF \cdot \frac{4}{4+CR} \cdot \frac{BW}{2^{SF}} \quad \text{[bits/s]} \tag{2.5}$$

Largura de Banda (BW)

A Largura de Banda (BW) define a faixa de frequência ocupada pelo sinal transmitido. Em LoRa, os valores mais comuns são 125 kHz, 250 kHz e 500 kHz. Larguras de banda maiores aumentam a taxa de transmissão, mas reduzem a sensibilidade do receptor. Inversamente, BW menores resultam em maior alcance e confiabilidade, porém com menor taxa de dados.

Frequência da Portadora (CF)

A Frequência da Portadora (CF) é o centro da faixa de operação utilizada pela modulação LoRa. No Brasil, conforme regulamentação da Anatel, o LoRa opera na banda ISM não licenciada de **915 MHz a 928 MHz** (Agência Nacional de Telecomunicações (ANATEL), 2017). O protocolo LoRaWAN define múltiplos canais dentro dessa faixa, permitindo comunicação bidirecional e flexível conforme a região e a densidade da rede.

Taxa de Dados e Carga Útil

A relação entre SF, BW e a taxa de bits é sintetizada na Tabela 2, considerando a banda ISM brasileira (915–928 MHz).

Tabela 2 – Taxas de dados e tamanho máximo da carga útil para a banda ISM 915–928 MHz (Brasil).

DR	**SF**	**BW (kHz)**	**Taxa de Bits (bit/s)**	**Carga Útil Máx. (bytes)**
DR0	12	125	250	51
DR1	11	125	440	51
DR2	10	125	980	51
DR3	9	125	1760	115
DR4	8	125	3125	242
DR5	7	125	5470	242
DR6	7	250	11000	242
DR7	7	500	21900	242

Fonte: adaptado de Idris, Karunathilake e Förster (2022), considerando a faixa ISM brasileira definida pela Anatel (Agência Nacional de Telecomunicações (ANATEL), 2017).

Conforme observado, fatores de espalhamento (SF) mais altos reduzem a taxa de bits e aumentam o alcance, enquanto larguras de banda maiores (BW) dobram a taxa de dados para o mesmo SF. Essa flexibilidade permite que o LoRaWAN equilibre alcance, robustez e consumo de energia de acordo com a aplicação e o ambiente de implantação.

Em suma, pode-se dizer que cada combinação específica de fator de espalhamento (SF) e largura de banda (BW), reflete em uma configuração que alterna entre alcance, robustez e velocidade de comunicação.

Tabela 3 – Relação entre *Data Rate*, fator de espalhamento e largura de banda no LoRaWAN.

DR	**SF**	**BW (kHz)**	**Taxa de bits (bit/s)**	**Interpretação**
DR0	12	125	250	Modo de longo alcance e alta robustez, adequado para sensores distantes e com baixa potência.
DR5	7	125	5470	Modo de curto alcance com maior velocidade de transmissão, indicado para áreas urbanas.
DR7	7	500	21900	Modo mais rápido, exige canal limpo e pouca interferência, ideal para cenários de alta densidade.

Fonte: o autor, adaptado de LoRa Alliance LoRa® Alliance (2015).

2.5 Segurança em IoT

Primeiramente, antes de definir o conceito de *fingerprinting*, é necessário compreender a importância da segurança não apenas em dispositivos IoT, mas em todo o sistema de comunicação. Segundo descrito por (SCHILLER et al., 2022), na década de 1980 a segurança da informação tinha como objetivo principal garantir a confidencialidade, disponibilidade e integridade das informações. Contudo, quando um terceiro obtém conhecimento da existência de uma comunicação, de sua origem ou de seu destino, a segurança pode ser comprometida. Sendo assim, o objetivo da segurança, nesse contexto, é definido da seguinte forma: uma mensagem é considerada confidencial quando apenas o remetente e o destinatário têm conhecimento de sua existência. O termo disponibilidade indica que a mensagem deve permanecer acessível e legível tanto pelo remetente quanto pelo destinatário a qualquer momento. Posteriormente, foi incorporado o conceito de responsabilidade, que assegura a capacidade de o destinatário comprovar a origem da mensagem e o remetente confirmar o recebimento. Dessa forma, o foco da segurança da informação evoluiu: inicialmente havia ênfase na disponibilidade, mas atualmente a prioridade é garantir confidencialidade, integridade e responsabilidade (SCHILLER et al., 2022).

De acordo com (HOSSAIN et al., 2024), sistemas IoT apresentam uma quantidade significativamente maior de interfaces de comunicação em comparação a sistemas digitais convencionais, em virtude da presença de dispositivos heterogêneos, múltiplos protocolos e integração com serviços em nuvem. Essa diversidade amplia as superfícies de ataque, ou seja, os pontos suscetíveis à exploração por agentes maliciosos. Assim, vulnerabilidades podem surgir nas comunicações entre dispositivos, controladores, gateways, usuários e servidores em nuvem, o que aumenta exponencialmente o número de vetores de ataque potenciais. Os autores ainda destacam três categorias principais de vulnerabilidades às quais os sistemas IoT estão particularmente expostos: vulnerabilidades em dispositivos finais, vulnerabilidades de comunicação e vulnerabilidades em serviços. De forma resumida, vulnerabilidades em dispositivos finais referem-se aos riscos associados a dispositivos de coleta, processamento e coordenação. Vulnerabilidades de comunicação abrangem falhas que podem ocorrer em protocolos como IPv4, IPv6, ZigBee, LoRaWAN e Wi-Fi, permitindo que uma vulnerabilidade em um protocolo se propague a outro. Por sua vez, vulnerabilidades em serviços englobam fragilidades em serviços web e APIs expostas ao público, frequentemente exploradas por ataques como *SQL injection, brute force e cross-site scripting*, especialmente quando não há práticas adequadas de autenticação e limitação de tentativas de acesso. Em síntese, (HOSSAIN et al., 2024) ressaltam que a segurança em IoT deve ser tratada de forma holística, abrangendo desde o hardware embarcado até os serviços em nuvem, de modo a mitigar a ampliação das superfícies de ataque e reduzir a probabilidade de exploração sistêmica em larga escala.

2.5.1 *Fingerprinting* de Radiofrequência (RF)

A identificação por *fingerprinting* tem ganhado relevância no contexto de dispositivos IoT e sistemas embarcados devido à necessidade crescente de mecanismos de autenticação seguros, não invasivos e compatíveis com plataformas de baixo custo. A ideia central consiste em explorar características físicas inerentes ao hardware, como imperfeições eletromagnéticas, variações de oscilador ou propriedades específicas dos sinais de RF, com o intuito de distinguir dispositivos individualmente, sem exigir modificações no *firmware* ou nos protocolos. Essa abordagem apresenta vantagens importantes como: consumo extremamente baixo de energia, ausência de sobrecarga computacional significativa e aplicabilidade mesmo em dispositivos limitados. (FENG et al., 2023)

Além das aplicações tradicionais em autenticação e detecção de intrusos, o *fingerprinting* também tem sido empregado para identificar dispositivos ocultos em ambientes sensíveis, contribuindo para mecanismos de segurança e privacidade. Diferentes modalidades de sinal podem ser exploradas para esse fim, incluindo radiação eletromagnética latente, sinais magnéticos e emissões de RF. No entanto, conforme destacado por Feng et al. (2023), muitos métodos do estado da arte ainda não oferecem uma solução genérica, robusta e aplicável a dispositivos IoT de baixo custo operando a distância, o que reforça a necessidade de técnicas mais leves, escaláveis e amplamente aplicáveis, motivação que fundamenta a proposta deste trabalho.

2.5.2 Relevância do *Fingerprinting* de RF

Além das vulnerabilidades estruturais discutidas anteriormente, os sistemas IoT estão sujeitos a uma ampla variedade de ataques direcionados à disponibilidade, integridade e confidencialidade das informações. Segundo (HOSSAIN et al., 2024), esses ataques podem ser classificados de acordo com o nível de dano causado às informações.

Como o foco deste trabalho é o uso do *fingerprinting*, a seguir são apresentados os principais tipos de ataques que podem ser mitigados com a aplicação correta dessa técnica.

- **Ataques por Interrupção (Interruption):** Ocorrem quando um invasor interfere intencionalmente nas comunicações entre dispositivos IoT, comprometendo a transmissão de dados entre nós e gateways. Como resultado, os serviços podem sofrer indisponibilidade e degradação de qualidade, afetando o desempenho da rede e a confiabilidade das aplicações IoT.
- **Ataque Man-in-the-Middle (MITM):** Nesse tipo de ataque, um nó malicioso posiciona-se entre dois dispositivos legítimos e intercepta a comunicação entre eles, fazendo-se passar por ambas as partes. Dessa forma, o atacante pode ler, modificar ou injetar dados sem que os dispositivos envolvidos percebam.

- **Ataques por Modificação (Modification):** Um nó malicioso altera o conteúdo das mensagens que trafegam pela rede, com o objetivo de enganar os dispositivos destinatários, induzindo-os a executar ações não autorizadas ou a revelar informações confidenciais.

- **Ataques por Falsificação (Fabrication):** Diferentemente dos ataques de modificação, a falsificação consiste na inserção de informações fraudulentas em uma mensagem legítima. O invasor cria ou injeta dados falsos, como cabeçalhos ou comandos, com o intuito de causar confusão e induzir o sistema a executar ações indevidas.

- **Ataques por Repetição de Mensagem (Replay Attack):** Ocorrem quando um intrusor captura e armazena mensagens legítimas transmitidas entre dispositivos, para reenviá-las posteriormente com o objetivo de se passar por um nó autorizado. Mesmo sem alterar o conteúdo das mensagens, o atacante pode reutilizar dados previamente transmitidos para obter acesso indevido ou autenticação não autorizada.

Em síntese, os ataques descritos por (HOSSAIN et al., 2024) demonstram que a segurança em sistemas IoT exige não apenas criptografia e autenticação robustas, mas também mecanismos de verificação de integridade, controle temporal e resiliência a interferências, a fim de garantir a confiabilidade das comunicações e a continuidade dos serviços.

Dessa forma, um passo inicial importante para garantir a segurança é conhecer e autenticar os dispositivos que integram a rede. Essa identificação pode ocorrer por meio de características únicas, digitais ou de hardware. Este processo é conhecido como *fingerprinting*. Segundo (ZHANG; GONG; QIAN, 2020), o *fingerprinting* de rádio frequência (RFFI) é uma abordagem promissora que utiliza as características físicas únicas dos dispositivos como identificadores. Essas características resultam de variações inevitáveis no processo de fabricação, que, mesmo com tecnologias avançadas, não podem ser totalmente eliminadas. Embora causem pequenas alterações na forma de onda das transmissões sem fio, essas variações permanecem dentro de limites que não afetam o funcionamento normal da comunicação. Por serem únicas, estáveis e difíceis de adulterar, tais imperfeições podem ser extraídas e usadas para identificar dispositivos em um sistema.

De acordo com (SOLTANIEH et al., 2020), para que um *fingerprint* seja confiável, ele deve apresentar as seguintes propriedades:

- **Universalidade:** todos os dispositivos devem possuir a característica analisada.

- **Unicidade:** nenhuma assinatura deve se repetir entre dispositivos diferentes;

- **Permanência:** as características devem ser estáveis ao longo do tempo e do ambiente.

- **Coletabilidade:** deve ser possível medir essas características com equipamentos disponíveis.
- **Robustez:** a assinatura deve resistir a variações externas, como temperatura, ruído, reflexões, etc.

O estudo em (SOLTANIEH et al., 2020) ainda destaca que o *fingerprinting* de RF têm sido amplamente aplicado em sistemas reais, como ADS-B na aviação, bluetooth, RFID e rádios *push-to-talk*, e que os métodos de extração de características são normalmente divididos em três abordagens:

- **Análises de transientes:** que observam o instante inicial da transmissão.
- **Estados estacionários:** que analisam o sinal contínuo durante a transmissão.
- **Métodos híbridos:** que exploram diferentes partes do sinal para fins de identificação.

Este trabalho adota uma abordagem de *fingerprinting* baseada em transientes, concentrando-se na fase inicial do preâmbulo do sinal LoRa, considerando como premissa que é neste momento em que as imperfeições ou diferenças do hardware se manifestam de forma mais evidente.

2.6 Aprendizagem de Máquina

Este trabalho emprega diferentes métodos de *machine learning* (ML) com o objetivo de investigar alternativas eficientes para identificação de dispositivos LoRa e comparar seu desempenho com técnicas já utilizadas no estado da arte. Atualmente, a aprendizagem de máquina está presente em inúmeros setores, incluindo reconhecimento facial, geolocalização, aplicações industriais, processamento de sinais e análise textual.

Segundo Mitchell (1997), ML é o estudo de algoritmos que melhoram seu desempenho automaticamente por meio da experiência. Suas aplicações abrangem mineração de dados, descoberta de conhecimento (ZHANG; ZHANG; YANG, 2003), classificação e reconhecimento de padrões, os quais, conforme Fukunaga (1990), baseiam-se em princípios estatísticos extraídos dos dados.

Os modelos de ML podem ser classificados em dois grupos principais: (i) métodos supervisionados, nos quais cada amostra possui um rótulo conhecido; e (ii) métodos não supervisionados, que buscam identificar estruturas e agrupamentos nos dados. Como o problema deste trabalho consiste em identificar qual dispositivo gerou uma dada transmissão LoRa, o foco recai sobre métodos supervisionados de classificação multiclasse.

Entre as diferentes abordagens existentes, destaca-se o *deep learning*. De acordo com Bengio, Courville e Vincent (2013), essas técnicas utilizam múltiplas camadas de transformações lineares e não lineares capazes de extrair características progressivamente mais abstratas. As Redes Neurais Artificiais (RNAs), por exemplo, são compostas por unidades interconectadas que processam entradas numéricas e aplicam funções de transferência, conforme descrito por Mitchell (1997) e Yao (1999). Apesar de seu poder de modelagem, tais métodos apresentam dois desafios relevantes para aplicações IoT: exigem grande quantidade de dados rotulados e demandam elevado custo computacional, dificultando a implementação em dispositivos de baixo consumo e baixa capacidade de processamento.

Diante disso, este trabalho explora principalmente métodos de *ensemble learning*, que combinam múltiplos classificadores para gerar predições mais robustas. Como discutido por Galar et al. (2012), ensembles tendem a reduzir variância e melhorar desempenho. Optiz e Maclin (1999) observa que métodos baseados em *bagging* e *boosting* podem superar classificadores individuais, sendo o primeiro mais estável e o segundo mais sensível à qualidade do conjunto de dados.

