



UNIVERSIDADE FEDERAL DO CEARÁ
CAMPUS DE QUIXADÁ
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO
MESTRADO ACADÊMICO EM COMPUTAÇÃO

ANTONIO DAVID TAVARES AMURIM

IDENTIFICAÇÃO DE DISPOSITIVOS ELÉTRICOS EM AMBIENTES DOMÉSTICOS

QUIXADÁ

2024

ANTONIO DAVID TAVARES AMURIM

IDENTIFICAÇÃO DE DISPOSITIVOS ELÉTRICOS EM AMBIENTES DOMÉSTICOS

Dissertação apresentada ao Curso de Mestrado Acadêmico em Computação do Programa de Pós-Graduação em Computação do Campus de Quixadá da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em computação. Área de Concentração: Ciência da Computação

Orientador: Prof. Dr. Elvis Miguel Galeas Stancanelli

QUIXADÁ

2024

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

A549i Amurim, Antonio David Tavares.
Identificação de dispositivos elétricos em ambientes domésticos / Antonio David Tavares Amurim. –
2024.
96 f. : il.

Dissertação (mestrado) – Universidade Federal do Ceará, Campus de Quixadá, Programa de Pós-
Graduação em Computação, Quixadá, 2024.

Orientação: Prof. Dr. Elvis Miguel Galeas Stancanelli.

1. Aprendizado de máquina. 2. Classificação. 3. Assinatura de carga. 4. Internet das coisas. 5.
Amostragem. I. Título.

CDD 005

ANTONIO DAVID TAVARES AMURIM

IDENTIFICAÇÃO DE DISPOSITIVOS ELÉTRICOS EM AMBIENTES DOMÉSTICOS

Dissertação apresentada ao Curso de Mestrado Acadêmico em Computação do Programa de Pós-Graduação em Computação do Campus de Quixadá da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em computação. Área de Concentração: Ciência da Computação

Aprovada em: 07 de novembro de 2024

BANCA EXAMINADORA

Prof. Dr. Elvis Miguel Galeas
Stancanelli (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Regis Pires Magalhães
Universidade Federal do Ceará (UFC)

Prof. Dr. Paulo Armando Cavalcante Aguilár
Universidade Federal do Ceará (UFC)

Prof. Dr. Thiago Werlley Bandeira da Silva
Universidade Federal do Ceará (UFC)

Dedico este trabalho a Deus. Pois, até aqui o
Senhor me ajudou.

AGRADECIMENTOS

Gostaria de expressar minha gratidão a Deus por ter me sustentado ao longo desta jornada. Foi Ele quem me deu a força e as condições necessárias para alcançar esta conquista.

Agradeço profundamente ao meu orientador, Prof. Dr. Elvis Stancanelli, por todo o incentivo, ensinamento, auxílio, confiança e dedicação ao longo deste trabalho. A parceria durante esse período foi fundamental para o meu desenvolvimento profissional.

Também sou grato ao Engenheiro de laboratório Abdul-Hamid, que desempenhou um papel crucial na construção deste projeto, oferecendo orientação e apoio no desenvolvimento das ideias.

Estendo meus agradecimentos a todos os professores que contribuíram para minha formação profissional, transmitindo o conhecimento necessário para a realização deste trabalho.

Agradeço ainda aos meus pais, Irenir Amurim e Carina Amurim, pelo apoio constante, assim como a toda minha família, que esteve ao meu lado durante essa trajetória.

Sou grato, também, à minha esposa, Sara Aline, pelo incentivo, força e compreensão ao longo desta jornada.

Por fim, agradeço a todas as pessoas que, mesmo não mencionadas, contribuíram de alguma forma, direta ou indiretamente, para a conclusão deste ciclo.

“Se eu vi mais longe, foi por estar sobre ombros
de gigantes.”

(Isaac Newton)

RESUMO

A Internet das Coisas (*IoT*) tem se destacado como uma área-chave na computação, impulsionando o avanço tecnológico global. A presença de cidades e edificações inteligentes se tornou parte do nosso cotidiano, e a ideia de que tudo está conectado à internet é cada vez mais comum. A automação de tarefas, o controle de tráfego e o monitoramento do consumo de energia são apenas algumas das áreas impactadas por essa tendência. Com o aumento do número de dispositivos elétricos, o gerenciamento dos recursos energéticos em grandes construções tornou-se mais complexo. Uma má gestão pode levar a desperdícios de energia elétrica, custos financeiros desnecessários e impactos ambientais negativos. Em suma, a dificuldade em identificar quais dispositivos estão em funcionamento pode resultar em desperdício de energia e redução da vida útil dos aparelhos. Neste contexto, o trabalho propõe uma solução para identificar dispositivos elétricos domésticos com vistas à implementação em um ambiente *IoT*. A proposta envolve a análise do sinal de corrente no domínio da frequência, o que permite reduzir o volume de dados processados. Foram utilizados modelos de classificação baseados em aprendizado de máquina para identificar o dispositivo responsável pelo sinal, aproveitando padrões previamente conhecidos para essa identificação. A análise dos sinais elétricos dos dispositivos e a comparação com padrões conhecidos demonstraram que os algoritmos investigados são eficazes na identificação dos dispositivos. Entre os algoritmos testados, o classificador *kNN* se destacou, alcançando aproximadamente 99% de acurácia e F1-score. Esses resultados indicam a alta precisão da abordagem proposta e sua eficácia na identificação de dispositivos elétricos em ambientes *IoT*.

Palavras-chave: aprendizado de máquina; classificação; assinatura de carga; internet das coisas; amostragem.

ABSTRACT

The Internet of Things (IoT) has emerged as a key area in computing, driving global technological advancement. Smart cities and buildings have become part of our daily lives, and the idea that everything is connected to the Internet is becoming increasingly common. Task automation, traffic control, and energy consumption monitoring are just some of the areas impacted by this trend. With the increase in the number of electrical devices, managing energy resources in large buildings has become more complex. Poor management can lead to energy waste, unnecessary financial costs, and negative environmental impacts. In short, the difficulty in identifying which devices are operating can result in energy waste and reduced appliance lifespan. In this context, this paper proposes a solution to identify household electrical devices with a view to implementation in an IoT environment. The proposal involves analyzing the current signal in the frequency domain, which allows reducing the volume of processed data. Machine learning-based classification models were used to identify the device responsible for the signal, taking advantage of previously known patterns for this identification. The analysis of the electrical signals of the devices and the comparison with known patterns demonstrated that the investigated algorithms are effective in identifying the devices. Among the tested algorithms, the kNN classifier stood out, achieving approximately 99% accuracy and F1-score. These results indicate the high accuracy of the proposed approach and its effectiveness in identifying electrical devices in IoT environments.

Keywords: machine learning; classification; load signature; internet of things; sampling.

LISTA DE FIGURAS

Figura 1 – Espectro do sinal $X_p(j\omega)$ amostrado com $\omega_s > 2\omega_M$	23
Figura 2 – Espectro do sinal $X_p(j\omega)$ amostrado com $\omega_s < 2\omega_M$	23
Figura 3 – Sinais de corrente de diferentes classes de dispositivos elétricos em funcionamento.	24
Figura 4 – Exemplos de ferramentas e dispositivos de <i>IoT</i> que atuam nas camadas de aplicação, rede e percepção, respectivamente.	26
Figura 5 – Ilustração da técnica de classificação de algoritmo <i>kNN</i> para vetores de tamanho 2, considerando duas classes (destacadas em azul e vermelho) e <i>k</i> igual a 5 vizinhos.	29
Figura 6 – Ilustração da técnica de classificação de algoritmo <i>SVM</i> para vetores de tamanho 2, considerando duas classes (destacadas em azul e vermelho). . .	30
Figura 7 – Demonstração da estrutura de uma rede neural <i>MLP</i> com camadas ocultas. .	33
Figura 8 – Visão geral do processo de monitoramento e identificação de dispositivos elétricos domésticos.	42
Figura 9 – Sequência dos procedimentos metodológicos empregados neste trabalho. . .	43
Figura 10 – Conexões adotadas para capturar amostras de cargas elétricas de dispositivos elétricos domésticos de forma não intrusiva.	44
Figura 11 – Arquitetura detalhada dos principais componentes do sensor <i>SCT-013-000</i> .	46
Figura 12 – Estrutura do conjunto de dados contendo as 20 primeiras componentes de frequência e os rótulos dos dispositivos.	48
Figura 13 – Estrutura do conjunto de dados contendo amostras no domínio do tempo e os rótulos dos dispositivos.	49
Figura 14 – Proposta de modelo de comunicação entre as diferentes partes do projeto de <i>IoT</i>	53
Figura 15 – Modelo de comunicação com computação de borda adicionada para realizar o cálculo da <i>FFT</i>	56
Figura 16 – Imagem capturada durante a montagem do experimento de amostragem. . .	58
Figura 17 – Medição do ventilador e do micro-ondas em seu local de funcionamento. . .	58
Figura 18 – Sinal de corrente do micro-ondas no domínio do tempo construído a partir das amostras capturadas.	60

Figura 19 – Sinal de corrente da geladeira no domínio do tempo construído a partir das amostras capturadas.	60
Figura 20 – Sinal de corrente do micro-ondas no domínio da frequência construído utilizando o algoritmo <i>FFT</i>	61
Figura 21 – Sinal de corrente da geladeira no domínio da frequência construído utilizando o algoritmo <i>FFT</i>	62
Figura 22 – Amplitudes discretizadas das 20 primeiras componentes de frequência do micro-ondas obtidas por meio da <i>FFT</i> das amostras capturadas.	63
Figura 23 – Percentual de distribuição de linhas associadas a cada dispositivo no conjunto de dados de frequência.	64
Figura 24 – Algumas amostras e características do conjunto de dados no domínio do tempo.	65
Figura 25 – Avaliação do número de vizinhos do <i>kNN</i> utilizando o conjunto de dados no domínio da frequência.	66
Figura 26 – Resultados de acurácia e F1-score dos classificadores utilizando o conjunto de dados no domínio da frequência.	68
Figura 27 – Resultados de acurácia dos classificadores ao empregar os conjuntos de dados nos domínios da frequência e do tempo.	70
Figura 28 – Pontuações F1-score dos classificadores ao empregar os conjuntos de dados nos domínios da frequência e do tempo.	71
Figura 29 – Sinal de corrente distorcido de uma lâmpada incandescente operando com outros dispositivos.	72
Figura 30 – Desempenho dos modelos na classificação de dados de dispositivos operando simultaneamente com outros em ambientes ruidosos.	73
Figura 31 – Desempenho dos modelos na classificação com dados filtrados de dispositivos operando simultaneamente com outros em ambientes ruidosos.	77
Figura 32 – Desempenho do modelo <i>kNN</i> ao tentar classificar o dispositivo de ar condicionado sem possuir conhecimento sobre o aparelho durante o treinamento.	78
Figura 33 – Comparativo das amplitudes das componentes de frequência dos dispositivos.	79
Figura 34 – Resultados de acurácia dos classificadores com a inclusão de novos dispositivos no conjunto de dados.	80
Figura 35 – Resultados de F1-score dos classificadores com a inclusão de novos dispositivos no conjunto de dados.	81