Entre os métodos avaliados neste estudo, quatro se destacam por sua ampla adoção e excelente desempenho em problemas de classificação multiclasse com dados tabulares:

- **Decision Tree (DT).** As árvores de decisão, formalizadas no método *Classification and Regression Trees* (CART) por Breiman et al. (1984), realizam particionamento recursivo do espaço de atributos. Trata-se de um modelo simples e interpretável, frequentemente utilizado como base teórica e prática para métodos ensemble, como *Random Forest* e *Gradient Boosting*. Neste trabalho, o DT foi incluído como classificador individual para fins comparativos.

- **Random Forest (RF).** O *Random Forest*, proposto por Breiman (2001), compõe múltiplas árvores de decisão treinadas sobre subconjuntos distintos de dados e atributos. As predições são combinadas por votação (ou média), resultando em um modelo robusto, pouco sensível a ruídos e apropriado para dados tabulares.

- **Gradient Boosting Machine (GBM).** Introduzido por Friedman (2001), o GBM constrói um conjunto de modelos fracos de forma sequencial, cada um focado em corrigir os erros do anterior. Essa abordagem tende a alcançar alta acurácia, especialmente quando os dados apresentam relações complexas.

- **XGBoost.** O XGBoost, desenvolvido por Chen e Guestrin (2016), é uma versão otimizada e escalável do *gradient boosting*, incorporando regularização explícita, paralelização e técnicas de redução de complexidade. Tornou-se um dos métodos mais populares em competições de aprendizagem de máquina devido à sua eficiência.

- **LightGBM.** O LightGBM, apresentado por Ke et al. (2017), utiliza estratégias como *Gradient-based One-Side Sampling* (GOSS) e *Exclusive Feature Bundling* (EFB) para acelerar o treinamento e reduzir o consumo de memória, mantendo desempenho comparável ou superior ao XGBoost. É especialmente adequado para grandes conjuntos de dados e tarefas com alta dimensionalidade.

Esses modelos foram escolhidos por três motivos principais: (i) apresentam boa relação entre custo computacional e desempenho, essencial em aplicações IoT; (ii) lidam adequadamente com *features* numéricas, como as métricas *Catch22*; e (iii) oferecem mecanismos de avaliação de importância de atributos, aspecto relevante para analisar quais características temporais contribuem mais para o processo de *fingerprinting* LoRa.

Métodos adicionais como MLP e Naive Bayes foram considerados inicialmente, mas sua performance inferior e menor interpretabilidade no contexto desta aplicação justificam sua não inclusão na comparação final. Da mesma forma, o método KNN, típico em tarefas de proximidade, não integra a análise final devido ao seu elevado custo computacional no momento da inferência e por não aparecer nos experimentos de resultados apresentados neste trabalho.

Por fim, embora modelos de regressão não sejam centrais para o problema de identificação de dispositivos, cabe mencionar que técnicas como *Ridge, Lasso, ElasticNet* e *Support Vector Regression* têm aplicações relevantes em análise de séries temporais e predição contínua. Pandangan e Talampas (2020) destaca, por exemplo, a utilização do *Random Forest Regressor* para problemas de predição, ressaltando a versatilidade do paradigma de ensembles.

3 Estado da Arte

3.1 Sistemas embarcados e Fingerprinting

A identificação de dispositivos por meio de *Radio Frequency Fingerprinting* (RFFI) tem se destacado como uma abordagem promissora para autenticação, detecção de intrusos e aumento da segurança em redes IoT. Em vez de depender apenas de credenciais criptográficas ou de software, o RFFI explora imperfeições físicas inevitáveis do hardware, que se manifestam no sinal transmitido e permitem distinguir dispositivos individualmente, ainda que utilizem o mesmo protocolo e firmware.

No contexto específico de dispositivos LoRa e redes LoRaWAN, o interesse por RFFI cresceu nos últimos anos, impulsionado pela necessidade de mecanismos de identificação que sejam compatíveis com as restrições de energia, custo e processamento típicas de sistemas IoT. Nesta seção, são apresentados os principais trabalhos relacionados, organizados por tipo de abordagem: métodos baseados em *Deep Learning*/CNN, abordagens fundamentadas em bases de dados virtuais e modelagem sintética, técnicas de portabilidade de modelos entre domínios e soluções que exploram diversidade espacial (MIMO).

3.2 Trabalhos Relacionados em RFFI para Dispositivos LoRa

O trabalho de Elmaghbub e Hamdaoui (2021) apresenta uma estrutura experimental que pretende auxiliar na compreensão de problemas associados à identificação dos dispositivos LoRa mediante as configurações de implantação. Ou seja, dado diferentes situações de implantação, qual seria a melhor maneira de otimizar os meios de identificação de dispositivos.

Para isso os autores recolheram um vasto conjunto de dados, em diferentes cenários, indoor e outdoor, com diferentes distâncias entre emissor e receptor, diferentes SFs para o LoRa e diferentes tipos de hardware para os receptores. Propuseram ainda uma nova técnica que explora distorções fora da banda de transmissão, provocadas por erros específicos de hardware individuais, para fornecer assinaturas únicas de dispositivos e aumentar a precisão da identificação.

O estudo experimental mostrou que modelos de identificação baseados em *Deep Learning* têm bom desempenho quando são treinados e testados sob as mesmas circunstâncias, sendo a representação em Domínio da Frequência (FFT) a mais eficaz. Contudo, quando eles são treinados e testados em diferentes condições, os modelos perdem precisão à medida que ocorrem mudanças nas condições de canal e qualquer decorrente alteração

da configuração LoRa ou hardware o receptor já não consegue classificar os dispositivos.

Em condições onde as coisas permanecem constantes, a representação FFT teve um bom desempenho, mas, uma vez que houve mudanças, o desempenho foi ruim. Dependendo da representação de entrada e das condições de teste, a precisão variou; por exemplo, com dados coletados no mesmo dia em cenários internos, a FFT atingiu mais de 80% de precisão, mas caiu para 5% quando testada com dados de dias diferentes. A representação IQ também registrou uma queda na precisão de 70% para 45% nas mesmas circunstâncias.

A partir dessa mesma base de dados este trabalho foi realizado, assim como os demais trabalhos abaixo, os quais também aplicaram estudos não só sobre essa base mas também em outras mais.

3.2.1 Abordagens Baseadas apenas em *Deep Learning* e CNN

Uma parte significativa da literatura recente em RFFI para LoRa utiliza arquiteturas de *deep learning*, em especial Redes Neurais Convolucionais (CNN), aplicadas a diferentes representações do sinal de rádio.

Shen et al. (2021) propõem um sistema de identificação de dispositivos LoRa baseado em *fingerprinting* de RF utilizando CNNs aplicadas a amostras IQ, FFT e espectrogramas. Um ponto central do trabalho é o tratamento do *Carrier Frequency Offset* (CFO), cuja correção mostrou-se fundamental para obter boas taxas de acerto. Em experimentos com 20 dispositivos LoRa, a acurácia com CNN e IQ bruto foi de 59,44% sem compensação de CFO, subindo para 83,36% após a correção. Para FFT, os resultados passaram de 51,62% para 87,36%, e com espectrogramas a combinação CNN + compensação de CFO atingiu 96,44%. Uma arquitetura híbrida (CNN + camadas totalmente conectadas) aplicada a espectrogramas alcançou a melhor acurácia, de 97,61%.

Apesar da alta precisão, a abordagem exige grande poder computacional, tanto na fase de treinamento quanto na inferência, além de etapas adicionais de pré-processamento (cálculo de FFT e espectrogramas) e grande volume de dados rotulados. Isso limita a viabilidade de implementação direta em dispositivos LoRa embarcados, que operam sob fortes restrições de energia, memória e capacidade de processamento.

Embora não se restrinja apenas a LoRa, o trabalho de Chillet et al. (2024), introduz a ferramenta RiFyFi_VDG para geração de bases de dados virtuais de RFFI, permitindo simular efeitos de CFO, ruído de fase, *IQ imbalance* e distorção de amplificador de potência. As bases geradas são então utilizadas para treinar modelos de *deep learning* inspirados em arquiteturas clássicas como AlexNet. Os autores demonstram que tais modelos conseguem aprender assinaturas de hardware de maneira controlada e reprodutível, permitindo estudos em larga escala de sensibilidade a imperfeições físicas.

Novamente, porém, o custo computacional permanece elevado: redes profundas,

grande número de amostras por transmissor e longos tempos de treinamento tornam a abordagem mais adequada a cenários de laboratório ou simulação do que a sistemas embarcados em campo, especialmente em redes LoRa de baixa potência.

3.2.2 Abordagens com Portabilidade de Modelos entre Domínios

Outro conjunto de trabalhos foca não apenas na acurácia em um cenário fixo, mas na capacidade de um modelo manter bom desempenho quando mudam o canal, o hardware receptor ou as condições de coleta, problema conhecido como portabilidade ou adaptação de domínio.

Gaskin et al. (2023) propõem o método Tweak, que busca tornar modelos de *deep learning* para RFFI mais portáveis entre domínios distintos (por exemplo, diferentes receptores, dias de coleta ou configurações de canal). A técnica combina aprendizado métrico e uma etapa leve de calibração, na qual um modelo pré-treinado é ajustado ao novo domínio usando apenas uma pequena quantidade de dados rotulados.

Em cenários de portabilidade de canal, o Tweak alcança acurácia media de 89,9%, e em portabilidade de hardware, 76,0%, superando abordagens convencionais de CNN treinadas do zero para cada caso, que permanecem tipicamente abaixo de 75%. Ainda assim, o método depende de uma fase inicial de treinamento profundo cara, e a necessidade de calibração com dados rotulados do domínio-alvo pode ser um entrave em sistemas que demandam escalabilidade ou mínima intervenção em campo.

3.2.3 Abordagens com Diversidade Espacial e MIMO

Uma linha complementar explora diversidade espacial para tornar o RFFI mais robusto a variações de canal. Em vez de utilizar apenas uma antena e um receptor (configuração SISO), são empregados sistemas com Múltiplas Entradas e Saídas (Multiple Input, Multiple Output - MIMO), para capturar diferentes perspectivas do mesmo sinal.

O método proposto por Basha et al. (2023), denomina-se *Channel-Resilient Fingerprinting* (CR-FP), que utiliza uma arquitetura MIMO para coletar simultaneamente múltiplos fluxos de dados de RF via diferentes antenas. Representações como IQ, espectrogramas e amplitude são combinadas em um modelo de *deep learning*, buscando maior robustez a desvanecimento e interferência.

Os resultados indicam ganhos significativos em relação a abordagens SISO tradicionais: em certos cenários, a acurácia aumenta em até 69%, e mesmo sob forte degradação de canal o método consegue identificar 20 dispositivos com acurácia de até 64%, enquanto a abordagem SISO se aproxima do desempenho aleatório. Em contrapartida, a exigência de uma infraestrutura MIMO, maior largura de banda de processamento e sincronização entre

múltiplos canais torna a solução pouco compatível com a realidade de dispositivos LoRa embarcados, que costumam operar com rádios simples e apenas uma cadeia de recepção.

3.3 Síntese e Comparação de Resultados

A partir da revisão realizada, é possível destacar alguns pontos em comum entre as abordagens estudadas:

- Métodos baseados em CNN e *deep learning* alcançam acurácias elevadas em ambientes controlados, mas exigem grande quantidade de dados rotulados, alto poder computacional e etapas de pré-processamento complexas.
- Abordagens baseadas em bases virtuais e modelagem sintética são valiosas para investigar efeitos físicos e limites teóricos, porém continuam centradas em arquiteturas de *deep learning* e não necessariamente traduzem, de forma direta, condições reais de operação em campo.
- Técnicas de portabilidade de modelos, como o Tweak, mitigam parcialmente o problema da mudança de domínio, mas ainda dependem de calibrar modelos profundos e de coletar amostras rotuladas em cada novo cenário.
- Soluções que exploram diversidade espacial com MIMO apresentam ganhos de robustez em relação ao canal, mas demandam infraestrutura de recepção mais cara e complexa, pouco compatível com dispositivos LoRa de baixo custo.

A Tabela 4 mostra uma comparação entre os trabalhos aqui citados e este trabalho, avaliando aspectos como **Acurácia, Portabilidade, Complexidade, Escalabilidade e Replicabilidade**:

Tabela 4 – Síntese comparativa das principais abordagens de RFFI para dispositivos LoRa.

Trabalho	Acurácia	Portabilidade	Complexidade	Escalabilidade	Replicabilidade
(ELMAGHBUB; HAMDAOUI, 2021)	Alta (80–90% em condições iguais)	Baixa	Alta	Média	Baixa
(SHEN et al., 2021)	Muito alta (até 97%)	Baixa	Alta	Baixa	Baixa
(CHILLET et al., 2024) (RiFyFi)	Alta (em simulação)	Baixa	Muito alta	Alta (simulada)	Média
(GASKIN et al., 2023) (Tweak)	Alta (75–89%)	Alta (domínios distintos)	Alta	Média	Média
(BASHA et al., 2023) (CR-FP MIMO)	Alta	Média	Muito alta	Baixa	Baixa
Este trabalho	Alta (63–87%)	Média–Alta	Baixa	Alta	Alta

Em conjunto, esses fatores evidenciam uma lacuna importante na literatura: a ausência de métodos de RFFI para LoRa que ofereçam um compromisso favorável entre desempenho, robustez e complexidade computacional, capazes de operar com recursos restritos e sem necessidade de redes neurais profundas. A abordagem proposta nesta dissertação busca justamente explorar esse espaço, utilizando características temporais extraídas via Catch22 combinadas com modelos de *machine learning* de menor complexidade,

investigando sua eficácia na identificação de dispositivos LoRa em diferentes condições de coleta.

4 Método Proposto

A metodologia proposta engloba diferentes etapas, desde a ativação inicial e coleta de dados dos dispositivos até a operação final de identificação do dispositivo, conforme ilustrado na Figura 2.