Figura 36 – Percentual de entradas classificadas como operação normal, baixa potência e
alta potência. 84

LISTA DE QUADROS

Quadro 1 – Matriz de confusão com as possíveis respostas de classificação considerando as classes positiva e negativa.	34
Quadro 2 – Comparação entre os trabalhos relacionados e o proposto.	40
Quadro 3 – Classificadores selecionados e seus F1-scores de treinamento após ajuste de hiperparâmetros e validação cruzada com <i>k-fold</i> igual a 10 utilizando o conjunto de dados de frequência.	67
Quadro 4 – Classificadores selecionados e seus F1-scores de treinamento após ajuste de hiperparâmetros e validação cruzada com <i>k-fold</i> igual a 10 utilizando o conjunto de dados temporais.	69

LISTA DE ABREVIATURAS E SIGLAS

<i>IoT</i>	<i>Internet of Things</i>
<i>FT</i>	<i>Fourier Transform</i>
<i>DFT</i>	<i>Discrete Fourier Transform</i>
<i>FFT</i>	<i>Fast Fourier Transform</i>
A	Ampère
SI	Sistema Internacional
<i>kNN</i>	<i>k-Nearest Neighbor</i>
<i>SVM</i>	<i>Support Vector Machine</i>
<i>GBM</i>	<i>Gradient Boosting Machine</i>
<i>MLP</i>	<i>Multilayer Perceptron</i>
VP	Verdadeiro Positivo
VN	Verdadeiro Negativo
FN	Falso Negativo
FP	Falso Positivo
<i>NILM</i>	<i>Non-Intrusive Load Monitoring</i>
<i>REDD</i>	<i>Reference Energy Disaggregation Data Set</i>
<i>TAR</i>	<i>Triangular Area Representation</i>
<i>PCA</i>	<i>Principal Component Analysis</i>
RNAs	Redes Neurais Artificiais
<i>SCT</i>	<i>Split-core Current Transformer</i>
<i>RF</i>	<i>Random Forest</i>
<i>GB</i>	<i>Gradient Boosting</i>
pp	pontos percentuais

SUMÁRIO

1	INTRODUÇÃO	16
1.1	Justificativa	17
1.2	Motivação	18
1.3	Objetivos	18
1.4	Estrutura da dissertação	18
2	FUNDAMENTAÇÃO TEÓRICA	20
2.1	Transformada rápida de Fourier	20
2.2	Teorema da amostragem de Nyquist	21
2.3	Assinatura de carga	23
2.4	IoT	25
2.4.1	<i>Construções inteligentes</i>	25
2.5	Aprendizado de máquina	27
2.5.1	<i>Algoritmos de classificação</i>	28
2.5.1.1	<i>kNN</i>	28
2.5.1.2	<i>SVM</i>	29
2.5.1.3	<i>Árvore de Decisão</i>	30
2.5.1.4	<i>Random Forest</i>	31
2.5.1.5	<i>Gradient Boosting</i>	32
2.5.1.6	<i>Multilayer Perceptron</i>	32
2.5.2	<i>Métricas de avaliação</i>	34
3	TRABALHOS RELACIONADOS	36
3.1	Non-Intrusive Electrical Appliances Monitoring and Classification using K-Nearest Neighbors	36
3.2	A feature-based model for the identification of electrical devices in smart environments	37
3.3	Multi-Scale Electrical Appliance Load Signature for Non-Intrusive Load Monitoring Classification	37
3.4	Multi-Label Classification Based on Random Forest Algorithm for Non-Intrusive Load Monitoring System	38

3.5	Non-Intrusive Load Disaggregation by Convolutional Neural Network and Multilabel Classification	39
3.6	Comparação entre trabalhos	39
4	METODOLOGIA	42
4.1	Amostragem e análise dos dispositivos	43
4.1.1	<i>Sensor SCT-013-000</i>	45
4.2	Construção do conjunto de dados	47
4.2.1	<i>Dados no domínio do tempo</i>	48
4.3	Treinamento e avaliação dos classificadores	49
4.4	Monitoramento dos dispositivos elétricos	51
4.5	Análise da solução em um ambiente <i>IoT</i>	52
4.5.1	<i>Implementação da FFT em dispositivos de IoT</i>	54
5	RESULTADOS E DISCUSSÕES	57
5.1	Amostragem	57
5.2	Análise dos dispositivos	59
5.3	Construção do conjunto de dados	62
5.3.1	<i>Dados no domínio do tempo</i>	63
5.4	Testes preliminares com os parâmetros fundamentais dos classificadores	64
5.5	Resultados de identificação dos dispositivos	66
5.5.1	<i>Comparação com domínio do tempo</i>	69
5.5.2	<i>Impacto da presença de dispositivos simultâneos na rede: Avaliação dos classificadores em ambientes com ruído</i>	71
5.5.2.1	<i>Empregando filtros para diminuir os ruídos</i>	74
5.5.3	<i>Classificação de dispositivos desconhecidos</i>	77
5.5.4	<i>Inclusão de novos dispositivos ao conjunto de dados</i>	79
5.6	Monitoramento dos dispositivos elétricos	82
6	CONCLUSÕES	85
6.1	Considerações finais	86
6.2	Trabalhos futuros	87
	REFERÊNCIAS	89

1 INTRODUÇÃO

Atualmente, observa-se a presença da tecnologia aplicada na maioria dos ambientes de convívio (Swarnkar *et al.*, 2023; Palomsba *et al.*, 2022; Barron, 2021; Meena; Dubey, 2021). Estudos relatam que cerca de dois bilhões de pessoas utilizam a *internet* diariamente para ler *e-mail*, interagir e se comunicar nas redes sociais, realizar pesquisas na *web*, entre outras atividades (Alberti *et al.*, 2019). Nessa perspectiva, fortalece-se a ideia de diferentes objetos conectados à *internet*, visando automatizar tarefas rotineiras e impactar positivamente na qualidade de vida das pessoas.

Seguindo essa linha de pensamento, a Internet das Coisas, conhecida como *Internet of Things (IoT)* em inglês, emerge como uma das grandes inovações dos últimos anos, abrangendo diversas áreas da computação com foco em redes e sistemas embarcados (Eceiza *et al.*, 2021; Dou; Wang, 2022; Sinha; Dhanalakshmi, 2022; Boursianis *et al.*, 2022). Essa inovação proporciona uma série de benefícios em diferentes ambientes, incluindo acessibilidade, controle de consumo energético, automação e desenvolvimento sustentável (Atlam *et al.*, 2018; Panori *et al.*, 2021; Tekler *et al.*, 2022). É evidente que o uso da tecnologia em diversos contextos está crescendo de forma gradual. Diante disso, o avanço dessa modernização implica diretamente na necessidade de conectar tudo, inclusive em locais de difícil acesso, a fim de viabilizar a comunicação entre os dispositivos (Huseien; Shah, 2022).

No entanto, à medida que a tecnologia e a Internet das Coisas (*IoT*) avançam, também cresce o consumo de energia elétrica. Isso ocorre porque a modernização traz consigo novidades em termos de dispositivos e funcionalidades que demandam eletricidade. Nesse cenário, a ausência de práticas eficientes de gerenciamento dos recursos energéticos tem impactos significativos tanto na sustentabilidade quanto nos custos financeiros relacionados à energia elétrica. É importante destacar que a maioria do consumo global de eletricidade provém de edifícios comerciais e residenciais (Rocha *et al.*, 2020). Portanto, construções de maiores dimensões tendem a apresentar os maiores consumos, devido ao grande número de dispositivos elétricos conectados (Yang *et al.*, 2021; Inibhunu; McGregor, 2020), o que também pode resultar em consideráveis desperdícios, como evidenciado por um estudo de caso realizado por Ferreira *et al.* (2018).

Diante desse cenário, surgem algumas questões cruciais: como gerenciar de maneira eficiente o consumo de energia em ambientes que utilizam uma vasta gama de dispositivos conectados à rede elétrica? Quais são os principais desafios para identificar e monitorar o uso

desses dispositivos, especialmente em construções de grande porte, onde o risco de desperdício energético é mais elevado? Essas perguntas apontam para a necessidade urgente de soluções que permitam não apenas o controle, mas também a otimização do uso de energia em ambientes cada vez mais dependentes da tecnologia.

Uma solução para abordar esse problema é a identificação e monitoramento de dispositivos elétricos, que pode ser realizado através de suas assinaturas de carga, representadas por arquétipos ou padrões nos sinais elétricos (Tundis *et al.*, 2019; Khan *et al.*, 2019; Chin; Zhang, 2021). Dessa forma, é possível *identificar* conexões inadequadas ou mesmo não autorizadas de dispositivos específicos à rede elétrica. Além disso, é viável *monitorar* o estado de funcionamento dos equipamentos, identificando possíveis anomalias ao comparar as variáveis reais com os parâmetros fornecidos pelo fabricante do dispositivo, considerando também o histórico de funcionamento, conforme abordado no artigo (Rebonatto *et al.*, 2023) voltado para o ambiente hospitalar. Para alcançar esses objetivos, algoritmos de Aprendizado de Máquina podem ser empregados para classificar dispositivos com base em suas assinaturas (Thepade; Kalbhor, 2018; Barker *et al.*, 2014; Abeykoon *et al.*, 2016).

Este trabalho visa automatizar a identificação de dispositivos elétricos domésticos em um ambiente de *IoT* por meio do reconhecimento de padrões no sinal de corrente elétrica. Adotando uma abordagem comparativa, diversos algoritmos renomados foram utilizados no domínio da frequência, visando reduzir o fluxo de dados em um ambiente com recursos computacionais e energéticos limitados. Os algoritmos de reconhecimento de padrões foram aplicados e avaliados utilizando métricas como acurácia e F1-score. O propósito é contribuir para a gestão eficiente de recursos energéticos, minimizando desperdícios e identificando anomalias no consumo de energia.

1.1 Justificativa

A crescente adoção da *IoT* em ambientes domésticos e comerciais, combinada com o aumento do consumo de energia elétrica, justifica a necessidade de desenvolver soluções que possibilitem o uso mais eficiente dos recursos energéticos. Com a proliferação de dispositivos conectados, há um risco crescente de desperdício de energia, especialmente em edifícios de grande porte. Além disso, a falta de sistemas eficazes de monitoramento pode resultar em altos custos operacionais e impactos ambientais adversos. Portanto, o desenvolvimento de métodos que permitam a identificação precisa e o monitoramento contínuo de dispositivos elétricos é

essencial para mitigar esses desafios e promover a sustentabilidade energética.

1.2 Motivação

A motivação para este trabalho surge da urgência de enfrentar os desafios associados ao crescente consumo de energia em ambientes modernos, que cada vez mais dependem de dispositivos conectados. A possibilidade de aplicar técnicas de aprendizado de máquina para otimizar o uso de energia em sistemas *IoT* oferece uma oportunidade única de contribuir para a sustentabilidade. Além disso, ao desenvolver uma solução que seja eficiente e aplicável em ambientes com recursos limitados, espera-se proporcionar um impacto positivo tanto na gestão energética quanto na redução de custos operacionais, o que pode beneficiar uma ampla gama de aplicações práticas.

1.3 Objetivos

O objetivo geral deste trabalho é identificar dispositivos elétricos domésticos através da análise do sinal de corrente no domínio da frequência em ambientes de *IoT*, empregando conceitos de aprendizado de máquina. Como objetivos específicos estão:

1. Desenvolver um banco de dados que relaciona dispositivos elétricos às suas componentes de frequência presentes no sinal de corrente.
2. Treinar e avaliar classificadores baseados em aprendizado de máquina para a identificação de dispositivos elétricos domésticos por meio da análise do sinal de corrente no domínio da frequência.
3. Monitorar o funcionamento dos dispositivos elétricos com o objetivo de identificar quaisquer irregularidades ou comportamentos anômalos.
4. Desenvolver uma solução viável para um ambiente de *IoT* levando em conta as limitações de recursos de processamento e memória.

1.4 Estrutura da dissertação

O restante deste trabalho está organizado da seguinte forma: no Capítulo 2 é apresentada a fundamentação teórica em que são explicados os principais conceitos relativos a esta proposta; no Capítulo 3, são detalhados os trabalhos relacionados em que são destacadas as análises dos modelos de classificação; no Capítulo 4, são explicados os procedimentos metodoló-

gicos necessários para a realização desta proposta de trabalho; No Capítulo 5, são comparados e analisados os resultados de desempenho de cada algoritmo e, por fim, no Capítulo 6 são tratadas as considerações finais bem como detalhes acerca dos próximos passos.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, são abordados de forma sucinta os principais fundamentos relevantes para compreender este trabalho. Na Seção 2.1, é feita uma breve discussão sobre a transformada rápida de Fourier. Na Seção 2.2, é apresentado de forma concisa o teorema da amostragem de Nyquist. Na Seção 2.3, são introduzidos os conceitos de assinatura de carga, os quais são empregados na identificação de dispositivos elétricos. Na Seção 2.4, são apresentados os paradigmas da *IoT*. Por fim, na Seção 2.5, são detalhados os conceitos de aprendizado de máquina, particularmente os algoritmos e métricas direcionados à classificação.

2.1 Transformada rápida de Fourier

A Transformada de Fourier, ou *Fourier Transform (FT)*, é um sistema linear invariante no tempo, definido como uma representação no domínio da frequência de um sinal original (Tripathy *et al.*, 2023). Nesse contexto, a *FT* é amplamente empregada para converter sinais para o domínio da frequência, possibilitando a identificação das frequências que compõem o sinal. Embora a transformada de Fourier não esteja estritamente limitada a funções temporais, por convenção, o domínio original costuma ser o tempo.

Considerando um sinal periódico contínuo, a transformada resultante consiste em um trem de impulsos no domínio da frequência, onde as áreas dos impulsos indicam quais frequências estão presentes no sinal e qual amplitude cada uma exerce no sinal (Oppenheim; Willsky, 2010). Assim, é possível constatar que um sinal no domínio do tempo pode ser representado como uma soma de exponenciais complexas, com suas frequências e amplitudes indicadas pela *FT* resultante. A Equação 2.1 apresenta a Transformada de Fourier como $X(j\omega)$, onde $x(t)$ representa o sinal no domínio do tempo a ser transformado, t corresponde à variável temporal, j é a unidade imaginária e ω indica a frequência angular, expressa em radianos por segundo.

$$X(j\omega) = \int_{-\infty}^{+\infty} x(t)e^{-j\omega t} dt \quad (2.1)$$

Da mesma forma, a *FT* pode ser aplicada a um sinal de tempo discreto, também conhecida como Transformada Discreta de Fourier, do inglês *Discrete Fourier Transform (DFT)*, conforme apresentado pelos autores Oppenheim e Willsky (2010). No entanto, o custo computacional para calcular a *DFT* é da ordem de n^2 , onde n representa o número de dados. Isso

significa que o tempo de processamento aumenta proporcionalmente ao quadrado da quantidade de dados, tornando impraticável a aplicação da transformada em conjuntos extensos de dados.

Diante dessa limitação, surgiu a Transformada Rápida de Fourier, conhecida como *Fast Fourier Transform (FFT)*, proposta por John Tukey em 1965 (Feng *et al.*, 2024). Em resumo, a *FFT* é um algoritmo eficiente que produz os mesmos resultados da *DFT* com uma complexidade de $n \log_2 n$, ou seja, em um tempo significativamente menor proporcional ao número de dados (Liu *et al.*, 2019). Assim, o algoritmo *FFT* é empregado nesta proposta no processo de conversão de sinais para o domínio da frequência.

A implementação do algoritmo *FFT* é amplamente disponível em diversas linguagens de programação, facilitando sua utilização em aplicações práticas. No entanto, é importante notar que o algoritmo *FFT* também possui implementações dedicadas em *hardware*, o que acelera ainda mais sua execução (Arivamudhan; Santhi, 2018). Circuitos especializados para a *FFT*, muitas vezes integrados em chips específicos, oferecem vantagens em termos de desempenho e eficiência energética (Liu *et al.*, 2022).

Esses circuitos são projetados para realizar as operações fundamentais da *FFT* de maneira otimizada, aproveitando características específicas da transformada. Essas implementações em *hardware* são comumente encontradas em aplicações que requerem processamento de sinais em tempo real, como em comunicações sem fio, processamento de áudio e vídeo, entre outras (Li *et al.*, 2000). Ao utilizar circuitos dedicados, é possível obter benefícios significativos em termos de velocidade de processamento e redução do consumo de energia em comparação com implementações puramente baseadas em *software* (Sakashita *et al.*, 2015).

2.2 Teorema da amostragem de Nyquist

Normalmente, sem considerar informações adicionais, não seria possível representar exclusivamente um sinal por meio de uma sequência de amostras uniformemente espaçadas. Isso ocorre porque existe um número incontável de sinais que podem passar pelos pontos definidos na sequência de amostras. No entanto, se um sinal for limitado em banda, ou seja, se sua *FT* for nula fora de um intervalo finito de frequências, e se as amostras forem escolhidas suficientemente próximas em relação à frequência fundamental do sinal, então essa sequência de amostras pode representar de forma única esse sinal. Isso possibilita a reconstrução do sinal em tempo contínuo (Oppenheim; Willsky, 2010). Essa conclusão é fundamental para o teorema da amostragem de Nyquist, que é aplicado no processo de amostragem realizado nesta dissertação.

Visando o teorema de Nyquist, a amostragem de um sinal será representada por um trem de impulsos periódico multiplicado pelo sinal de tempo contínuo $x(t)$ que será demonstrado. Em resumo, o trem de impulsos periódico $p(t)$ é conhecido como função de amostragem, o período T , como o período de amostragem, e a frequência fundamental de $p(t)$, $\omega_s = 2\pi/T$, como a frequência de amostragem (Oppenheim; Willsky, 2010). Dessa forma, a função de amostragem $p(t)$ é apresentada na Equação 2.2, em que $\delta(t - nT)$ representa um impulso localizado no instante t , com nT determinando o deslocamento temporal de cada impulso.

$$p(t) = \sum_{n=-\infty}^{+\infty} \delta(t - nT) \quad (2.2)$$

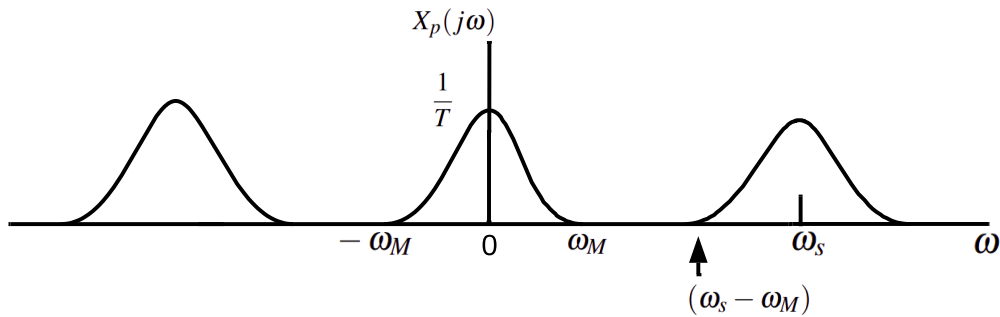
Segundo Oppenheim e Willsky (2010), é possível afirmar que ao multiplicar $x(t)$ por um impulso unitário é feita a seleção da amostra do sinal no ponto em que o impulso está localizado, ou seja, $x(t)\delta(t - t_0) = x(t_0)\delta(t - t_0)$. Logo, conclui-se a Equação 2.3, em que $x_p(t)$ é equivalente a um trem de impulsos com as amplitudes dos impulsos iguais às amostras de $x(t)$ em intervalos espaçados de T . Ainda, aplicando algumas propriedades matemáticas demonstradas em Oppenheim e Willsky (2010) e considerando que a convolução com um impulso simplesmente desloca um sinal, ou seja, $X(j\omega)\delta(\omega - \omega_0) = X(j(\omega - \omega_0))$, conclui-se a Equação 2.4. Nesse sentido, $X_p(j\omega)$ é uma função de ω periódica definida por uma sobreposição de réplicas deslocadas de $X(j\omega)$ multiplicadas por $1/T$.