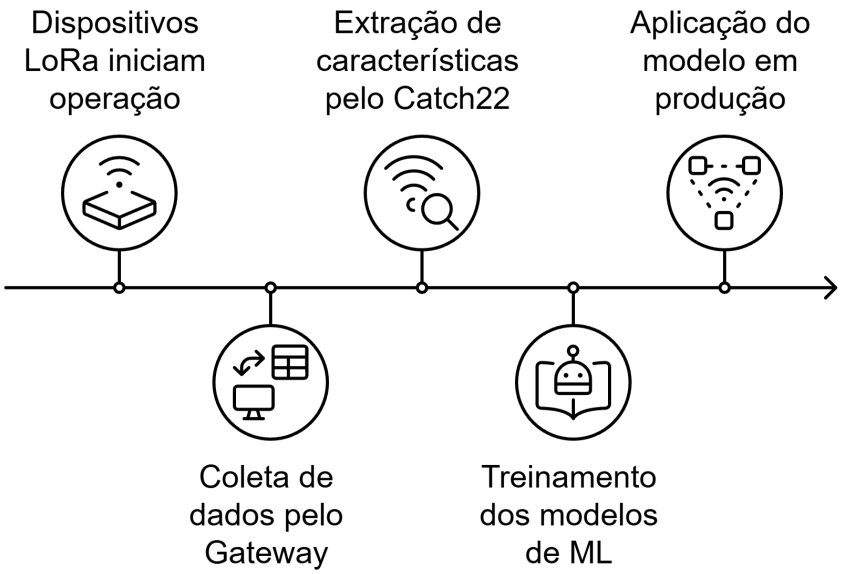


Figura 2 – Processo de identificação dos dispositivos LoRa.

Fonte: o autor.

1. **Início da operação dos dispositivos LoRa:** Neste estágio inicial, os dispositivos LoRa são integrados à rede para formar o sistema IoT. Cada dispositivo é configurado com parâmetros específicos, incluindo frequência, largura de banda e fator de espalhamento. Uma vez configurados, os dispositivos iniciam uma transmissão periódica de sinais, com base nos requisitos da aplicação. Esses sinais transmitidos são então capturados e armazenados para análise posterior do algoritmo.

2. **Período de coleta de dados no Gateway:** Uma vez que os dispositivos estejam operacionais, um subconjunto inicial dos dados coletados é analisado. Dado que os sinais podem ser afetados por vários fatores no cenário de aplicação, esta etapa é crucial para capturar dados representativos de cada dispositivo integrado. O objetivo é garantir um número suficiente de pacotes para o processo de *fingerprinting*, mantendo a qualidade e a consistência dos dados usados na análise.

3. **Extração de características usando a biblioteca Catch22 (LUBBA et al., 2019):** Os sinais IQ coletados passam por processamento para extrair características das séries temporais. Usando a biblioteca de extração de características Catch22, são obtidas 22 características únicas derivadas dos pacotes de sinal de cada dispositivo, servindo como métricas distintas. Esses dados extraídos formam vetores de características que podem ser usados como assinatura de radiofrequência dos transmissores. As subseções 4.1 e 4.2 fornecem uma explicação detalhada do processo de coleta de dados e dos recursos extraídos, respectivamente.

4. **Treinamento de modelo:** Uma vez que os vetores de características são gerados, os modelos de machine learning são treinados para classificar dispositivos com base em suas características de sinal únicas. O processo de implementação do treinamento é detalhado na subseção 4.4.

5. **Implantação do modelo e operação diária:** Após a avaliação, o modelo de melhor desempenho é implantado para a classificação diária de dispositivos LoRa. Ele analisa dados de sinal recém-coletados, permitindo uma solução de *fingerprinting* RF escalável para monitoramento contínuo de dispositivos, detecção de anomalias e segurança aprimorada do sistema IoT. Todo o fluxo de trabalho do sistema é projetado para manter baixo custo de processamento e aplicabilidade, garantindo ao mesmo tempo um desempenho tão eficaz quanto o de modelos de *fingerprinting* mais complexos.

4.1 Base de Dados

Os conjuntos de dados de *fingerprinting* de RF para LoRa têm sido essenciais para o avanço da pesquisa em identificação de dispositivos e segurança de IoT. Existem vários conjuntos de dados na literatura, cada um projetado para configurações experimentais e objetivos de pesquisa específicos.

Diversos conjuntos de dados públicos relacionados a LoRa e LoRaWAN têm sido publicados recentemente, cada um voltado a uma finalidade específica dentro do ecossistema IoT. O dataset *LoRaWAN Traffic Analysis* (POVALAC; KRAL, 2023) reúne tráfego real capturado em quatro cidades europeias, contemplando análise de beacons da Classe B, conformidade regulatória e padrões de comunicação, sendo valioso para estudos de segurança do protocolo e comportamento de rede, mas não contém amostras de sinal em banda base necessárias para *fingerprinting* de camada física. De forma semelhante, o *LoRa Signal Quality and GPS Positioning Dataset* (ESCOBAR et al., 2023) disponibiliza séries temporais de RSSI, SNR e localização geográfica de dispositivos móveis em um arquipélago espanhol, permitindo pesquisas em geolocalização e modelagem de propagação,

porém não inclui dados de IQ ou preâmbulos capturados diretamente do rádio, essenciais para caracterização de assinaturas de hardware.

Já os datasets de Aernouts et al. (AERNOUTS et al., 2018) oferecem medições massivas de RSSI provenientes de múltiplas estações-base em cenários urbanos e rurais, focados em *fingerprinting* para localização *outdoor*, mas igualmente limitados à camada MAC e sem acesso ao sinal bruto. Em contraste, o conjunto de dados disponibilizado por Elmaghbub et al. (ELMAGHBUB; HAMDAOUI, 2021), utilizado neste trabalho, fornece amostras IQ completas capturadas diretamente do preâmbulo LoRa em condições controladas e replicáveis, abrangendo múltiplos dispositivos, cenários *indoor* e *outdoor*, diferentes fatores de espalhamento e variações temporais. Por disponibilizar o sinal em sua forma mais pura, antes de qualquer demodulação ou processamento, esse *dataset* é o único entre os analisados que permite investigar assinaturas físicas intrínsecas ao hardware, sendo portanto o mais adequado para experimentos de identificação de dispositivos via *fingerprinting* de radiofrequência.

O conjunto de dados inclui amostras IQ no domínio do tempo e representações FFT de transmissões LoRa, coletadas usando uma plataforma de teste que compreende dispositivos IoT e Rádios Definidos por Software (Software Defined Radios - SDRs). O cenário de coleta desta base utilizou 25 dispositivos Pycom IoT idênticos e receptores USRP B210, todos operando a 915 MHz, com sinais amostrados na taxa de 1 MS/s. Cada transmissão gerou arquivos de dados IQ brutos e representações baseadas em FFT no formato ".dat". Os arquivos binários são codificados em Float32 com valores complexos intercalados, onde índices ímpares armazenam componentes em fase (I) e índices pares armazenam componentes em quadratura (Q). Além disso, arquivos de metadados foram gerados, fornecendo informações importantes como taxa de amostragem, registro de data e hora, dia da gravação, frequência da portadora e outros parâmetros relevantes para a configuração experimental.

O conjunto de dados completo consiste em 16.300 arquivos, excedendo 1,2 TB de dados. No entanto, para este estudo, focamos especificamente no Cenário 1: Configuração indoor em dias diferentes, que é estruturado da seguinte forma: O diretório *Diff Days Indoor Setup* consiste em cinco subdiretórios, cada um correspondendo a um dia de gravação diferente. Dentro do subdiretório de cada dia, há 25 pastas específicas do dispositivo, uma para cada dispositivo Pycom IoT. Cada pasta do dispositivo contém 20 gravações SigMF, representando 10 transmissões. Ou seja, para cada transmissão, há um arquivo de dados binários armazenando as amostras IQ e um arquivo de metadados correspondente.

Segundo (ELMAGHBUB; HAMDAOUI, 2021), com o objetivo de permitir a avaliação do desempenho, minimizando os impactos de variáveis externas, foi definido um cenário experimental *indoor* controlado, no qual foram conduzidos os experimentos e realizada a coleta de dados. Esses experimentos ocorreram em um ambiente ocupado

(representando uma sala típica) durante cinco dias consecutivos.

Todos os dispositivos foram configurados para transmitir uma mesma mensagem a partir da mesma posição física, localizada a uma distância de 5 metros do receptor. Dessa forma, buscou-se garantir que todos os transmissores estivessem sujeitos às mesmas condições de canal. Cada transmissor gerou, a cada dia, 10 transmissões com duração de 20 segundos, espaçadas por intervalos de 1 minuto. Como resultado, foram obtidas aproximadamente 200 milhões de amostras complexas por dispositivo, por dia (ELMAGHBUB; HAMDAOUI, 2021).

Ao selecionar este cenário específico, reduzimos o conjunto de dados a um tamanho gerenciável, preservando a diversidade nas condições de gravação. Isso permite experimentação eficaz e extração de recursos, suportando fingerprinting de RF e outros aplicativos relacionados. A Figura 3 representa a organização do conjunto de dados e demais cenários existentes.

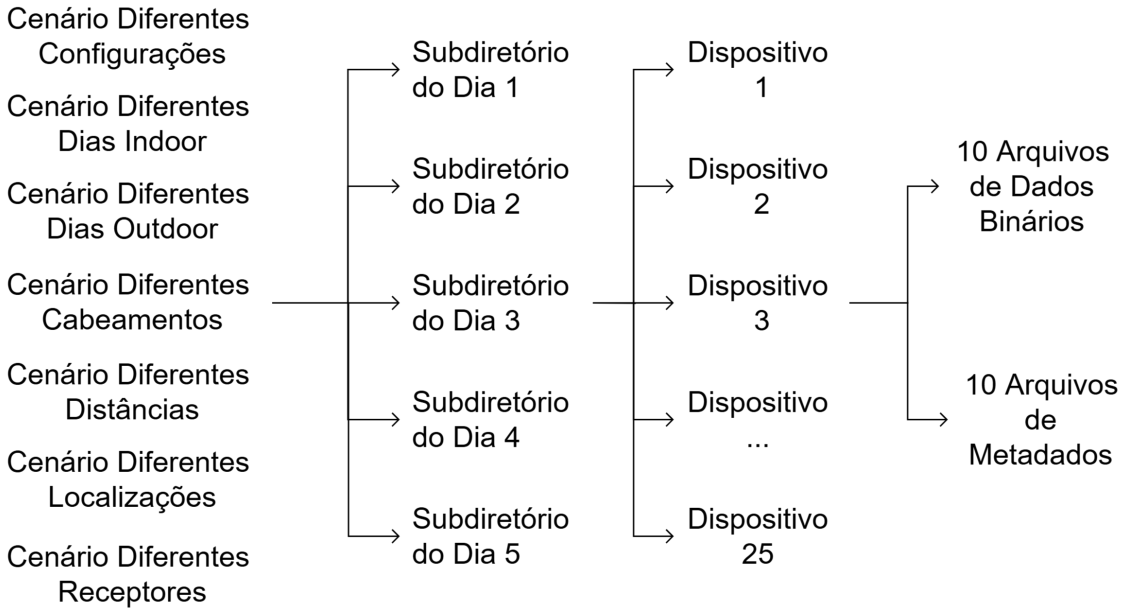


Figura 3 – Estrutura da coleta de dados

Fonte: o autor.

Além do hardware utilizado, as configurações específicas da modulação LoRa são cruciais para a replicabilidade dos experimentos de *fingerprinting*. Todos os dispositivos transmissores foram operados no modo *Raw-LoRa*, utilizando as configurações detalhadas na Tabela 5.

Tabela 5 – Parâmetros de configuração LoRa e especificações do hardware de captura.

Parâmetro	**Especificação**
Dispositivos Transmissores	23 LoPy4, 2 FiPy (chips Semtech SX1276)
Fator de Espalhamento (SF)	7
Largura de Banda (BW)	125 kHz
Taxa de Codificação (CR)	4/5
Tamanho do Preâmbulo	8 símbolos
Potência de Transmissão	20 dBm
Frequência Central (CF)	915 MHz
Taxa de Amostragem (SDR)	1 MS/s
Hardware do Receptor	USRP B210 com Antena Vert900 Ettus

Fonte: Adaptado de (ELMAGHBUB; HAMDAOUI, 2021).

4.2 *Catch 22*

O *Catch22* (*CAnonical Time-series CHaracteristics*) é uma biblioteca de análise de séries temporais *open-source* desenvolvida em linguagem C, com integrações para Python, R, Julia e Matlab. Publicada em 2019 por (LUBBA et al., 2019), a biblioteca surgiu da necessidade de equilibrar a capacidade discriminativa de características estatísticas com a eficiência computacional necessária para aplicações práticas.

A seleção das métricas baseou-se em uma versão filtrada da biblioteca *Highly Comparative Time-Series Analysis* (HCTSA), que possui 4.791 características. Através de uma análise comparativa em 93 conjuntos de dados de classificação (totalizando cerca de 147.000 séries temporais), os autores identificaram um subconjunto canônico de apenas 22 características fundamentais. Esta redução drástica permitiu uma aceleração de aproximadamente 1.000 vezes na velocidade de processamento em comparação com a biblioteca completa, com uma perda de acurácia média de apenas 7% (LUBBA et al., 2019).

Para o contexto desta dissertação, tal eficiência é fundamental, pois permite a extração de assinaturas de RF sem a necessidade de hardware de alto desempenho ou GPUs, diferenciando-se de soluções baseadas em *Deep Learning*. As 22 características originais, acrescidas da **Média** e do **Desvio Padrão** (totalizando 24 dimensões neste estudo), estão listadas a seguir:

1. **DN_HistogramMode_5:** Modo da distribuição padronizada (histograma de 5 bins).
2. **DN_HistogramMode_10:** Modo da distribuição padronizada (histograma de 10 bins).
3. **SB_BinaryStats_mean_longstretch1:** Maior período de valores consecutivos

acima da média.