$$x_p(t) = \sum_{n=-\infty}^{+\infty} x(nT)\delta(t - nT) \quad (2.3)$$

$$X_p(j\omega) = \frac{1}{T} \sum_{k=-\infty}^{+\infty} X(j(\omega - k\omega_s)) \quad (2.4)$$

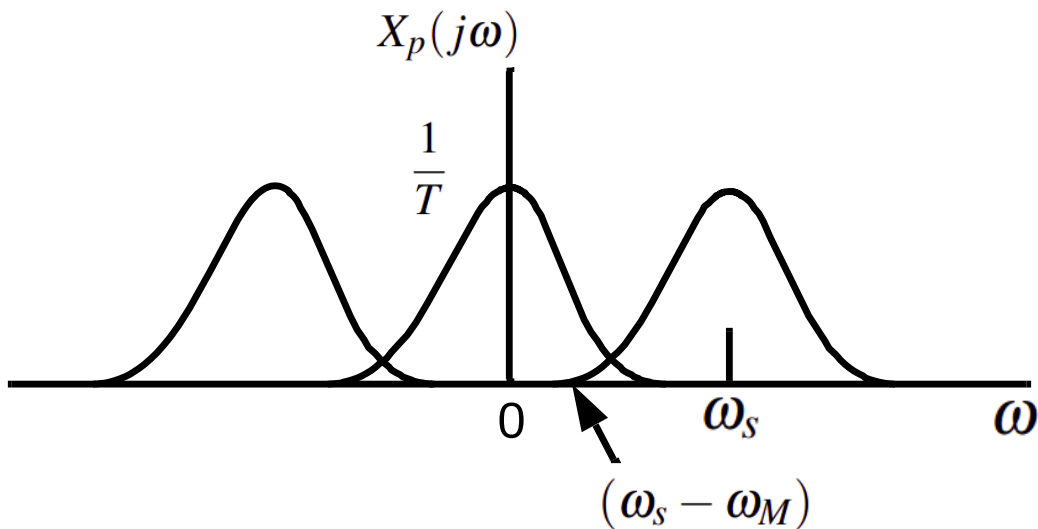
Como mostrado na Figura 1, não existe sobreposição entre as réplicas deslocadas de $X(j\omega)$, pois, $\omega_s > 2\omega_M$, onde ω_M é a maior frequência presente no sinal contínuo $x(t)$. Logo, o sinal pode ser precisamente retratado por múltiplos inteiros da frequência de amostragem (Oppenheim; Willsky, 2010). Seguindo essa lógica, se $\omega_s > 2\omega_M$, $x(t)$ pode ser reconstruído de forma contínua a partir de $x_p(t)$, utilizando um filtro passa-baixas com ganho T e uma frequência de corte maior que ω_M e menor que $\omega_s - \omega_M$, podendo então aplicar o teorema da amostragem. Em contrapartida, na Figura 2 há sobreposição, pois, $\omega_s < 2\omega_M$. Neste caso, ocorre o fenômeno chamado de *aliasing*, onde a amostragem não satisfaz as condições do teorema de Nyquist, não possibilitando a recuperação do sinal contínuo.

Figura 1 – Espectro do sinal $X_p(j\omega)$ amostrado com $\omega_s > 2\omega_M$



Fonte: Adaptada de Oppenheim e Willsky (2010)

Figura 2 – Espectro do sinal $X_p(j\omega)$ amostrado com $\omega_s < 2\omega_M$



Fonte: Adaptada de Oppenheim e Willsky (2010)

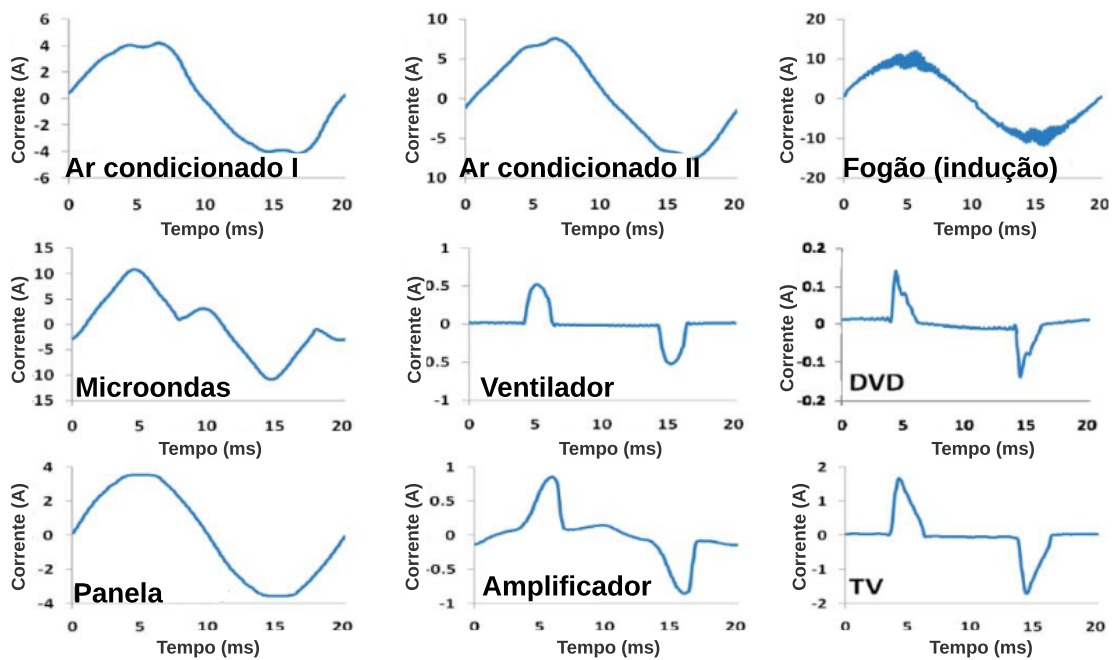
2.3 Assinatura de carga

A assinatura de carga pode ser definida como uma característica específica de uma carga, que permite observar um padrão distinto no consumo de um dispositivo elétrico (Liang *et al.*, 2009). Nesse contexto, as características observáveis geralmente englobam variáveis como tensão, potência e corrente, incluindo aspectos como formato de onda, média, variância, valores de pico e vale, representações gráficas do sinal, frequência, entre outras.

Nesta proposta, as assinaturas utilizadas serão características do sinal de corrente no domínio da frequência. A corrente elétrica pode ser definida como o movimento ordenado de

portadores de cargas elétricas (Feitosa *et al.*, 2021). Em resumo, a corrente é expressa como o número de elétrons passando através de uma secção condutora de eletricidade por unidade de tempo. Além disso, sua unidade de medida é calculada em Ampère (A), conforme o Sistema Internacional (SI). Deste modo, os dispositivos serão identificados por meio do seu sinal de corrente elétrica. Para exemplificar, a Figura 3 mostra o sinal de corrente de alguns dispositivos elétricos em funcionamento, onde é possível perceber os diferentes valores e formatos de onda próprios de cada dispositivo.

Figura 3 – Sinais de corrente de diferentes classes de dispositivos elétricos em funcionamento.



Fonte: Adaptada de Liang *et al.* (2009)

Como mostrado na Figura 3, os gráficos de corrente dos dispositivos de ar condicionado I e II se diferenciam um pouco em termos de amplitude, porém, demonstram formatos de onda semelhantes, indicando uma relação por serem dispositivos da mesma classe. Em contrapartida, eles se distinguem claramente de outros dispositivos com propriedades distintas. Em uma análise adicional, é possível perceber que dispositivos como a *TV* e o *DVD* possuem algumas características semelhantes em seus sinais, devido à presença de circuitos eletrônicos em grande parte de sua composição. No entanto, os valores de pico apontam diferenças que possibilitam suas identificações distintas.

2.4 IoT

O desenvolvimento de dispositivos tecnológicos conectados à *internet* e a intercomunicação entre eles está experimentando um notável crescimento. Nesse contexto, a demanda por uma conectividade contínua, mesmo em locais remotos, está aumentando. Sob essa perspectiva, a *IoT* emerge como uma significativa inovação no campo da tecnologia da informação, trazendo diversas vantagens, tais como automação, facilidade de acesso e controle do consumo energético, entre outros (Atlam *et al.*, 2018).

A arquitetura de comunicação *IoT* é dividida em camadas com funcionalidades específicas. Além disso, cada camada se caracteriza pelos dispositivos e tecnologias utilizadas. Sobretudo, a estrutura da *IoT* implementa as camadas de Percepção, Rede e Aplicação (Mahmoud *et al.*, 2015; Lin *et al.*, 2017), como mostra a Figura 4. Logo, essas camadas da *IoT* possuem seus próprios obstáculos e ferramentas que contribuem na resolução das dificuldades encontradas. Vale salientar que, a arquitetura de camadas *IoT* conta com derivações, possuindo mais ou menos camadas, Wang *et al.* (2016) mostram um exemplo de arquitetura contendo um número maior de subdivisões.