4. **DN_OutlierInclude_p_001_mdrmd:** Intervalos de tempo entre eventos extremos sucessivos acima da média.
5. **DN_OutlierInclude_n_001_mdrmd:** Intervalos de tempo entre eventos extremos sucessivos abaixo da média.
6. **CO_f1ecac:** Primeiro cruzamento da função de autocorrelação em $1/e$,
7. **CO_FirstMin_ac:** Primeiro mínimo da função de autocorrelação.
8. **SP_Summaries_welch_rect_area_5_1:** Potência total na menor quinta parte das frequências no espectro de potência de Fourier.
9. **SP_Summaries_welch_rect_centroid:** Centroide do espectro de potência de Fourier.
10. **FC_LocalSimple_mean3_stderr:** Erro médio de previsão em uma média móvel de 3 amostras.
11. **CO_trev_1_num:** Estatística de reversibilidade temporal $(x_{t+1} - x_t)^3$ para t.
12. **CO_HistogramAMI_even_2_5:** Informação mútua automática com $m = 2$ e $\tau = 5$.
13. **IN_AutoMutualInfoStats_40_gaussian_fmmi:** Primeiro mínimo da função de informação mútua automática.
14. **MD_hrv_classic_pnn40:** Proporção de diferenças sucessivas que excedem 0.04σ.
15. **SB_BinaryStats_diff_longstretch0:** Maior período de decréscimos incrementais sucessivos.
16. **SB_MotifThree_quantile_hh:** Entropia de Shannon de duas letras sucessivas em uma simbolização de 3 letras equiprovável.
17. **FC_LocalSimple_mean1_tauresrat:** Mudança no comprimento de correlação após diferenciação iterativa.
18. **CO_Embed2_Dist_tau_d_expfit_meandiff:** Ajuste exponencial das distâncias sucessivas no espaço de incorporação em 2D.
19. **SC_FluctAnal_2_dfa_50_1_2_logi_prop_r1:** Proporção de flutuações de longo prazo que escalam com DFA (amostragem de 50%).
20. **SC_FluctAnal_2_rsrangefit_50_1_logi_prop_r1:** Proporção de flutuações de longo prazo que escalam com ajustes de faixa linearmente reescalados.

21. **SB_TransitionMatrix_3ac_sumdiagcov:** Traço da covariância da matriz de transição entre símbolos em um alfabeto de 3 letras.

22. **PD_PeriodicityWang_th0_01:** Medida de periodicidade.

As características do *Catch22* podem ser agrupadas em categorias funcionais, conforme ilustrado na Figura 4:

- **Forma de Distribuição:** Examina a distribuição dos valores da série temporal, independente da ordem temporal (ex: histogramas e momentos).
- **Cronometragem de Eventos Extremos:** Analisa o momento e a frequência de valores atípicos em relação à duração da série.
- **Autocorrelação Linear:** Quantifica as dependências lineares e a estrutura de memória do sinal.
- **Autocorrelação Não Linear:** Captura relacionamentos complexos e dependências de ordem superior.
- **Simbólico:** Transforma os dados em representações discretas para analisar padrões de transição.
- **Dimensionamento Autoafim:** Identifica correlações de longo alcance através de análise de flutuação.

Uma descrição detalhada de cada categoria e a formulação matemática das métricas podem ser consultadas em (LUBBA et al., 2019).

4.3 Procedimento de extração de características

Este estudo integra o ambiente *MATLAB* com a biblioteca *catch22* para a extração de características de séries temporais, aproveitando a capacidade do *MATLAB* para visualização de dados, pré-processamento e manipulação eficiente de grandes volumes de amostras IQ. As ferramentas integradas do *MATLAB* permitem um fluxo de trabalho otimizado, facilitando operações computacionalmente intensivas (MATHWORKS, 2024).

O pré-processamento foi realizado reconstruindo o sinal complexo a partir das amostras IQ. No código, a reconstrução do sinal é feita com $data = complex(I, Q)$, e a envoltória ou envelope é obtida com $abs(data)$, que equivale a:

$$A(t) = \sqrt{I^2(t) + Q^2(t)}.$$

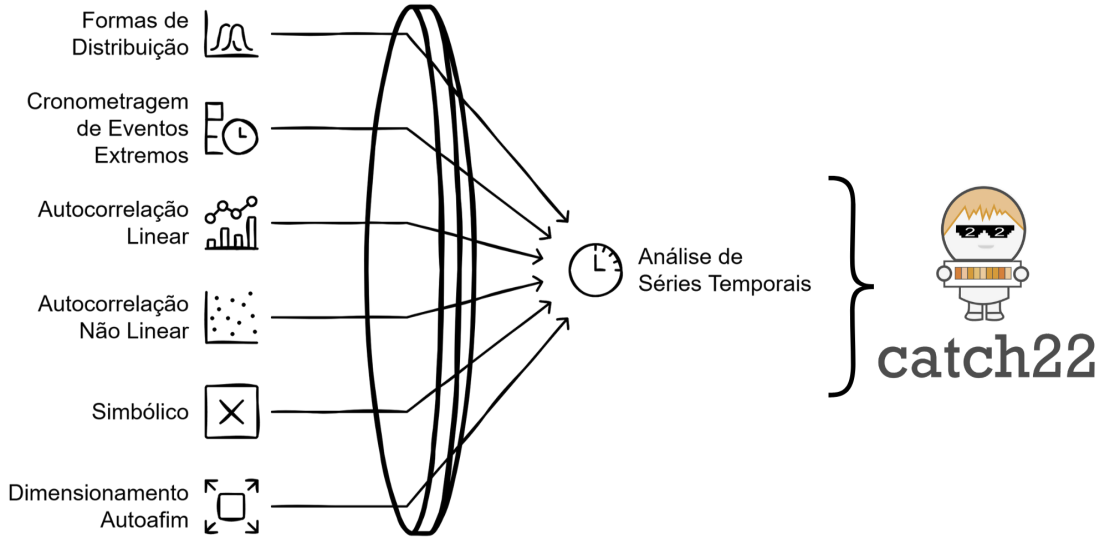


Figura 4 – Visão geral dos grupos de análise do Catch22.

Fonte: o autor.

Em seguida, a componente real equivalente é calculada por $x = abs(data). * cos(angle(data))$, resultando em:

$$x(t) = A(t)\cos(\theta(t)),$$

onde $A(t)$ é a envoltória ou envelope do sinal IQ e $\theta(t)$ a fase instantânea. Essa representação compacta preserva as informações dinâmicas de amplitude e fase relevantes para a tarefa de identificação, ao mesmo tempo em que reduz a dimensionalidade e facilita a extração de características com o conjunto de *features* do *catch22*.

A detecção do início dos pacotes foi realizada aplicando um limiar empírico ($\delta = 0.005$) à média da magnitude do sinal em janelas deslizantes de 1000 amostras. O valor foi definido empiricamente a partir da análise do nível de ruído e da amplitude média dos pacotes, garantindo uma identificação confiável do transiente de ativação. Pequenas variações nesse parâmetro não comprometem o desempenho do sistema, uma vez que a amplitude do transiente é significativamente superior ao ruído de fundo, assegurando a robustez do método.

A definição do limiar ($\delta = 0.005$) não tem apenas o objetivo de detectar o início dos pacotes, mas também de isolar a porção do sinal que concentra as assinaturas únicas de cada transmissor. Esse trecho inicial, correspondente ao transiente de ativação do oscilador e do circuito de modulação, reflete diretamente as imperfeições analógicas de cada dispositivo, como instabilidades de frequência, atrasos de fase e pequenas não linearidades. Tais imperfeições são características intrínsecas de hardware e permanecem praticamente invariantes ao longo do tempo, constituindo a base para a identificação por impressão

digital de RFFI. Assim, o uso do limiar permite capturar a região mais informativa do sinal, reduzindo o volume de dados analisados e, ao mesmo tempo, preservando as variações específicas que diferenciam cada dispositivo LoRa.

Durante esta etapa, também foram realizados experimentos variando o tamanho da janela de análise e o valor do limiar δ, com o objetivo de selecionar os parâmetros que maximizassem a robustez e a reprodutibilidade da extração de características. Cada experimento foi conduzido separadamente para os dados de um único dia, considerando o custo computacional da operação e o grande número de transmissões. Ainda assim, o tempo de processamento obtido foi significativamente inferior ao de abordagens baseadas em espectrogramas, validando a viabilidade do método para cenários de larga escala. O procedimento completo e a versão final selecionada para os experimentos subsequentes são resumidos no Algoritmo 1.

Algorithm 1 Extração de Características de Dados IQ em Multiplos Arquivos

Require: $device\_list$ ▷ Lista de dispositivos
Require: $device\_files$ ▷ Lista de arquivos IQ agrupados por dispositivo
Require: packet_size ▷ Tamanho do pacote LoRa (em amostras)
Require: window_size ▷ Tamanho da janela de análise (em amostras)
Require: offset ▷ Offset do inicio do pacote (em amostras)
Require: δ ▷ Threshold do inicio do pacote

Inicializa $features\_matrix \leftarrow [\,]$ ▷ Matriz de características Agregadas para todos os dispositivos
for cada dispositivo em $device\_list$ **do**
    Inicializa: $device\_features \leftarrow [\,]$
    for cada arquivo em $device\_files$ **do**
        Abrir o arquivo e carregar IQ_data
        Calcula $x[j]$ para todo j: $x[j] \leftarrow |IQ\_data[j]| \cdot \cos(\text{angulo}(IQ\_data[j]))$
        Inicializa: $i \leftarrow 1$, $p \leftarrow 0$
        Calcula: $n\_frames \leftarrow \lfloor (|x|/\text{packet\_size}) \rfloor - 1$
        while $p < n\_frames$ **do**
            while média$(|x[i : i + 1000]|) < \delta$ **do**
                Incrementa i ▷ Encontra o ínicio do pacote
            end while
            $packet \leftarrow x[i + \text{offset} : i + \text{packet\_size}]$ ▷ Extrai o pacote
            $features \leftarrow \text{catch22}(packet[1 : \text{window\_size}])$ ▷ Extrai as características
            Adiciona $features$ e o rótulo do dispositivo a $device\_features$
            Atualiza: $i \leftarrow i + \text{step\_size}$
            Incrementa p
        end while
    end for
    Adiciona $device\_features$ a $features\_matrix$
end for
return $features\_matrix$

Como citado anteriormente a variação do tamanho da janela de análise foi crucial

para determinar o modelo com melhor acurácia, então para isso foi criado grupos de análise onde variou-se as janelas em (10.000, 20.000, 40.000 e 80.000 amostras) e testou a combinação de dados provenientes de múltiplos dias de coleta. A Tabela 6 resume os cenários avaliados.

Tabela 6 – Cenários avaliados

Cenário	**Dataset**	**Janela de análise**
Grupo 1	Dataset dia 1	20.000
Grupo 2	Dataset dia 2	20.000
Grupo 3	Dataset dia 3	20.000
Grupo 4	Dataset dia 4	20.000
Grupo 5	Dataset dia 5	20.000
Grupo 6	Dataset dias 1 a 5 combinados	20.000
Grupo 7	Dataset dia 1	10.000
Grupo 8	Dataset dia 1	40.000
Grupo 9	Dataset dia 1	80.000

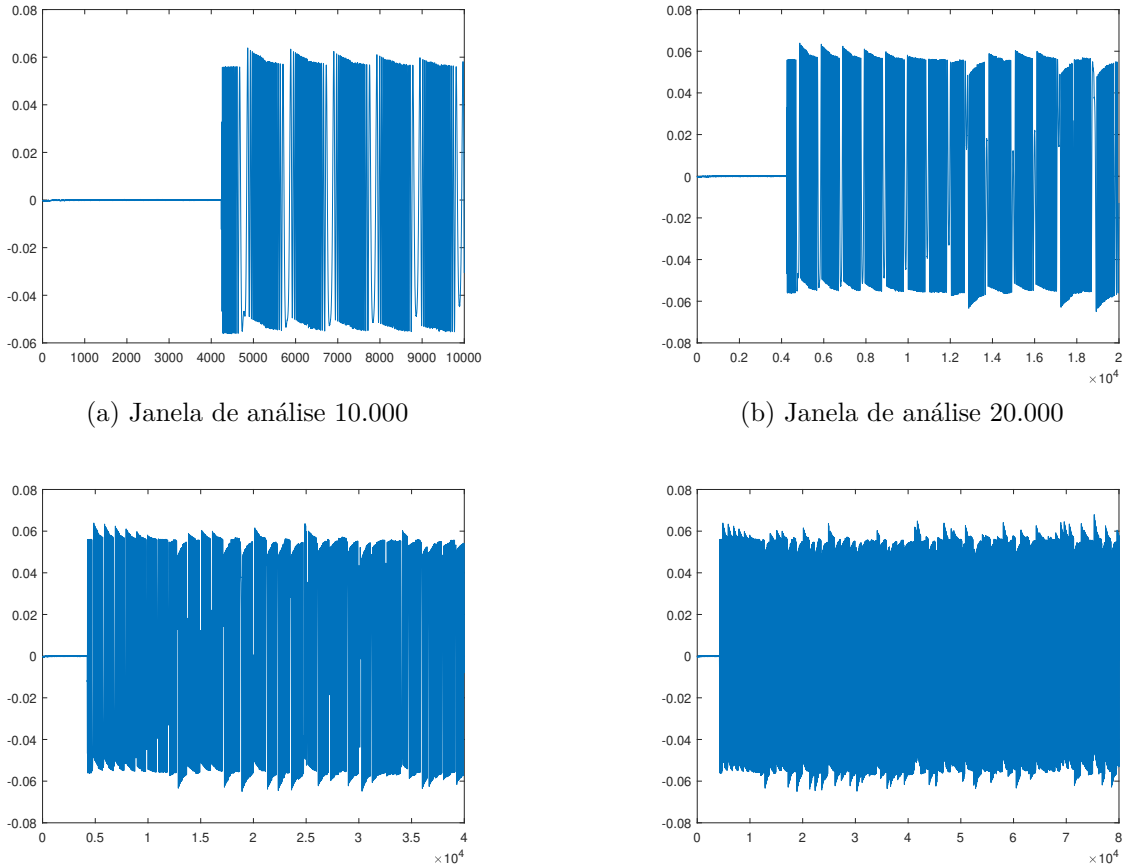
A variável *window_analysis*, utilizada neste trabalho e nos códigos de processamento, representa o número de amostras discretas selecionadas a partir de cada pacote LoRa para o processo de extração de características por meio da biblioteca *Catch22*. Essas amostras encontram-se no domínio do tempo e, portanto, não possuem unidade temporal direta, pois correspondem à quantidade de pontos extraídos da sequência de sinais *in-phase* e *quadrature* (IQ).

Considerando que os sinais foram coletados com uma taxa de amostragem de 1 milhão de amostras por segundo (1 MS/s), é possível converter a variável *window_analysis* em uma janela temporal equivalente, conforme expresso na Equação 4.1.

$$\text{Tempo da janela (s)} = \frac{\text{window\_analysis}}{\text{frequência de amostragem}} \tag{4.1}$$

Com base nessa relação, as janelas utilizadas nos experimentos — 10.000, 20.000, 40.000 e 80.000 amostras — correspondem, respectivamente, a 10 ms, 20 ms, 40 ms e 80 ms de sinal contínuo analisado por transmissão.

A variação do tamanho da janela de análise teve como objetivo investigar o impacto da quantidade de amostras sobre o processo de *fingerprinting*. Uma janela maior pode conter mais informações discriminativas do sinal, mas também tende a incluir ruídos ou redundâncias, enquanto uma janela muito curta pode não capturar características suficientes para a correta identificação do transmissor. A Figura 5 ilustra as diferenças entre as janelas de análise extraídas de um mesmo sinal.



(a) Janela de análise 10.000

(b) Janela de análise 20.000

(c) Janela de análise 40.000

(d) Janela de análise 80.000

Figura 5 – Diferentes janelas de análise extraídas de um mesmo sinal

Fonte: o autor.

A decisão de restringir a análise a um curto período do sinal, isto é, uma fração do preâmbulo que tem como objetivo não apenas reduzir o tempo de processamento, mas também aumentar a eficiência na extração de características únicas. Essa escolha se fundamenta no fato de que as maiores discrepâncias entre os sinais ocorrem no instante inicial de transmissão (T0), quando o transmissor é ativado e inicia a geração do preâmbulo, refletindo imperfeições específicas de hardware. Ao analisarmos a Tabela 1 pode-se observar a predominância da janela de 20.000 amostras, isso se deve ao fato que ela teve o melhor resultado perante as demais, conforme os dados apresentados no Capítulo 5.

4.4 Modelos de Aprendizagem de Máquina

Para o desenvolvimento e teste dos modelos de aprendizado de máquina, foi utilizado o ambiente de execução *Google Colaboratory Pro*, que oferece suporte a recursos de processamento otimizados e maior capacidade de memória. As especificações técnicas do ambiente utilizado estão descritas na Tabela 7.

Tabela 7 – Especificações técnicas do ambiente Google Colab Pro.

Processador	Intel(R) Xeon(R) CPU @ 2.20GHz (7ª geração)
Núcleos	4 Cores
Memória Cache	56,32 MB
Memória RAM	≈ 53,5 GB

Diferentemente de abordagens anteriores, que se baseavam predominantemente em redes neurais profundas, técnicas de alta complexidade e elevado custo computacional, este trabalho prioriza métodos de menor custo e maior eficiência, sem comprometer o desempenho de classificação. Assim, optou-se por empregar algoritmos baseados em *ensemble learning*, conhecidos por sua capacidade de combinar múltiplos classificadores de forma cooperativa, reduzindo variância e aumentando a precisão final do modelo.

O primeiro passo consistiu na organização dos algoritmos de leitura e pré-processamento dos arquivos *.csv* gerados no ambiente MATLAB, que foram exportados e armazenados no Google Drive para facilitar sua importação pelo Colab. Após a importação, as bibliotecas necessárias foram carregadas e os modelos configurados para execução automatizada dos experimentos.

Inicialmente, diferentes classificadores individuais foram avaliados, como *Decision Tree* e *K-Nearest Neighbors* (KNN). No entanto, verificou-se que esses modelos apresentaram desempenho inferior em comparação com os métodos baseados em *ensemble learning*. Dessa forma, este estudo concentrou-se na avaliação de cinco modelos principais: **Decision Tree (DT)** - Apenas para comparação, **Random Forest (RF)**, **Gradient Boosting Machine (GBM)**, **LightGBM** e **XGBoost**.

Para cada modelo, foi realizada uma etapa de ajuste fino dos hiperparâmetros utilizando o método *Grid Search Cross Validation (Grid-SearchCV)*, com validação cruzada de 10 dobras. Esse processo foi conduzido inicialmente sobre o conjunto de dados referente ao dia 1, selecionado como base de calibração.

Ao final de cada experimento, foram calculadas as matrizes de acurácia e de confusão, que possibilitaram avaliar o desempenho de cada classificador quanto à capacidade de identificação correta dos dispositivos. As análises quantitativas e qualitativas desses resultados são apresentadas detalhadamente no Capítulo 5.

4.4.1 Configuração Experimental

Os experimentos foram implementados em Python, utilizando as bibliotecas *Scikit-learn*, *XGBoost*, *LightGBM* e *TensorFlow*, responsáveis pelas etapas de treinamento, validação e avaliação dos modelos. O principal objetivo foi comparar o desempenho de diferentes algoritmos de classificação aplicados às características estatísticas extraídas pela

biblioteca *Catch22*, identificando o método mais adequado ao problema de *fingerprinting* de dispositivos LoRa.

Modelos Individuais

Foram testados classificadores tradicionais, configurados da seguinte forma:

- **Árvore de Decisão (Decision Tree):** critério de entropia como métrica de divisão e semente aleatória fixa para reprodutibilidade.
- **K-Nearest Neighbors (K-NN):** versões com pesos uniformes e ponderados por distância, com $k = 3$ vizinhos.
- **Rede Neural Multicamadas (MLPClassifier):** camada oculta de 32 neurônios, função de ativação logística, taxa de aprendizado inicial de 0,001, *momentum* de 0,8 e até 10.000 iterações.
- **Naive Bayes (GaussianNB):** suavização de variância com parâmetro padrão de 10^{-9}.
- **Máquina de Vetores de Suporte (SVM):** otimização de parâmetros via *GridSearchCV*, explorando os núcleos *poly* e *rbf* com múltiplas combinações de C e γ.

Modelos de Ensemble

Em seguida, foram avaliados diferentes algoritmos baseados em árvores e gradiente de *boosting*, amplamente utilizados em tarefas de classificação supervisionada:

- **Random Forest (RF):** 300 estimadores, profundidade máxima de 30, divisão mínima de 5 amostras e ausência de *bootstrap*.
- **Gradient Boosting Machine (GBM):** taxa de aprendizado de 0,1, profundidade máxima de 7, 200 estimadores e amostragem de 80%.
- **LightGBM:** execução em GPU, taxa de aprendizado de 0,1, profundidade máxima de 7, 200 árvores e até 50 folhas por árvore.
- **XGBoost:** método de construção de árvores baseado em histograma (*tree_method='hist'*), 200 estimadores e profundidade máxima de 5.

Treinamento e Validação

Cada conjunto de dados foi dividido em 70% para treinamento e 30% para teste, utilizando amostragem estratificada para preservar a proporção de classes. Todos os modelos foram treinados e avaliados separadamente sobre os grupos definidos na Tabela 1.

A principal métrica de avaliação foi a **acurácia**, calculada conforme a Equação 4.2. Adicionalmente, foram geradas **matrizes de confusão** para análise visual dos resultados de classificação.

$$\text{Acurácia} = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.2}$$

em que TP e TN representam, respectivamente, os acertos verdadeiros positivos e negativos, e FP e FN correspondem aos erros de classificação.

Geração e Organização dos Resultados

Os resultados de cada execução foram salvos automaticamente em arquivos de texto e imagens. Foram geradas:

- Tabelas de desempenho contendo a acurácia obtida por cada modelo.
- Gráficos de importância das 24 características extraídas via Catch22, indicando o impacto de cada variável na decisão do modelo.
- Matrizes de confusão dos modelos com melhor e pior desempenho, apresentadas nas Figuras 10 e 11, localizadas no Anexo A.

A Figura 6 apresenta uma visão geral do fluxo experimental desenvolvido, desde a leitura e pré-processamento dos dados até a geração dos resultados finais.

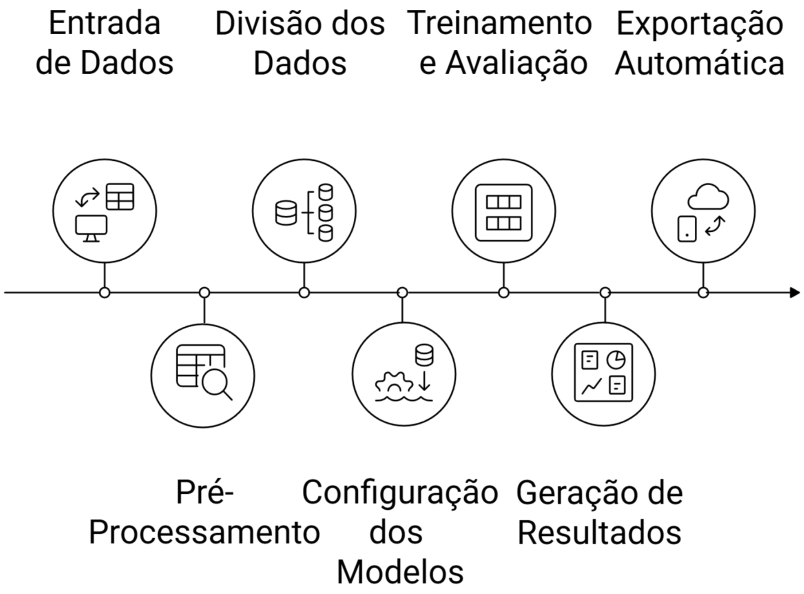


Figura 6 – Fluxo geral dos experimentos realizados.

Fonte: o autor.

Essa estrutura experimental possibilitou a comparação sistemática de múltiplos classificadores sob as mesmas condições, permitindo avaliar a robustez e a eficiência de cada método no processo de identificação de dispositivos LoRa com base em *fingerprinting*.

5 Análise e Interpretação dos Resultados

Este capítulo aprofunda a discussão sobre os resultados obtidos nos experimentos conduzidos, considerando tanto as métricas quantitativas apresentadas anteriormente quanto as observações qualitativas a respeito do comportamento dos modelos de aprendizagem de máquina aplicados à identificação de dispositivos LoRa.

5.1 Análise de Complexidade Computacional

A baixa complexidade da técnica proposta é um diferencial central deste trabalho frente ao estado da arte. Enquanto métodos baseados em Redes Neurais Convolucionais (CNN) realizam o aprendizado de características de ponta a ponta (*end-to-end*), exigindo milhões de operações de ponto flutuante (FLOPs) e memória RAM significativa para o armazenamento de tensores, a abordagem com *Catch22* simplifica o problema em duas etapas eficientes:

1. **Extração de Características:** A maioria das 22 métricas do *Catch22* possui complexidade de tempo linear ou quase linear em relação ao comprimento da série temporal (N). Isso permite que janelas de 20.000 amostras sejam processadas rapidamente em implementações baseadas em C.
2. **Classificação:** Ao reduzir o sinal bruto de 20.000 dimensões para um vetor de apenas 22 características estatísticas, o classificador final (como o LightGBM ou Random Forest) opera sobre uma matriz extremamente leve.

Comparativamente, modelos como o de (SHEN et al., 2021), embora possam atingir acurácias superiores, demandam um custo computacional que inviabiliza o uso em *gateways* IoT de baixo custo. A técnica aqui apresentada atinge um equilíbrio ótimo (*trade-off*), mantendo uma acurácia próxima a 87% com uma pegada computacional ordens de grandeza menor que modelos de *Deep Learning*.

Além da eficiência algorítmica, a técnica proposta destaca-se pela sua independência de unidades de processamento gráfico (GPUs). Enquanto o estado da arte em *fingerprinting* de rádio frequentemente recorre a Redes Neurais Profundas (como ResNet ou VGG), estas demandam arquiteturas de hardware com suporte a processamento massivamente paralelo para viabilizar a inferência em tempo razoável.

A não utilização de GPUs nesta metodologia justifica-se por três fatores críticos:

- **Viabilidade Econômica em Gateways IoT:** Dispositivos de borda e *gateways* LoRaWAN são projetados para baixo custo. A inclusão de GPUs ou aceleradores de IA (como TPUs) elevaria o custo unitário da infraestrutura, inviabilizando implantações em larga escala.
- **Eficiência Energética:** O processamento em CPU utilizando a biblioteca Catch22 (implementada em C) consome significativamente menos energia por amostra processada do que o acionamento de núcleos de processamento gráfico, o que é vital para dispositivos que operam em regimes de restrição energética.
- **Redução da Latência de Pipeline:** Em modelos de *Deep Learning*, a latência não advém apenas do cálculo, mas também do *overhead* de transferência de dados entre a memória do sistema e a memória de vídeo (VRAM). Ao manter todo o fluxo de extração de características e classificação na CPU, elimina-se esse gargalo de comunicação, permitindo uma resposta mais ágil na identificação do dispositivo.

Em suma, a transição de modelos de aprendizado profundo (Deep Learning) para uma abordagem baseada em características canônicas e aprendizado de máquina tradicional não representa apenas uma simplificação estatística, mas uma escolha de engenharia voltada à escalabilidade. Ao eliminar a dependência de hardware especializado e reduzir a demanda por recursos de memória e energia, a técnica proposta viabiliza a implementação da segurança por *fingerprinting* diretamente na borda da rede. Isso permite que a identificação e autenticação de dispositivos LoRa ocorram de forma sustentável, atendendo aos requisitos de baixo custo e longa vida útil inerentes ao ecossistema da Internet das Coisas.

5.2 Desempenho Comparativo entre Modelos

Os resultados de acurácia média obtidos em cada um dos cenários são apresentados na Tabela 8. Foram avaliados quatro modelos baseados em *ensemble learning*: *Gradient Boosting Machine* (GBM), *LightGBM*, *Random Forest* (RF) e *XGBoost*. Adicionalmente, foram considerados classificadores singulares como *K-Nearest Neighbors* (KNN), *Multi-Layer Perceptron* (MLP), *Naïve Bayes* (NB) e *Decision Tree* (DT). Dentre esses, o *Decision Tree* apresentou o melhor desempenho médio e foi, portanto, incluído na Tabela de resultados como referência comparativa. Cada grupo corresponde a uma configuração distinta de tamanho de janela ou estratégia de agrupamento, conforme descrito na Tabela 1.

Tabela 8 – Análise de desempenho — acurácia média dos modelos por grupo de experimento

Cenário	**DT**	**GBM**	**LightGBM**	**RF**	**XGBoost**
Grupo 1	69,33%	82,35%	**83,43%**	82,30%	82,75%
Grupo 2	72,84%	85,02%	**87,15%**	86,38%	86,38%
Grupo 3	67,77%	80,31%	**82,44%**	81,30%	82,01%
Grupo 4	74,81%	83,80%	**85,79%**	84,57%	85,33%
Grupo 5	68,43%	80,43%	**82,67%**	81,30%	81,45%
Grupo 6	57,32%	73,35%	**75,40%**	73,54%	74,03%
Grupo 7	51,77%	62,24%	**63,06%**	62,81%	62,61%
Grupo 8	70,44%	82,21%	**83,38%**	81,93%	82,87%
Grupo 9	59,94%	75,69%	76,96%	**77,05%**	76,14%

Fonte: o autor.