2.4.1 Construções inteligentes

Construções inteligentes são estruturas que utilizam os conceitos de *IoT* e implementam projetos a fim de conectar diferentes coisas ao mundo virtual. Deste modo, esses ambientes inteligentes agregam valores às pessoas e aos negócios, oferecendo serviços tecnológicos, melhorando a sustentabilidade e utilizando os recursos de modo eficiente e eficaz (Alberti *et al.*, 2019). De fato, qualquer ambiente físico pode se tornar inteligente, compreendendo de forma precisa o que acontece em determinados espaços físicos ou seres vivos.

Nesse raciocínio, o conceito de prédios inteligentes pode se expandir para a ideia de cidades inteligentes, ou seja, um aglomerado de diferentes espaços que empregam tecnologia e conceitos de *IoT*. Em resumo, cidades inteligentes se caracterizam por dispor de infraestruturas tecnológicas na maioria de seus segmentos. Além disso, oferecem serviços e automatização de tarefas para os cidadãos em inúmeras áreas como pontos de *Wi-Fi* (*IEEE* 802.11), energia renovável, defesa cibernética, aplicativos governamentais, áreas climatizadas, transportes com funcionalidades inteligentes, entre outros (Özcan *et al.*, 2017).

Atualmente, as soluções e serviços *IoT* são fundamentais para viabilizar ambientes

Figura 4 – Exemplos de ferramentas e dispositivos de *IoT* que atuam nas camada de aplicação, rede e percepção, respectivamente.



Fonte: Elaborada pelo próprio autor.

inteligentes em residências, edifícios e cidades (Farahzadi *et al.*, 2018). Todavia, são constantemente encontrados problemas de pouca memória, latência na comunicação e processamento lento, devido à disposição de poucos recursos computacionais encontrados nos dispositivos *IoT* que precisam ser baratos e portáteis, às vezes adaptados a ambientes remotos.

Contudo, os avanços tecnológicos trazem consigo soluções visando melhorar a comunicação e o fluxo de dados entre camadas *IoT*. Para exemplificar, pode ser mencionado o conceito de Nuvem das Coisas que proporciona recursos de comunicação e armazenamento em Nuvem voltados a ambientes *IoT*. Outro exemplo é a ideia de computação em nevoeiro, que aproxima o processamento em Nuvem dos dispositivos *IoT*, oferecendo um maior poder de computação em uma distância fisicamente próxima.

Além disso, a implementação de uma comunicação simples, com poucos dados e protocolos leves, é essencial para o bom desempenho de um dispositivo *IoT*. Nesta proposta, os conceitos de *IoT* são utilizados para fundamentar a lógica empregada na aquisição e no processamento dos dados. Desta forma, busca-se uma comunicação simples e com baixo fluxo

de dados, visando implantar a ideia em uma construção inteligente.

2.5 Aprendizado de máquina

Com o avanço das tecnologias de comunicação e, conseqüentemente, a crescente demanda por dados, a análise inteligente das informações ganha cada vez mais importância, tornando-se essencial para o progresso da computação (Osisanwo *et al.*, 2017). Nesse contexto, a combinação da inteligência artificial com a capacidade das máquinas de aprender e se adaptar a diferentes contextos por meio de dados tornou-se uma realidade, utilizada para automatizar tarefas, tomar decisões com base na análise dos dados, classificar e identificar padrões (Miah *et al.*, 2019). Esta abordagem explora os conceitos de aprendizado de máquina, em particular o reconhecimento de padrões. Dessa forma, modelos de classificação são aplicados para identificar características específicas nos sinais de corrente dos dispositivos.

O aprendizado de máquina diz respeito ao desenvolvimento de algoritmos e técnicas que permitem aos computadores aprenderem a partir de dados (Jadhav *et al.*, 2023). Em vez de programar explicitamente regras específicas para realizar uma tarefa, os sistemas de aprendizado de máquina são projetados para aprender com exemplos e experiências passadas, permitindo que façam previsões ou tomem decisões com base nesses dados. Existem vários tipos de aprendizado de máquina, sendo os principais:

- **Aprendizado supervisionado:** Os algoritmos são treinados em um conjunto de dados rotulados, ou seja, onde cada exemplo do conjunto de dados é acompanhado por um rótulo que indica a resposta correta. O objetivo é aprender uma função que mapeia as entradas para as saídas corretas. Alguns exemplos de algoritmos de aprendizado supervisionado incluem regressão linear, árvores de decisão e redes neurais (Baharuddin *et al.*, 2022).
- **Aprendizado não supervisionado:** Os algoritmos são treinados em um conjunto de dados não rotulado, e o sistema é deixado para encontrar padrões ou estruturas nos dados por conta própria. Isso é útil para descobrir percepções ou padrões em grandes conjuntos de dados e para tarefas como segmentação de clientes, agrupamento e redução de dimensionalidade (Subramanian *et al.*, 2021).
- **Aprendizado por reforço:** Neste paradigma, o sistema aprende a tomar decisões sequenciais através da interação com um ambiente. O agente recebe retornos positivos ou negativos com base nas ações que executa e usa esse retorno para aprender a política ótima para alcançar seus objetivos. Isso é comumente usado em jogos e robótica (Kaloiev; Krastev,

2023).

Além desses, existem outras abordagens e técnicas, como aprendizado parcialmente supervisionado, aprendizado por transferência, redes neurais profundas, etc. Nesta proposta, são utilizados apenas algoritmos de aprendizado supervisionado (Richardson *et al.*, 2022). Esses algoritmos são detalhados a seguir.

2.5.1 Algoritmos de classificação

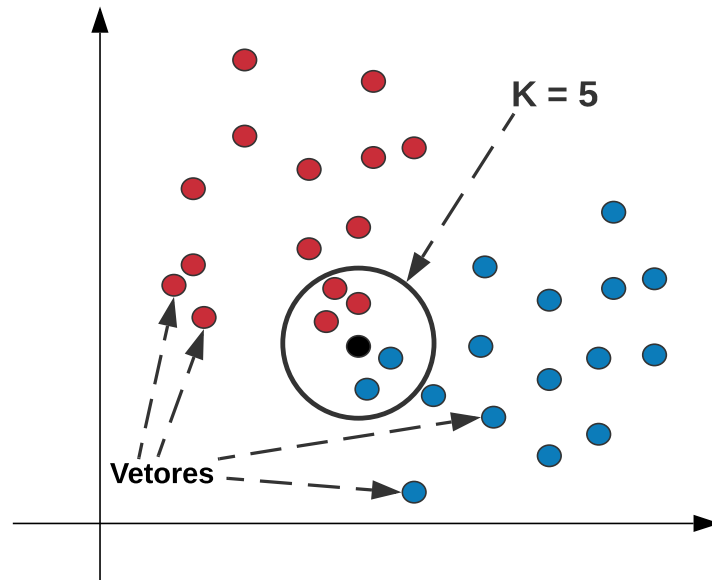
Conforme mencionado por Zhang *et al.* (2017), a classificação ou identificação de padrões é uma área de pesquisa crucial na mineração de dados, com constantes evoluções e novas abordagens sendo propostas. Nesta seção, a atenção será concentrada nas técnicas de aprendizado supervisionado, nas quais o modelo é treinado com base em resultados previamente definidos (Singh *et al.*, 2016). Portanto, será detalhado o funcionamento de alguns classificadores a seguir.

2.5.1.1 *kNN*

O *k-Nearest Neighbor (kNN)* é um algoritmo de aprendizado supervisionado no qual o parâmetro k representa um número específico. A abordagem do *kNN* é simples: para classificar um vetor, examinam-se as classes dos k vetores mais próximos a ele; a classe de um vetor é determinada pela moda das classes dos k vetores mais próximos (Dey *et al.*, 2018). A Figura 5 ilustra o conceito do algoritmo *kNN*.

Na Figura 5, é visível uma disposição de vetores de amostra em um plano, com alguns marcados em vermelho e outros em azul para representar diferentes classes. O objetivo do algoritmo é categorizar o vetor preto localizado no centro do círculo. Esse círculo é definido para englobar os k vizinhos mais próximos do vetor em questão, sendo k igual a 5 neste caso. Durante a execução do algoritmo, os vizinhos são selecionados através do cálculo da distância entre o vetor a ser classificado e os demais vetores. O método de cálculo de distância é um parâmetro que pode ser ajustado no algoritmo. Portanto, ao observar que a maioria dos k vizinhos selecionados pertence à classe vermelha, conclui-se que o vetor preto é categorizado como pertencente à classe vermelha pelo algoritmo.

Figura 5 – Ilustração da técnica de classificação de algoritmo *kNN* para vetores de tamanho 2, considerando duas classes (destacadas em azul e vermelho) e k igual a 5 vizinhos.



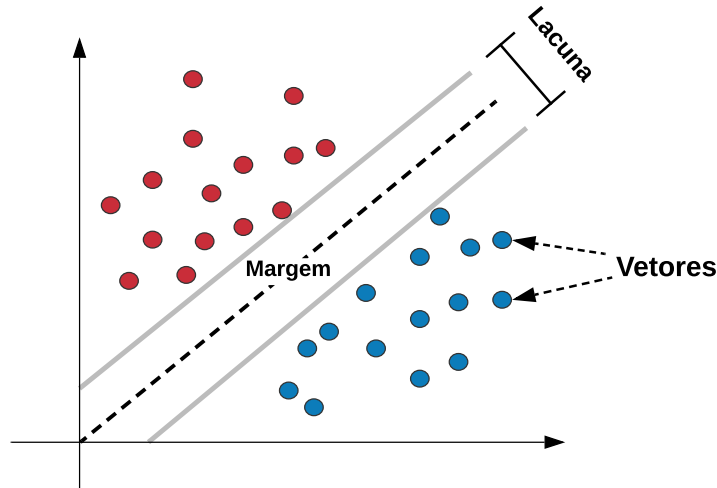
Fonte: Elaborada pelo próprio autor.