Observa-se que o modelo *LightGBM* apresentou as melhores médias de acurácia na maioria dos cenários, atingindo o desempenho máximo de **87,15%** no Grupo 2 (janela de 20.000 amostras). O *XGBoost* manteve resultados próximos, confirmando a robustez das abordagens de *boosting* em comparação com os métodos baseados em *bagging*, como o *Random Forest*.

Em contrapartida, o desempenho do *Decision Tree* mostrou-se consistentemente inferior, reforçando a importância dos métodos *ensemble* para a redução de variância e o aumento da estabilidade dos modelos. Esses mecanismos, baseados na combinação de múltiplas árvores, permitem capturar interações complexas entre variáveis e suavizar decisões locais, o que explica o ganho de desempenho observado.

Outro ponto relevante é a sensibilidade dos modelos às condições do conjunto de dados. O Grupo 6, que combinou dados de múltiplos dias de coleta, demonstrou estabilidade intermediária, com acurácias superiores a 70%, evidenciando a capacidade dos modelos em generalizar padrões temporais mesmo sob variabilidade de canal. Já os Grupos 7, 8 e 9, que exploram diferentes janelas de observação, apresentaram leve queda de desempenho, um indicativo de que as variações absolutas de amplitude e fase, responsáveis pelas assinaturas físicas dos dispositivos, podem sofrer interferências conforme o número de amostras analisadas. Esse efeito reforça a importância de selecionar janelas curtas e representativas suficientes, centradas no transiente inicial, para maximizar a discriminação entre transmissores.

A Figura 7 sintetiza visualmente esses resultados, permitindo comparar o comportamento dos cinco modelos ao longo dos diferentes grupos experimentais.

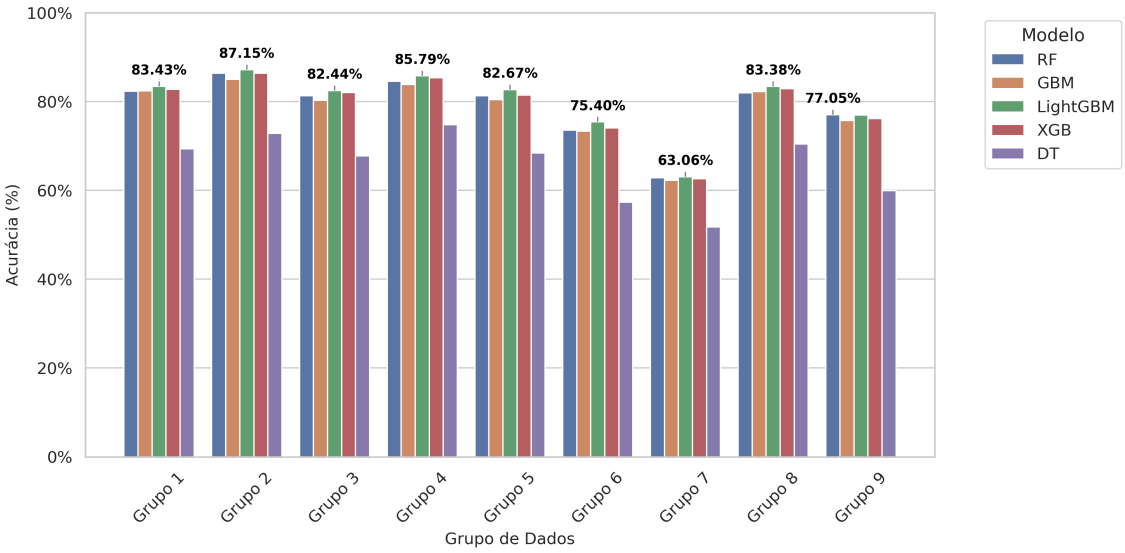


Figura 7 – Gráfico comparativo de Acurácias dos Modelos x Grupo de dados.

Fonte: o autor.

5.3 Influência dos Parâmetros de Configuração

A variação do parâmetro *window_analysis*, correspondente à quantidade de amostras por pacote analisado, exerceu influência direta sobre o desempenho dos classificadores. Como ilustrado na Figura 5, as janelas curtas capturam a fase inicial do sinal, onde se concentram as imperfeições de osciladores, o ruído de partida e as distorções de sincronismo, elementos fundamentais para a formação da assinatura de radiofrequência (*fingerprint*) de cada dispositivo.

Conforme demonstrado na Tabela 8, a janela de 20.000 amostras (20 ms) apresentou o melhor desempenho médio entre todos os experimentos, especialmente para os modelos baseados em *boosting*, como o *LightGBM* e o *XGBoost*. Esse resultado indica que esse intervalo de observação captura de forma equilibrada as características físicas relevantes do transiente inicial, sem introduzir redundância excessiva. Os cinco primeiros grupos (Grupos 1 a 5) representam dados coletados em dias distintos, mantendo a mesma configuração de janela de 20.000 amostras. Essa variação temporal permite avaliar a robustez dos modelos frente a mudanças naturais nas condições do ambiente e nos parâmetros físicos dos transmissores, como temperatura, umidade e pequenas flutuações nos osciladores locais. O comportamento relativamente estável das acurácias entre esses grupos indica que as características extraídas pelo conjunto *Catch22* preservam informações discriminativas consistentes ao longo do tempo, reforçando sua adequação para aplicações reais de identificação de dispositivos LoRa.

Por outro lado, a janela mais curta, de 10.000 amostras (10 ms), apresentou o pior desempenho geral, sugerindo que a quantidade de amostras analisadas não foi

suficiente para representar completamente o comportamento do transmissor durante o início da transmissão. Já as janelas mais longas (40.000–80.000 amostras) provocaram leve degradação da acurácia, possivelmente devido à inclusão de trechos estacionários do sinal, nos quais as assinaturas de hardware tornam-se menos pronunciadas e mais suscetíveis a ruído e interferência.

Esse comportamento está em conformidade com as observações de Zhang, Gong e Qian (2020), que relataram melhora de precisão ao restringir a análise à fase transiente do sinal, onde predominam as não linearidades e desvios de frequência específicos de cada transmissor.

5.4 Análise das Características Extraídas

A análise de importância das características foi conduzida com base na média normalizada das importâncias obtidas pelos modelos do tipo *ensemble*, considerando todos os grupos experimentais. O objetivo dessa análise foi identificar quais métricas do conjunto *Catch22* mais contribuíram para a tarefa de identificação de dispositivos LoRa, permitindo compreender a relevância física de cada atributo na diferenciação entre transmissores.

A Figura 8 apresenta a distribuição de importância relativa das 24 variáveis analisadas. Os resultados foram obtidos a partir da média ponderada das importâncias dos classificadores *LightGBM*, *XGBoost*, *Random Forest* e *Gradient Boosting*, normalizados em relação ao somatório total de cada grupo. Esse procedimento visou atenuar possíveis vieses de modelo, resultando em uma medida global de relevância das variáveis.

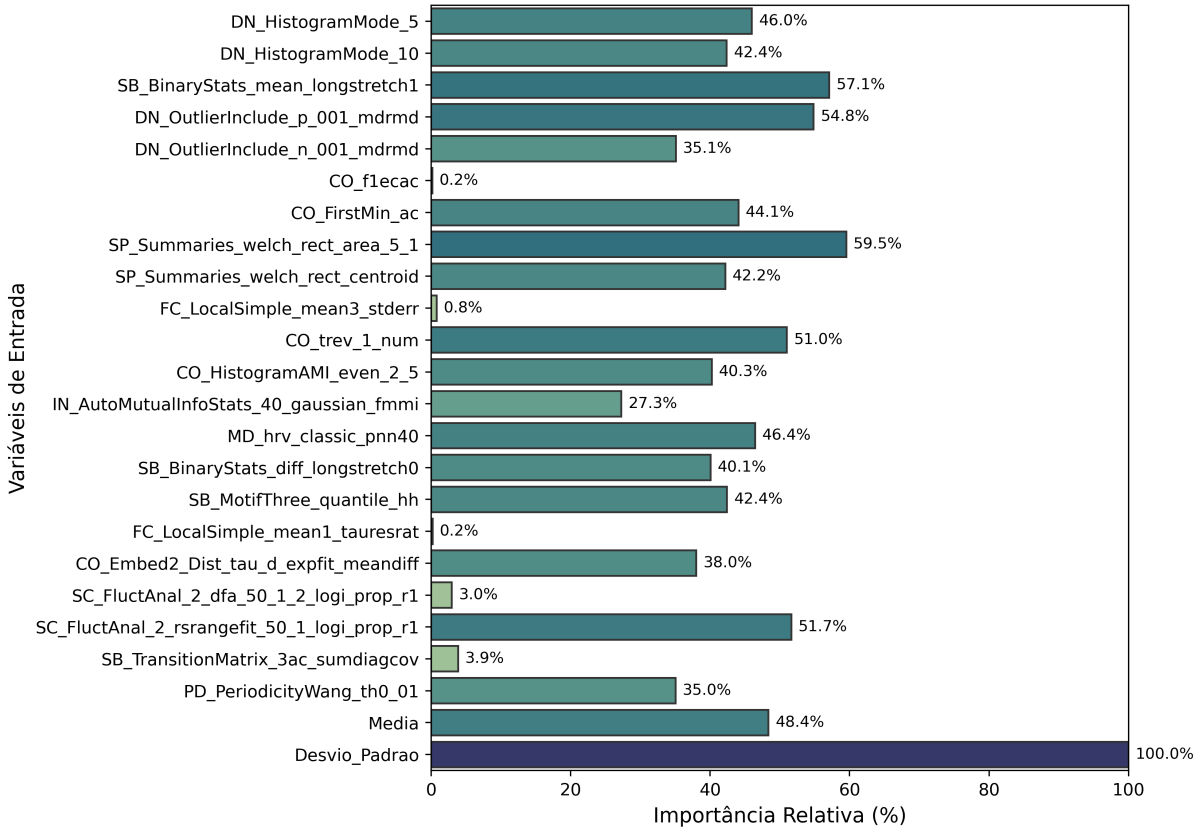


Figura 8 – Importância global das características (*Catch22*) — média entre todos os grupos experimentais.

Fonte: o autor.

Observa-se que as métricas `SB_BinaryStats_mean_longstretch1`, `DN_OutlierInclude_p_001_mdrmd` e `SP_Summaries_welch_rect_area_5_1` figuram entre as mais relevantes, indicando que a duração de trechos binários, a presença de valores atípicos e a energia espectral retangular têm papel significativo na diferenciação entre dispositivos. Essas métricas refletem variações de amplitude e irregularidades estatísticas no domínio do tempo e da frequência, manifestações diretas das imperfeições de hardware dos transmissores LoRa.

Outros atributos, como `CO_FirstMin_ac` (primeiro mínimo da autocorrelação) e `SC_FluctAnal_2_rsrangefit_50_1_logi_prop_r1`, também apresentaram importância considerável. Esses descritores estão associados à regularidade temporal e à persistência estatística dos sinais, fornecendo indícios sobre a estabilidade dos osciladores e a consistência de fase. Em contrapartida, métricas de natureza mais genérica, como `CO_f1ecac` e `FC_LocalSimple_mean1_tauresrat`, exibiram contribuições menores, sugerindo baixa capacidade discriminativa isoladamente.

A predominância das métricas relacionadas à densidade espectral e aos padrões binários de longa duração corrobora a hipótese de que os dispositivos LoRa podem ser

identificados a partir de assinaturas temporais e estatísticas simples, sem necessidade de decomposição espectral complexa ou arquiteturas profundas de aprendizado. Esse resultado reforça a viabilidade da abordagem proposta, que combina a extração leve de características pelo *Catch22* com classificadores de baixo custo computacional.

Além disso, o atributo `Desvio_Padrao` destacou-se como a variável de maior peso relativo, indicando que a dispersão estatística das amostras dentro da janela analisada carrega forte correlação com as imperfeições intrínsecas do transmissor. Isso se alinha ao comportamento esperado de sistemas de *fingerprinting*, nos quais pequenas flutuações na amplitude e fase são suficientes para distinguir emissores com circuitos analógicos levemente distintos.

Em síntese, os resultados dessa análise indicam que o conjunto de características do *Catch22*, mesmo sem transformações no domínio da frequência, é capaz de capturar nuances relevantes do comportamento físico dos transmissores. A predominância de atributos relacionados à irregularidade temporal e à variabilidade estatística demonstra que a identificação de dispositivos LoRa pode ser realizada com alta eficiência utilizando apenas um subconjunto reduzido de métricas, o que abre caminho para aplicações embarcadas e de tempo real em dispositivos IoT.

5.5 Desempenho Cruzado dos Modelos nos Diferentes Grupos de Dados

A Figura 9 apresenta um mapa de calor contendo as acurácias obtidas por cada modelo nos nove grupos de dados avaliados. Essa visualização evidencia o comportamento dos classificadores frente às variações entre os cenários experimentais, permitindo identificar padrões de estabilidade, sensibilidade e generalização.

Observa-se que os métodos baseados em gradiente, especialmente *LightGBM* e *XGBoost*, apresentam desempenho superior e relativamente estável na maior parte dos grupos, com acurácias típicas entre 0,82 e 0,87. Esses modelos demonstram boa capacidade de adaptação às diferenças entre os conjuntos, incluindo mudanças no dia de coleta, características do canal e variações inerentes aos dispositivos.

O *Random Forest* apresentou desempenho semelhante, embora com ligeira variação entre grupos, enquanto o *GBM* exibiu comportamento intermediário, com quedas perceptíveis em grupos mais desafiadores, como o Grupo 7. Já o *Decision Tree*, por ser um classificador mais simples, apresentou a maior oscilação e as menores acurácias, variando de 0,51 a 0,74, confirmando sua limitação em capturar variações finas das características extraídas.

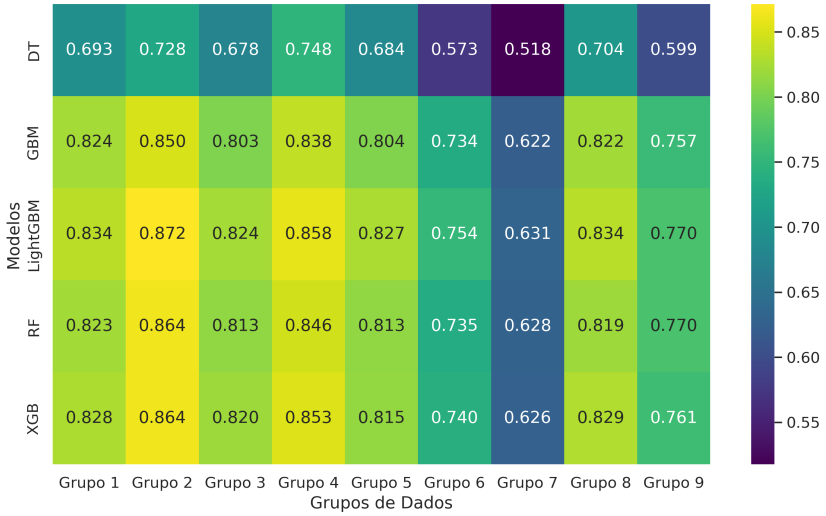


Figura 9 – Mapa de calor das acurácias obtidas por cada modelo em cada grupo de dados.

Fonte: o autor.

De maneira geral, os resultados indicam que os modelos mais robustos compartilham padrões de desempenho semelhantes entre grupos, sugerindo que as características derivadas via Catch22 preservam informações discriminativas relevantes dos sinais LoRa. Isso reforça a consistência do método proposto, mesmo frente a mudanças nas condições dos dados.

5.6 Discussão Geral e Considerações Finais

Os resultados apresentados nesta seção consolidam a validade da abordagem proposta neste trabalho. O desempenho médio obtido pelos modelos *ensemble*, com acurácias superiores a 80%, demonstra que a extração de características estatísticas do domínio temporal é suficiente para distinguir dispositivos LoRa de forma robusta e eficiente.

Em comparação com abordagens baseadas em redes neurais convolucionais (CNNs), que frequentemente exigem milhões de parâmetros e longos tempos de treinamento, os métodos testados aqui obtiveram resultados competitivos com uma fração do custo computacional. Enquanto modelos como o de Shen et al. (2021) atingem acurácias próximas a 97% com CNNs aplicadas a espectrogramas, o método proposto neste trabalho alcança desempenho comparável utilizando apenas 24 atributos estatísticos e classificadores leves, viabilizando sua aplicação em sistemas embarcados.

Por fim, destaca-se que a redução drástica da complexidade computacional não comprometeu a acurácia nem a interpretabilidade do modelo. Pelo contrário, a natureza estatística das características extraídas possibilita uma compreensão direta das propriedades físicas subjacentes a cada dispositivo, tornando o método proposto não apenas mais eficiente, mas também mais explicável e aplicável em cenários reais de autenticação e segurança em IoT.

6 Conclusão

Com o aumento do acesso a internet surgiu a necessidade de interligar diferentes aparelhos e sistemas, com o intuito de ajudar a vida humana no cotidiano, criando sistemas de rede chamados de Internet das Coisas (IoT). As aplicações de IoT são encontradas em diversas áreas, como agricultura, indústrias, cidades e até mesmo residências. Ao considerar tais aplicações, uma das áreas englobadas discute a segurança nos meios de comunicação dos sistemas IoT.

Sendo assim, este estudo propôs um método de baixa complexidade para identificação de impressão digital de RF em redes LoRa, utilizando amostras IQ e extração de características com a biblioteca *Catch22*. O objetivo foi enfrentar as limitações de estudos existentes que se apoiam em modelos de *deep learning*, os quais são computacionalmente exigentes, demandam grande volume de dados de treinamento e são demorados, com o intuito de auxiliar nos processos que englobam segurança em sistemas embarcados. Os resultados obtidos demonstram que o método proposto oferece um *trade-off* equilibrado entre precisão e eficiência, tornando-o adequado para implantação em ambientes IoT com recursos restritos.

As questões de pesquisa podem ser respondidas da seguinte forma:

Q1: Como identificar dispositivos LoRa de forma eficiente e precisa, considerando as restrições computacionais em sistemas IoT? A abordagem realizada obteve cerca de 80% de acurácia na classificação utilizando apenas breves segmentos iniciais do sinal LoRa recebido, reduzindo o overhead computacional em comparação com métodos baseados em CNN aplicados a espectrogramas. Isso viabiliza o processamento em tempo real em sistemas embarcados, configurando uma alternativa prática para implantações IoT em larga escala.

Q2: De que maneira métricas analíticas, como as extraídas pelo *Catch22*, podem constituir alternativa viável às abordagens baseadas em CNN, que demandam elevado uso de recursos computacionais? Ao contrário de modelos CNN, que exigem treinamento extensivo e GPUs de alto desempenho, a *Catch22* extrai 22 características-chave de séries temporais com esforço computacional mínimo. Essa estratégia reduz tanto o tempo de processamento quanto o consumo de energia, possibilitando uma identificação de dispositivos mais rápida e eficiente, sem sacrificar significativamente a acurácia.

Q3: Quais melhorias de segurança e monitoramento podem ser alcançadas com a abordagem proposta para identificação de dispositivos em redes IoT? O método proposto fortalece a segurança de redes IoT ao permitir autenticação de dispositivos a baixo custo por meio de *fingerprinting* de RF. Isso provê proteção contra *spoofing* e acesso não autorizado,

além de aprimorar as capacidades de monitoramento em tempo real dos sistemas de rede. Além disso, a adaptabilidade do modelo viabiliza sua aplicação em ambientes dinâmicos, nos quais as características dos dispositivos podem evoluir ao longo do tempo.

De modo geral, este estudo demonstrou que o *fingerprinting* de RF baseada em *Catch22* e técnicas de *machine learning* constituem uma solução viável para a identificação segura e escalável de dispositivos LoRa. Trabalhos futuros poderão explorar otimizações adicionais na seleção de características e avaliar o desempenho do modelo em cenários externos com maior variabilidade de sinal.

Referências

ADELANTADO, Ferran; VILAJOSANA, Xavier; TUSET-PEIRO, Pere; MARTINEZ, Borja; MELIA-SEGUI, Joan; WATTEYNE, Thomas. Understanding the limits of lorawan. *IEEE Communications Magazine*, Institute of Electrical and Electronics Engineers Inc., v. 55, 2017. Citado na página 30.

AERNOUTS, Michiel; BERKVENS, Rafael; VLAENDEREN, Koen Van; WEYN, Maarten. *Sigfox and LoRaWAN Datasets for Fingerprint Localization in Large Urban and Rural Areas*. Zenodo, 2018. Disponível em: <https://zenodo.org/record/1405484>. Citado na página 46.

Agência Nacional de Telecomunicações (ANATEL). *Ato n.º 14448, de 23 de dezembro de 2017: Requisitos Técnicos para Avaliação da Conformidade de Produtos para Telecomunicações*. 2017. Acesso em: 24 fev. 2025. Disponível em: <https://informacoes.anatel.gov.br/legislacao/atos-de-certificacao-de-produtos/2017/1139-ato-14448>. Citado 3 vezes nas páginas 29, 31 e 32.

AL-FUQAHA, Ala; GUIZANI, Mohsen; MOHAMMADI, Mehdi; ALEDHARI, Mohammed; AYYASH, Moussa. Internet of things: A survey on enabling technologies, protocols, and applications. *IEEE Communications Surveys and Tutorials*, Institute of Electrical and Electronics Engineers Inc., v. 17, 2015. Citado 2 vezes nas páginas 20 e 24.

ARIF, Muhammad; MAYA, Juan Augusto; ANANDAN, Narendiran; PéREZ, Dailys Arronde; TONELLO, Andrea M.; ZANGL, Hubert; RINNER, Bernhard. Resource-efficient ubiquitous sensor networks for smart agriculture: A survey. *IEEE Access*, v. 12, p. 193332–193364, 2024. Citado na página 20.

BASHA, Nora; HAMDAOUI, Bechir; SIVANESAN, Kathiravetpillai; GUIZANI, Mohsen. Channel-resilient deep-learning-driven device fingerprinting through multiple data streams. *IEEE Open Journal of the Communications Society*, Institute of Electrical and Electronics Engineers Inc., v. 4, p. 118–133, 2023. ISSN 2644125X. Citado 3 vezes nas páginas 21, 41 e 42.

BENGIO, Yoshua; COURVILLE, Aaron; VINCENT, Pascal. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 35, 2013. Citado na página 37.

BREIMAN, Leo. Random forests. *Machine Learning*, v. 45, n. 1, p. 5–32, 2001. Citado na página 37.

BREIMAN, Leo; FRIEDMAN, Jerome; OLSHEN, Richard; STONE, Charles. *Classification and Regression Trees*. [S.l.]: Wadsworth International Group, 1984. Citado na página 37.

CHEN, Tianqi; GUESTRIN, Carlos. Xgboost: A scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. [S.l.: s.n.], 2016. p. 785–794. Citado na página 37.

CHILLET, Alice; GERZAGUET, Robin; DESNOS, Karol; GAUTIER, Matthieu; LOHAN, Elena Simona; NOGUES, Erwan; VALKAMA, Mikko. Understanding radio frequency fingerprint identification with rifyfi virtual databases. *IEEE Open Journal of the Communications Society*, Institute of Electrical and Electronics Engineers Inc., v. 5, p. 3735–3752, 2024. ISSN 2644125X. Citado 2 vezes nas páginas 40 e 42.

DING, Jie; NEMATI, Mahyar; RANAWEERA, Chathurika; CHOI, Jinho. Iot connectivity technologies and applications: A survey. *IEEE Access*, v. 8, p. 67646–67673, 2020. Citado na página 20.

DOTSON, Aaron D.; CENEK, Martin; MICHAELSON, Gregory. Open-source iot framework for mobile household water reuse system. *IEEE Global Humanitarian Technology Conference*, p. 1–5, 2019. Citado na página 20.

ELMAGHBUB, Abdurrahman; HAMDAOUI, Bechir. Comprehensive rf dataset collection and release: A deep learning-based device fingerprinting use case. In: *2021 IEEE Globecom Workshops, GC Wkshps 2021 - Proceedings*. [S.l.]: Institute of Electrical and Electronics Engineers Inc., 2021. ISBN 9781665423908. Citado 7 vezes nas páginas 21, 22, 39, 42, 46, 47 e 48.

ESCOBAR, Juan José López; FONDO-FERREIRO, Pablo; GONZáLEZ-CASTAñO, Francisco Javier; GIL-CASTIñEIRA, Felipe. *LoRa signal quality and GPS positioning time series dataset*. Zenodo, 2023. Disponível em: <https://zenodo.org/records/7919213>. Citado na página 45.

FARHAD, Arshad; PYUN, Jae-Young. Lorawan meets ml: A survey on enhancing performance with machine learning. *Sensors*, v. 23, n. 15, 2023. ISSN 1424-8220. Disponível em: <https://www.mdpi.com/1424-8220/23/15/6851>. Citado na página 30.

FAROOQ, Muhammad; WASEEM, Muhammad; MAZHAR, Sadia; KHAIRI, Anjum; KAMAL, Talha. A review on internet of things (iot). *International Journal of Computer Applications*, v. 113, p. 1–7, 03 2015. Citado na página 25.

FENG, Justin; ZHAO, Tianyi; SARKAR, Shamik; KONRAD, Dominic; JACQUES, Timothy; CABRIC, Danijela; SEHATBAKHSH, Nader. Fingerprinting iot devices using latent physical side-channels. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, Association for Computing Machinery, New York, NY, USA, v. 7, n. 2, jun. 2023. Disponível em: <https://doi.org/10.1145/3596247>. Citado na página 34.

FRIEDMAN, Jerome H. Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, v. 29, n. 5, p. 1189–1232, 2001. Citado na página 37.

FUKUNAGA, K. *Introduction to Statistical Pattern Recognition*. [S.l.]: Academic Press, 1990. Citado na página 36.

GALAR, Mikel; FERNANDEZ, Alberto; BARRENECHEA, Edurne; BUSTINCE, Humberto; HERRERA, Francisco. A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches. *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, v. 42, 2012. Citado na página 37.

GASKIN, Jared; ELMAGHBUB, Abdurrahman; HAMDAOUI, Bechir; WONG, Weng Keen. Deep learning model portability for domain-agnostic device fingerprinting.

IEEE Access, Institute of Electrical and Electronics Engineers Inc., v. 11, p. 86801–86823, 2023. ISSN 21693536. Citado 3 vezes nas páginas 21, 41 e 42.

HASSAN, Qusay F. *Internet of things A to Z : technologies and applications*. 1. ed. Wiley-IEEE Press, 2018. ISBN 9781119456735; 1119456738. Disponível em: <libgen.li/file.php?md5=f34f806d8a67145e118e22843ee678e1>. Citado 2 vezes nas páginas 27 e 28.

HAXHIBEQIRI, Jetmir; POORTER, Eli De; MOERMAN, Ingrid; HOEBEKE, Jeroen. A survey of lorawan for iot: From technology to application. *Sensors*, v. 18, n. 11, 2018. ISSN 1424-8220. Disponível em: <https://www.mdpi.com/1424-8220/18/11/3995>. Citado na página 30.

HOSSAIN, Mahmud; KAYAS, Golam; HASAN, Ragib; SKJELLUM, Anthony; NOOR, Shahid; ISLAM, S. M.Riazul. *A Holistic Analysis of Internet of Things (IoT) Security: Principles, Practices, and New Perspectives*. [S.l.]: Multidisciplinary Digital Publishing Institute (MDPI), 2024. Citado 3 vezes nas páginas 33, 34 e 35.

IDRIS, Sadiq; KARUNATHILAKE, Thenuka; FöRSTER, Anna. Survey and comparative study of lora-enabled simulators for internet of things and wireless sensor networks. *Sensors*, v. 22, n. 15, 2022. ISSN 1424-8220. Disponível em: <https://www.mdpi.com/1424-8220/22/15/5546>. Citado 3 vezes nas páginas 29, 30 e 32.

IEEE. *IEEE Standard for Low-Rate Wireless Networks–Amendment 2: Low Power Wide Area Network (LPWAN) Extension to the Low-Energy Critical Infrastructure Monitoring (LECIM) Physical Layer (PHY)*. 2020. Citado 3 vezes nas páginas 20, 28 e 29.

JABBAR, Waheb A.; SUBRAMANIAM, Thanasrii; ONG, Andre Emelio; SHU'IB, Mohd Iqmal; WU, Wenyan; de Oliveira, Mario A. Lorawan-based iot system implementation for long-range outdoor air quality monitoring. *Internet of Things*, v. 19, p. 100540, 2022. ISSN 2542-6605. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2542660522000427>. Citado na página 30.