2.5.1.2 SVM

A *Support Vector Machine (SVM)* é uma técnica de aprendizado supervisionado frequentemente empregada para classificação, embora também seja utilizada em regressão em menor escala. Seu principal foco reside na classificação linear, embora seja capaz de realizar classificações não lineares de forma eficiente, por meio de um mapeamento implícito de suas entradas em espaços de características de alta dimensão, conhecido como o truque do *kernel* (Mahesh, 2020). A ideia central do *SVM* é estabelecer margens entre as classes, de modo que essas margens sejam desenhadas de forma a maximizar a distância entre elas e as classes, garantindo que a margem entre os rótulos seja traçada de modo centralizado. Isso assegura que a classificação não seja tendenciosa para nenhum dos rótulos e minimiza possíveis erros. A ideia do *SVM* é exemplificada na Figura 6.

Como mostrado na Figura 6, vetores de duas classes distintas estão posicionados em um plano, destacados em vermelho e azul. Em seguida, é estabelecida uma margem que separa as duas classes de forma que a lacuna entre a linha que define a margem e os vetores seja maximizada. No entanto, em cenários reais, pode haver várias classes, o que implica em mais do que apenas duas dimensões. Nesse caso, a mesma abordagem é aplicada selecionando

Figura 6 – Ilustração da técnica de classificação de algoritmo SVM para vetores de tamanho 2, considerando duas classes (destacadas em azul e vermelho).



Fonte: Elaborada pelo próprio autor.

hiperplanos com a máxima distância possível entre as classes (Dey *et al.*, 2018).

2.5.1.3 Árvore de Decisão

O classificador Árvore de Decisão é fundamentado em uma estrutura em forma de árvore composta por vários pontos de decisão. Em cada ponto de decisão, é feita uma escolha sobre qual ramo seguir até alcançar um resultado. Essa árvore é construída utilizando os dados de treinamento, o que permite que o algoritmo classifique novos dados com base nas decisões tomadas durante o treinamento (Slijepcevic *et al.*, 2023). Além disso, é importante destacar que esse método não se restringe apenas à classificação, sendo também aplicável a problemas de regressão. Ele serve como base para outros algoritmos, como o *Random Forest*.

Além de sua aplicação em classificação e regressão, o classificador Árvore de Decisão é apreciado pela sua capacidade de fornecer interpretabilidade aos resultados. Isso ocorre porque a estrutura de árvore é facilmente compreensível, permitindo uma análise clara das decisões tomadas em cada ponto (Lu *et al.*, 2022). Cada nó da árvore representa uma característica e cada ramo representa uma decisão com base nessa característica. Portanto, ao seguir os ramos da árvore, é possível entender intuitivamente o raciocínio por trás de cada classificação ou predição. Essa característica torna o classificador Árvore de Decisão valioso em cenários onde a explicação do modelo é tão importante quanto sua precisão.

Outro aspecto importante do classificador Árvore de Decisão é sua capacidade de lidar com dados tanto numéricos quanto categóricos (Zhang, 2021). Isso significa que ele não requer pré-processamento extensivo dos dados, como normalização ou codificação de variáveis categóricas. Além disso, o algoritmo é robusto em relação a valores ausentes nos dados, podendo lidar com eles de maneira eficaz durante a construção da árvore de decisão. Isso simplifica o processo de modelagem e torna o classificador Árvore de Decisão uma escolha atraente em muitas situações práticas.

2.5.1.4 *Random Forest*

Os métodos de *ensemble* são técnicas de aprendizado que combinam múltiplos modelos (Ramteke; Maidamwar, 2023). O *Random Forest*, um dos métodos mais populares de *ensemble* em aprendizado de máquina, é construído sobre o princípio de *Bagging*. Esse princípio envolve a criação de múltiplos subconjuntos de dados de treinamento, gerados por amostragem com reposição, e o treinamento de um modelo em cada subconjunto. As previsões finais são obtidas por meio da média ou votação dos modelos individuais, o que reduz a variância e melhora a robustez do modelo. O classificador *Random Forest* utiliza múltiplas árvores de decisão, cada uma treinada em uma amostra aleatória dos dados de treinamento, com reposição. Em outras palavras, várias árvores de decisão são construídas independentemente, e a previsão final é obtida por meio da agregação das previsões individuais das árvores (Prihatno *et al.*, 2021).

Uma das características do *Random Forest* é sua capacidade de lidar com grandes conjuntos de dados com muitas variáveis. Ele é menos propenso a sobreajuste em comparação com uma única árvore de decisão, tornando-o uma escolha robusta em muitas situações práticas (Guo *et al.*, 2019). Além disso, o método pode ser facilmente paralelizado, o que o torna eficiente para grandes conjuntos de dados e ambientes distribuídos.

Outro ponto relevante é a sua flexibilidade em lidar com diferentes tipos de dados, como variáveis categóricas e numéricas, sem a necessidade de pré-processamento extensivo. Isso simplifica o fluxo de trabalho de modelagem e permite que o *Random Forest* seja aplicado em uma variedade de problemas de aprendizado de máquina, tanto para classificação quanto para regressão (Dhanka; Maini, 2021). No entanto, é importante ajustar corretamente os parâmetros do algoritmo, como o número de árvores e a profundidade máxima de cada árvore, para alcançar o melhor desempenho possível.

2.5.1.5 Gradient Boosting

As técnicas de *boosting* surgiram com a ideia de combinar de forma iterativa diversos modelos simples, chamados de aprendizes fracos, para formar um aprendiz forte com maior capacidade de previsão (Touzani *et al.*, 2018). Inicialmente aplicado principalmente a problemas de classificação, o conceito de *boosting* foi expandido para a regressão com o desenvolvimento do método *Gradient Boosting Machine (GBM)*.

O método *GBM* inicia o processo com uma estimativa inicial, geralmente uma árvore de decisão, que busca minimizar a função de erro. Em cada iteração, uma nova árvore de decisão é ajustada e adicionada ao modelo anterior, visando reduzir ainda mais o erro global (Yan; Wen, 2021). O número de iterações é um parâmetro ajustável, definido pelo usuário.

Uma das principais vantagens do *GBM* é sua capacidade de capturar relações complexas entre as variáveis de entrada e a variável de saída. Ele é robusto em relação a pontos discrepantes e lida bem com diferentes tipos de dados, como variáveis categóricas e numéricas (Su, 2020). Além disso, o *GBM* é menos suscetível a sobreajuste em comparação com outros métodos de *ensemble* devido à abordagem de ajuste sequencial das árvores de decisão.

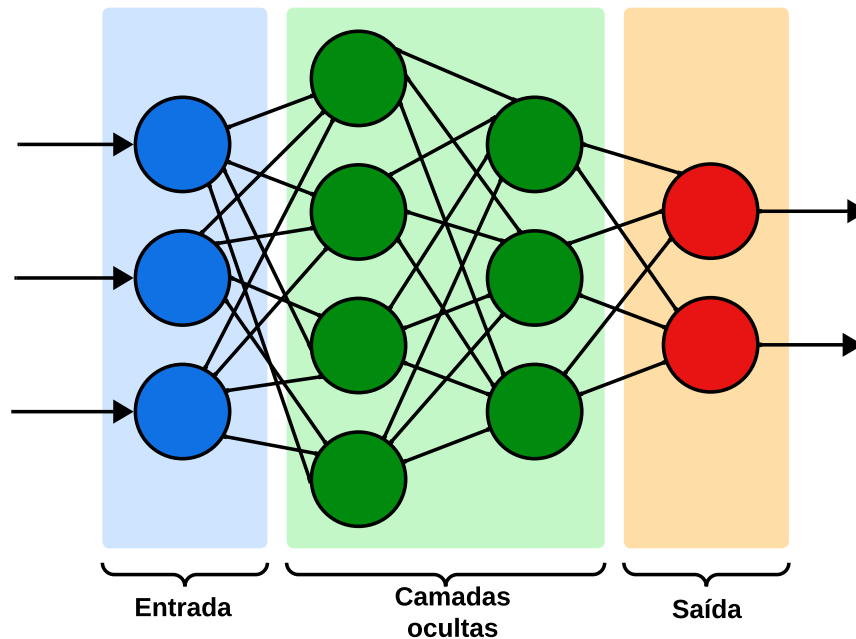
No entanto, é importante observar que o *GBM* pode ser sensível aos parâmetros de ajuste e ao número de iterações. Um ajuste inadequado desses parâmetros pode levar a um tempo de treinamento prolongado ou a um modelo com desempenho subótimo (Niu *et al.*, 2021). Portanto, é essencial realizar uma seleção cuidadosa dos parâmetros e, se necessário, aplicar técnicas de otimização para encontrar a configuração ideal do modelo.

2.5.1.6 Multilayer Perceptron

O *Multilayer Perceptron (MLP)* é um tipo de rede neural artificial composta por várias camadas de neurônios, incluindo uma ou mais camadas ocultas entre a camada de entrada e a camada de saída (Tang *et al.*, 2016). Cada neurônio em uma camada está conectado a todos os neurônios da camada anterior e da camada seguinte, formando uma rede densamente conectada, como mostra a Figura 7. Durante o treinamento do *MLP*, os pesos das conexões entre os neurônios são ajustados usando um algoritmo de retropropagação, que calcula gradientes para atualizar os pesos de modo a minimizar uma função de custo. A função de custo geralmente mede a diferença entre as saídas previstas da rede e as saídas reais dos dados de treinamento.

Uma característica importante do *MLP* é sua capacidade de aprender representações

Figura 7 – Demonstração da estrutura de uma rede neural *MLP* com camadas ocultas.



Fonte: Elaborada pelo próprio autor.

complexas dos dados de entrada, permitindo modelar relações não lineares entre as características. Isso o torna adequado para uma variedade de tarefas de aprendizado de máquina, como classificação e regressão (Jain *et al.*, 2018). Além disso, o *MLP* oferece flexibilidade em termos de arquitetura, permitindo que os usuários ajustem o número de camadas ocultas, o número de neurônios em cada camada e as funções de ativação dos neurônios. Esses hiperparâmetros podem influenciar significativamente o desempenho e a capacidade de generalização do modelo, exigindo ajustes cuidadosos durante o processo de treinamento e validação (AlBasiouny *et al.*, 2022).

No entanto, é importante notar que os *MLPs* podem ser sensíveis ao problema de sobreajuste, especialmente em conjuntos de dados pequenos ou mal balanceados. Estratégias como regularização, validação cruzada, seleção de modelos e a técnica de *dropout* podem ser empregadas para mitigar esse problema e melhorar o desempenho do *MLP* em uma variedade de cenários de aplicação (Mishra; Huhtala, 2019). O *dropout* funciona desativando aleatoriamente uma fração dos neurônios durante cada iteração do treinamento, o que força a rede a não depender excessivamente de neurônios específicos e, assim, ajuda a reduzir o sobreajuste.

2.5.2 Métricas de avaliação

Na literatura, foram estabelecidas métricas de avaliação para mensurar o desempenho dos modelos de classificação. Essas métricas são derivadas da matriz de confusão, que resume os resultados das classificações reais e as previsões feitas pelo classificador, conforme mostrado no Quadro 1 (Fayzrakhmanov *et al.*, 2018).

Quadro 1 – Matriz de confusão com as possíveis respostas de classificação considerando as classes positiva e negativa.

Real	Predito	
	Classe Positiva	Classe Negativa
Classe Positiva	Verdadeiro Positivo (VP)	Falso Negativo (FN)
Classe Negativa	Falso Positivo (FP)	Verdadeiro Negativo (VN)

Fonte: Adaptado de Fayzrakhmanov *et al.* (2018)

Conforme observado, os resultados corretos incluem tanto os Verdadeiro Positivo (VP) quanto os Verdadeiro Negativo (VN), indicando que o método previu corretamente se um elemento pertence à classe positiva ou negativa, respectivamente. Por outro lado, os Falso Negativo (FN) representam informações que são verdadeiramente positivas, mas foram erroneamente classificadas como negativas. Da mesma forma, os Falso Positivo (FP) correspondem a dados que são verdadeiramente negativos, mas foram incorretamente classificados como positivos (Chicco; Jurman, 2020).

No contexto da avaliação de classificadores, a acurácia, representada pela Equação 2.5, é uma métrica comumente utilizada que indica a proporção de predições corretas feitas pelo modelo em relação ao total de predições (Almeida *et al.*, 2018). Apesar de ser amplamente adotada, a acurácia pode não ser tão confiável em conjuntos de dados desbalanceados, onde uma classe é dominante em relação à outra.

$$\text{Acurácia} = \frac{VP + VN}{VP + VN + FP + FN} \quad (2.5)$$

A métrica de precisão, demonstrada pela Equação 2.6, foca em detectar erros associados aos falsos positivos. Essa métrica é particularmente útil em cenários onde é crucial minimizar os casos em que o modelo prevê incorretamente uma instância como pertencente a uma classe específica, quando na verdade não é (Philippou *et al.*, 2020).

$$\text{Precisão} = \frac{VP}{VP + FP} \quad (2.6)$$

A revocação, definida pela Equação 2.7, tem como objetivo detectar erros relacionados aos falsos negativos. Ela indica a proporção de instâncias de uma classe que foram corretamente identificadas pelo modelo em relação ao total de instâncias que realmente pertencem a essa classe (Philippou *et al.*, 2020). A revocação é especialmente relevante em situações em que é crucial identificar corretamente todas as instâncias de uma classe, minimizando assim os casos em que o modelo deixa de reconhecer instâncias verdadeiramente positivas.

$$\text{Revocação} = \frac{VP}{VP + FN} \quad (2.7)$$

O F1-score, apresentado na Equação 2.8, é uma média harmônica entre a precisão e a revocação, o que significa que ele penaliza valores extremos em qualquer uma das métricas. Essa abordagem oferece uma medida balanceada, levando em conta tanto os erros de falsos positivos quanto os de falsos negativos. Isso evita que um modelo tenha um bom desempenho em uma das métricas enquanto falha significativamente na outra, garantindo uma avaliação mais equilibrada. Como resultado, o F1-score fornece uma análise mais completa e robusta do desempenho do modelo (Kate *et al.*, 2023).

$$\text{F1-score} = 2 \cdot \frac{\text{Precisão} \cdot \text{Revocação}}{\text{Precisão} + \text{Revocação}} \quad (2.8)$$

3 TRABALHOS RELACIONADOS

Na literatura, são encontrados trabalhos que propõem soluções visando a identificação de dispositivos elétricos por meio das assinaturas de carga presentes nos sinais elétricos (tensão, corrente ou potência). Nesta seção são listados e discutidos trabalhos que apresentam formas de identificação de dispositivos elétricos utilizando modelos de aprendizado de máquina e, além disso, que realizam análise de desempenho dos algoritmos por meio de métricas como acurácia e F1-score.

3.1 Non-Intrusive Electrical Appliances Monitoring and Classification using K-Nearest Neighbors

Os autores Khan *et al.* (2019) detalham o monitoramento e classificação do sinal elétrico de dispositivos elétricos realizado. Esse projeto tem como intuito incentivar mudanças no comportamento de consumo que facilitam a conservação de energia, com a justificativa de que os medidores de eletricidade existentes fornecem pouca ou nenhuma informação sobre os consumos energéticos dos dispositivos de forma individual. Em vista disso, é empregado o método *Non-Intrusive Load Monitoring (NILM)* no processo de amostragem e aplicado o algoritmo *kNN*, que realiza a classificação observando especificamente os 5 vizinhos mais próximos.

Na metodologia de treinamento foi usado o conjunto de dados *Reference Energy Disaggregation Data Set (REDD)*, separando 90% dos dados para treino do modelo e 10% para teste. Em resumo, são apresentados os resultados da classificação de 7 dispositivos elétricos. Na análise dos resultados, embora o classificador tenha confundido os dispositivos eletrônicos retornando uma acurácia de 50%, nos demais casos teve um bom desempenho com acurácias entre 90% e 100%. Todavia, a não avaliação de outros algoritmos surge como uma das limitações do trabalho que impossibilita a obtenção de resultados melhores. Além disso, outra limitação se pendura no fato de que os autores utilizam apenas um valor para o parâmetro k , além de não justificarem essa escolha.

Nesta proposta, também é empregado o algoritmo *kNN* para classificar os sinais de corrente e, conseqüentemente, identificar o dispositivo proprietário. Contudo, são realizados experimentos para definir o melhor valor de k possível, fazendo comparações entre resultados do classificador com diferentes valores para k . Além disso, nesta proposta empregam-se outros tipos de classificadores, visando desvendar qual técnica se sobressai no contexto de classificação

de sinais elétricos.

3.2 A feature-based model for the identification of electrical devices in smart environments

Os autores Tundis *et al.* (2019) tratam a identificação de dispositivos elétricos como uma das soluções para constatar anomalias no consumo de energia. Para tanto, fazem a comparação das variáveis reais com as informadas pelo fabricante do dispositivo reconhecido. No processo de identificação, foram utilizados modelos de aprendizado de máquina como *Random Forest*, *LogitBoost*, *Bagging*, *Árvore de Decisão*, *Naive Bayes* e *SVM*. Além disso, as métricas empregadas foram acurácia, precisão, revocação, positivos verdadeiros e falsos, além da importância de cada atributo.

Em seguida, os autores mostram, de forma detalhada, toda a arquitetura do projeto e qual a função de cada bloco da estrutura. Por fim, o melhor resultado de acurácia é alcançado pelo classificador *Random Forest*, com 96,51% de acertos. Em contrapartida, o pior resultado de acurácia é conseguido pelo modelo *SVM*, com 90,11% de acertos. Como é possível notar, até mesmo os piores resultados mantiveram um alto nível de acertos.

Nesta proposta, também são empregadas várias ideias de algoritmos a fim de constatar o melhor para esse contexto de classificação. Todavia, é utilizado o domínio da frequência na construção do conjunto de dados, visando diminuir o número de datagramas na comunicação entre dispositivos de *IoT*. Ainda, são realizados testes para definir as configurações de parâmetros dos modelos. Além disso, as métricas empregadas são acurácia e F1-score que em seu cálculo inclui as métricas de precisão e revocação.

3.3 Multi-Scale Electrical Appliance Load Signature for Non-Intrusive Load Monitoring Classification

No trabalho de Chou *et al.* (2022) é detalhada a identificação de dispositivos elétricos por meio da assinatura de multirresolução denominada *Triangular Area Representation (TAR)*. Para tanto, é utilizado o modo *NILM* na extração da assinatura. A *TAR* é calculada a partir da trajetória V-I dos sinais de corrente e tensão de um ciclo em regiões de estado estacionário. Essas trajetórias são baseadas em imagens representadas por matrizes 2D com alto espaço dimensional. Em seguida, a técnica de correção de fase de *Fourier* é utilizada para evitar baixa precisão ou incompatibilidade na classificação devido à seleção do ponto de partida de um sinal de período.

Logo após, o *Principal Component Analysis (PCA)* é aplicado à assinatura *TAR* para criar um recurso de espaço de baixa dimensão com apenas componentes importantes.

Depois que as imagens de trajetórias são construídas e processadas, o modelo de classificação *kNN* é empregado para reconhecimento do dispositivo. Nessa ideia, foram feitos testes com diferentes valores para o parâmetro *k*. No fim, o melhor resultado foi alcançado com *k* igual a 1, com 97,43% de acurácia. Além disso, cada classe de dispositivo atingiu uma taxa média de sucesso superior a 96,94%.

Analisando os resultados, a classificação por trajetória de sinais elétricos se mostra um caminho promissor na identificação de dispositivos. Contudo, a representação de imagens por matrizes multidimensionais pode acarretar inúmeros problemas de processamento e memória em um ambiente *IoT* com recursos limitados. Nesta proposta, são utilizadas ideias que buscam diminuir o número de dados visando a implantação em um ambiente *IoT*. Além disso, é realizado um comparativo entre resultados de vários classificadores, buscando a melhor escolha para esse cenário.

3.4 Multi-Label Classification Based on Random Forest Algorithm for Non-Intrusive Load Monitoring System

O trabalho dos autores Wu *et al.* (2019) propõe um método de classificação multirrótulo usando o modelo *Random Forest* como algoritmo de aprendizado para identificação de carga não intrusiva. A escolha do classificador ocorreu por meio da análise de resultados de vários modelos como *MLP*, *SVC*, *kNN*, entre outros. Vale ressaltar que a classificação multirrótulo pode ser usada para determinar a quais categorias os dados pertencem. Nessa perspectiva, esta classificação pode ajudar a identificar os estados de operação de cargas independentes de sinais mistos sem desagregação.