KE, Guolin; MENG, Qi; FINLEY, Thomas; WANG, Taifeng; CHEN, Wei; MA, Weidong; YE, Qiwei; LIU, Tie-Yan. Lightgbm: A highly efficient gradient boosting decision tree. In: *Advances in Neural Information Processing Systems (NeurIPS)*. [S.l.: s.n.], 2017. p. 3146–3154. Citado na página 38.

LIU, Y.; CHENG, C.; GU, T.; JIANG, T.; LI, X. Scheme smart grid. *Smart Grid*, v. 16, n. 3, p. 836–842, 2016. Citado na página 21.

LORA® ALLIANCE. *A technical overview of LoRa ® and LoRaWAN ™ What is it?* EUA, 2015. Citado 3 vezes nas páginas 20, 30 e 32.

LOUNIS, Karim; ZULKERNINE, Mohammad. Attacks and defenses in short-range wireless technologies for iot. *IEEE Access*, v. 8, p. 88892–88932, 2020. Citado na página 20.

LUBBA, C. H.; SETHI, S. S.; KNAUTE, P.; SCHULTZ, S. R.; FULCHER, B. D.; JONES, N. S. catch22: Canonical time-series characteristics. *Data Mining and Knowledge Discovery*, v. 33, p. 1821–1852, 2019. Citado 4 vezes nas páginas 22, 45, 48 e 50.

MANSOUR, Mohammad; GAMAL, Amal; AHMED, Ahmed I.; SAID, Lobna A.; ELBAZ, Abdelmoniem; HERENCSAR, Norbert; SOLTAN, Ahmed. Internet of things: A comprehensive overview on protocols, architectures, technologies, simulation tools, and future directions. *Energies*, v. 16, n. 8, 2023. ISSN 1996-1073. Disponível em: <https://www.mdpi.com/1996-1073/16/8/3465>. Citado 3 vezes nas páginas 24, 25 e 26.

MATHWORKS, Inc. The. *MATLAB version: 24.1.0.2603908 (R2024a)*. 2024. Accessed: August 01, 2024. Disponível em: <https://www.mathworks.com>. Citado na página 50.

MITCHELL, Tom M. *Machine Learning*. [S.l.]: McGraw-Hill Science/Engineering/Math, 1997. Citado 2 vezes nas páginas 36 e 37.

MUMTAZ, Shahid; ALSOHAILY, Ahmed; PANG, Zhibo; RAYES, Ammar; TSANG, Kim Fung; RODRIGUEZ, Jonathan. Massive internet of things for industrial applications: Addressing wireless iiot connectivity challenges and ecosystem fragmentation. *IEEE Industrial Electronics Magazine*, v. 11, n. 1, p. 28–33, 2017. Citado na página 20.

OLIVEIRA, Luiz; RODRIGUES, Joel J. P. C.; KOZLOV, Sergei A.; RABêLO, Ricardo A. L.; ALBUQUERQUE, Victor Hugo C. de. Mac layer protocols for internet of things: A survey. *Future Internet*, v. 11, n. 1, 2019. ISSN 1999-5903. Disponível em: <https://www.mdpi.com/1999-5903/11/1/16>. Citado na página 25.

OPTIZ, David; MACLIN, Richard. Popular ensemble methods: An empirical study. *Journal of Artificial Intelligence Research*, v. 11, 1999. Citado na página 37.

PANDANGAN, Zaeefa A.; TALAMPAS, Marc Caesar R. Hybrid lorawan localization using ensemble learning. *GIoTS 2020 - Global Internet of Things Summit, Proceedings*, Institute of Electrical and Electronics Engineers Inc., 6 2020. Citado na página 38.

POVALAC, Ales; KRAL, Jan. *LoRaWAN Traffic Analysis Dataset*. Zenodo, 2023. Disponível em: <https://zenodo.org/records/7919213>. Citado na página 45.

SCHILLER, Eryk; AIDOO, Andy; FUHRER, Jara; STAHL, Jonathan; ZIöRJEN, Michael; STILLER, Burkhard. Landscape of iot security. *Computer Science Review*, v. 44, p. 100467, 2022. ISSN 1574-0137. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1574013722000120>. Citado na página 33.

SHEN, G.; ZHANG, J.; MARSHALL, A.; PENG, L.; WANG, X. Radio frequency fingerprint identification for lora using spectrogram and cnn. In: *IEEE INFOCOM*. [S.l.: s.n.], 2021. Citado 5 vezes nas páginas 21, 40, 42, 59 e 66.

SILVA, Jonathan de Carvalho; RODRIGUES, Joel J. P. C.; ALBERTI, Antonio M.; SOLIC, Petar; AQUINO, Andre L. L. Lorawan — a low power wan protocol for internet of things: A review and opportunities. *International Multidisciplinary Conference on Computer and Energy Science (SpliTech)*, p. 1–6, 2017. Citado na página 21.

SOLTANIEH, Naeimeh; NOROUZI, Yaser; YANG, Yang; KARMAKAR, Nemai Chandra. A review of radio frequency fingerprinting techniques. *IEEE Journal of Radio Frequency Identification*, v. 4, n. 3, p. 222–233, 2020. Citado 2 vezes nas páginas 35 e 36.

VASHI, Shivangi; RAM, Jyotsnamayee; MODI, Janit; VERMA, Saurav; PRAKASH, Chetana. Internet of things (iot): A vision, architectural elements, and security issues. In: *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*. [S.l.: s.n.], 2017. p. 492–496. Citado na página 25.

YAO, Xin. Evolving artificial neural networks. 1999. Citado na página 37.

ZANELLA, Andrea; BUI, Nicola; CASTELLANI, Angelo; VANGELISTA, Lorenzo; ZORZI, Michele. Internet of things for smart cities. *IEEE Internet of Things Journal*, Institute of Electrical and Electronics Engineers Inc., v. 1, 2014. Citado 2 vezes nas páginas 20 e 24.

ZHANG, Liwu; GONG, Liangliang; QIAN, Hankun. An effiective iot device identification using machine learning algorithm. In: *2020 IEEE 6th International Conference on Computer and Communications (ICCC)*. [S.l.: s.n.], 2020. p. 874–877. Citado 2 vezes nas páginas 35 e 63.

ZHANG, Shichao; ZHANG, Chengqi; YANG, Qiang. Data preparation for data mining. *Applied Artificial Intelligence*, v. 17, p. 375–381, 2003. Citado na página 36.

ANEXO A – Matrizes de Confusão

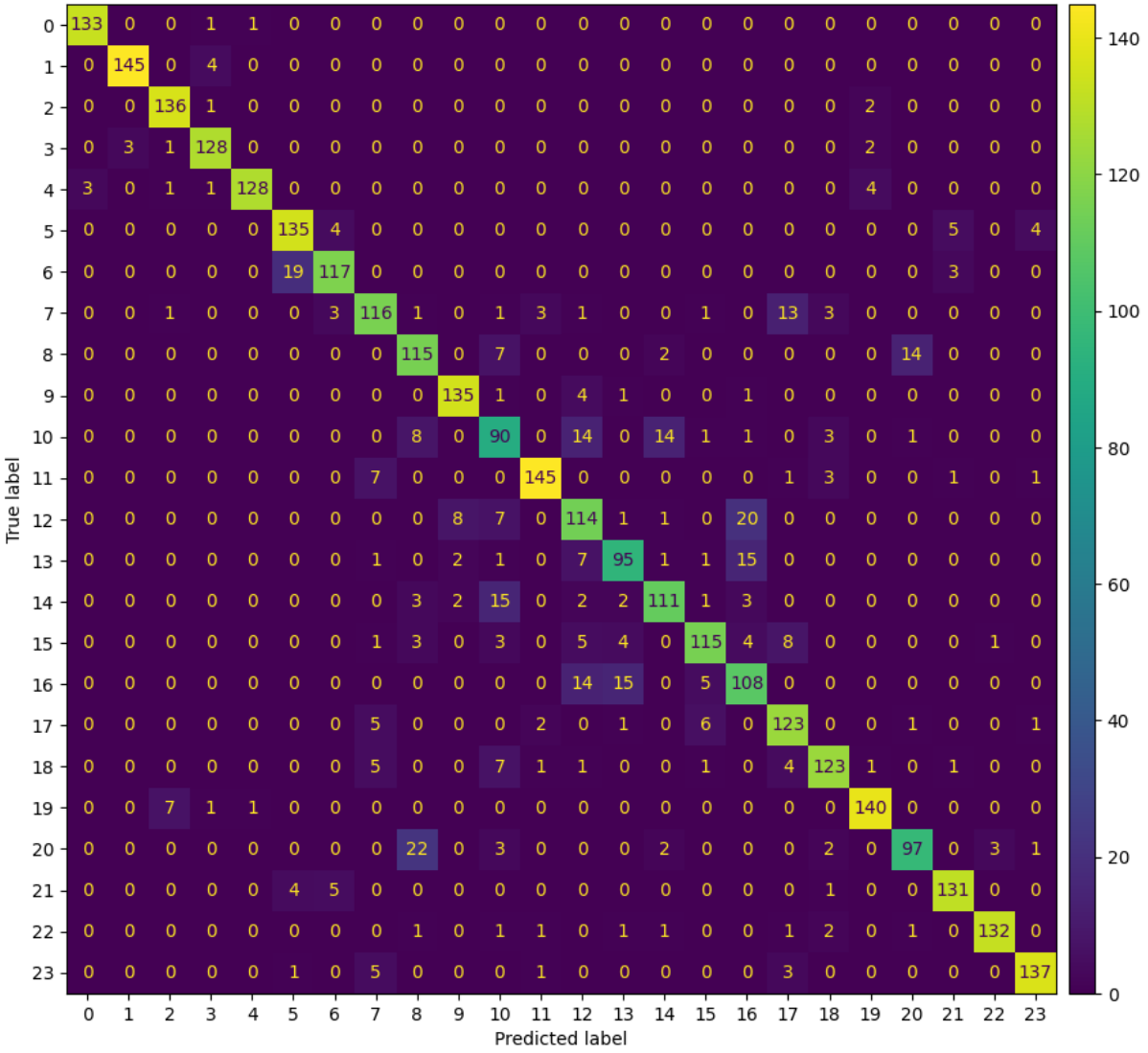


Figura 10 – Matriz de confusão do modelo LightGBM para o Grupo de dados 2.

Fonte: o autor.

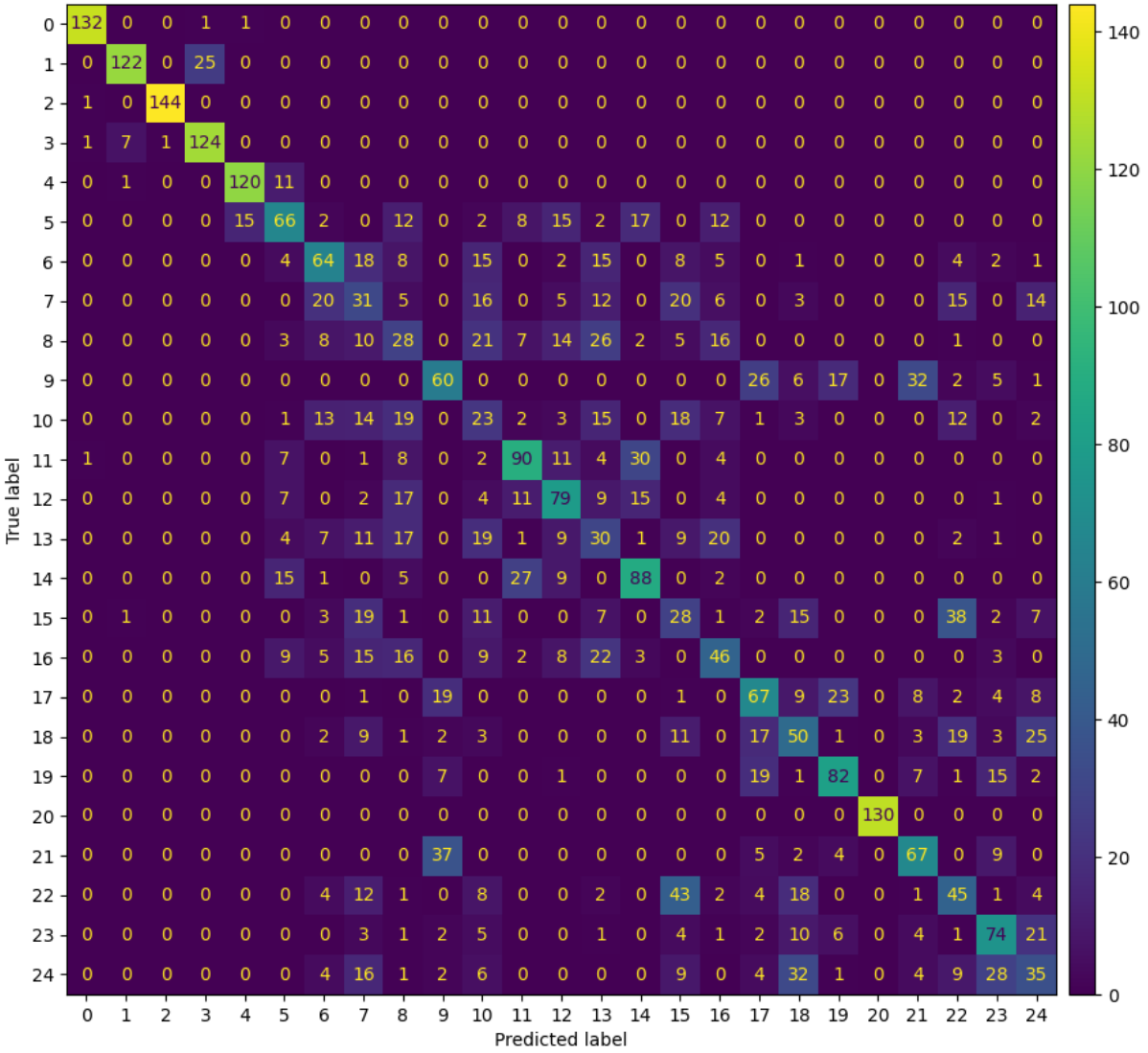


Figura 11 – Matriz de confusão do modelo *Decision Tree* para o Grupo de dados 7.

Fonte: o autor.