De início, os experimentos foram conduzidos em um ambiente real; posteriormente, foi usado um conjunto de dados públicos para os testes. Ademais, foram utilizados valores do sinal de corrente no domínio do tempo, visando a identificação do dispositivo por assinaturas presentes no formato da onda. Em síntese, os resultados de acurácia e F1-score do *Random Forest* chegaram a 97% e 98%, respectivamente.

Semelhantemente, esta proposta realiza a análise de vários métodos de classificação, objetivando encontrar o mais adequado para o contexto de identificação de dispositivos elétricos. Ainda, utiliza as mesmas métricas de acurácia e F1-score na avaliação dos resultados. Porém,

aplica a ideia de classificação dos sinais no domínio da frequência, visando diminuir o fluxo de dados em um ambiente *IoT*.

3.5 Non-Intrusive Load Disaggregation by Convolutional Neural Network and Multilabel Classification

Os autores Massidda *et al.* (2020) descrevem uma nova metodologia de monitoramento e desagregação de cargas utilizando o método *NILM*, fundamentada no reconhecimento dos estados de ativação de eletrodomésticos. Nessa perspectiva, a ideia baseia-se na observação de que um usuário suficientemente experiente é capaz de reconhecer esses estados de ativação examinando um gráfico de sinais elétricos agregado e que uma rede neural como a proposta pode emular essa habilidade.

Nesse contexto, a abordagem de utilizar aprendizado profundo obteve resultados promissores com os melhores valores girando em torno de 99% de acurácia e 97% de F1-score. É importante frisar que foram aproveitadas técnicas mais recentes usadas para segmentação semântica de imagens, introduzindo um modelo baseado em redes convolucionais, aplicável em conjuntos de dados com baixa taxa de amostragem.

Nesta proposta, não foram utilizadas técnicas complexas de aprendizado profundo, visto que, o foco de uma possível implantação em um ambiente *IoT* direcionou os testes para modelos de classificação mais simples. Além disso, esta dissertação utiliza o domínio da frequência como solução para o alto fluxo de dados, visando diminuir o tempo de treinamento e classificação. Contudo, os métodos de avaliação se assemelham com os dos autores, além da utilização de métodos de aprendizado em cargas não intrusivas.

3.6 Comparação entre trabalhos

O Quadro 2 confronta os trabalhos, relacionados e o proposto, em relação aos quesitos de métodos de classificação, variáveis, conjunto de dados, métricas e acurácia. Deste modo, é possível analisar, de forma resumida, os algoritmos de classificação utilizados, quais variáveis foram utilizadas para identificar os dispositivos, quais conjuntos de dados foram empregados, as métricas usadas na avaliação e os resultados de acurácia alcançados nos experimentos.

No que se refere aos métodos de classificação, Khan *et al.* (2019) e Chou *et al.* (2022) optaram por usar apenas o *kNN*, o que indica uma limitação, pois uma análise mais aprofundada

Quadro 2 – Comparação entre os trabalhos relacionados e o proposto.

Trabalho	Khan <i>et al.</i> (2019)	Tundis <i>et al.</i> (2019)	Chou <i>et al.</i> (2022)	Wu <i>et al.</i> (2019)	Massidda <i>et al.</i> (2020)	Esta proposta
Métodos de Classificação	<i>kNN</i>	<i>Random Forest, LogisticBoost, Bagging, Árvore de Decisão, Naive Bayes e SVM</i>	<i>kNN</i>	<i>Random Forest, MLP, SVC e kNN</i>	<i>RNAs</i>	<i>kNN, SVM, Árvore de Decisão, Random Forest, Gradient Boosting e MLP</i>
Variáveis	Sinal de potência no domínio do tempo	Potência média de pico, potência média ativa, máxima potência, tempo de potência ativa, etc	Imagens 2D de trajetórias dos sinais de corrente e tensão	Sinal de corrente no domínio do tempo	Imagens do sinal de potência no domínio do tempo	Sinal de corrente no domínio da frequência
Conjunto de Dados	<i>Reference Energy Disaggregation Data Set (REDD)</i>	Conjunto de dados público disponibilizado sob a <i>Open Database License (ODbL)</i>	<i>Plug-Load Appliance Identification Dataset (PLAID)</i>	Conjunto de dados próprio e conjunto de dados público <i>BLUED</i>	<i>UK-DALE</i>	Conjunto de dados próprio
Métricas	Acurácia, precisão, revocação e F1-score	Acurácia, precisão, revocação e positivos verdadeiros e falsos	Acurácia, precisão, revocação e F1-score	Acurácia e F1-score	Acurácia, precisão, revocação e F1-score	Acurácia e F1-score
Acurácia	90% a 100%	90% a 96%	97%	97% a 98%	97% a 99%	99%

Fonte: Elaborado pelo próprio autor.

de outros classificadores poderia obter um resultado mais satisfatório no contexto do projeto. Diferentemente, os autores Massidda *et al.* (2020) optaram pelo emprego do aprendizado profundo, realizando um estudo com várias Redes Neurais Artificiais (RNAs) objetivando o melhor resultado de identificação. Em contrapartida, os demais trabalhos, incluindo este, aplicaram vários modelos de aprendizado de máquina buscando comparar os resultados com o intuito de selecionar o melhor algoritmo.

Com relação às variáveis observadas para classificar e identificar os dispositivos, os trabalhos de Khan *et al.* (2019) e Tundis *et al.* (2019) analisam o formato e características presentes no sinal de potência dos dispositivos. Por outro lado, os autores Chou *et al.* (2022) e Massidda *et al.* (2020) fazem a classificação por imagens dos sinais de corrente e tensão. Os autores Wu *et al.* (2019) e esta proposta realizam a identificação por meio de amostras do sinal de corrente. Todavia, esta dissertação optou pelo domínio da frequência com vistas a diminuir a quantidade de dados.

Em relação ao conjunto de dados, este trabalho optou por construir um conjunto próprio, buscando reduzir a quantidade de dados ao utilizar o domínio da frequência. De forma semelhante, os autores Wu *et al.* (2019) utilizaram um conjunto de dados próprio, além de um

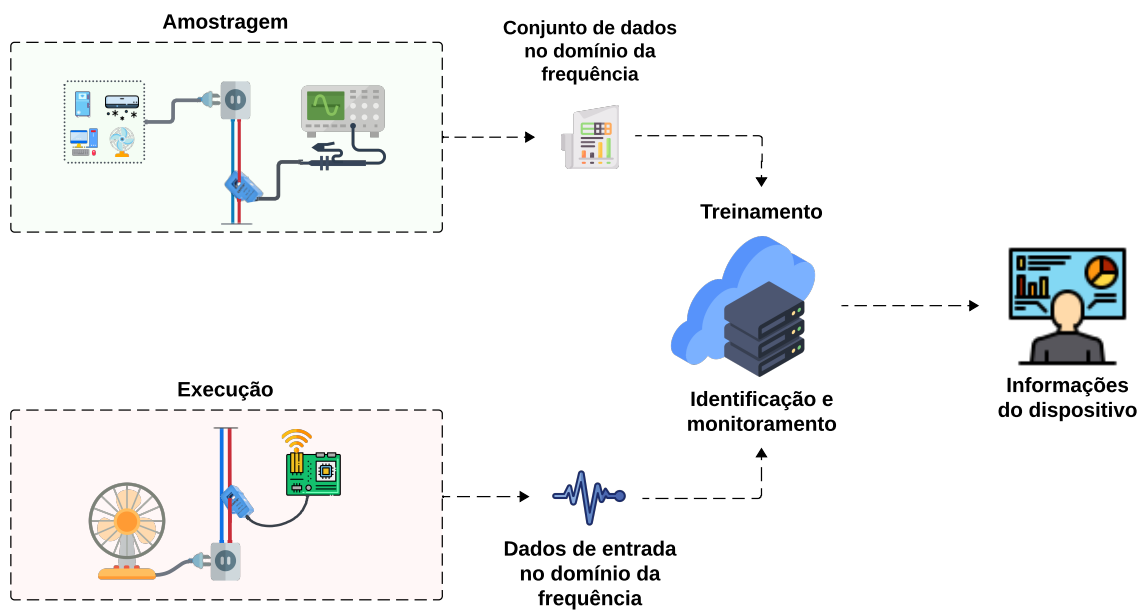
conjunto público. Os demais trabalhos utilizaram apenas conjuntos de dados públicos.

Considerando as métricas de avaliação, todos os trabalhos optaram de maneira semelhante por utilizar a acurácia como principal métrica. Nesse contexto, o trabalho de Khan *et al.* (2019) alcançou resultados de acurácia entre 90% e 100%, dependendo do dispositivo. Semelhantemente, o trabalho de Tundis *et al.* (2019) obteve acurácias entre 90% e 96%. Os autores Chou *et al.* (2022) conseguiram uma acurácia média de 97%. Os trabalhos de Wu *et al.* (2019) e Massidda *et al.* (2020) obtiveram uma acurácia mínima de 97% e uma máxima de 98% e 99%, respectivamente. Esta dissertação alcançou uma acurácia média de, aproximadamente, 99%, como será mostrado no Capítulo 5.

4 METODOLOGIA

Neste capítulo, foram abordadas as etapas necessárias para a realização deste estudo. A Figura 8 mostra uma visão geral do projeto completo. Na etapa de amostragem, foram coletados sinais de corrente de dispositivos elétricos domésticos, conectados individualmente a uma tomada. Utilizando um sensor não intrusivo e um osciloscópio, capturamos as amostras de corrente elétrica. Em seguida, essas amostras foram pré-processadas e convertidas em dados no domínio da frequência, criando um conjunto de dados que associa o nome do dispositivo às 20 primeiras componentes de frequência do seu sinal de corrente. Este conjunto foi utilizado para treinar modelos de aprendizado de máquina, armazenados em um servidor central. Esses modelos pré-treinados podem então identificar os dispositivos elétricos durante a etapa de execução.

Figura 8 – Visão geral do processo de monitoramento e identificação de dispositivos elétricos domésticos.



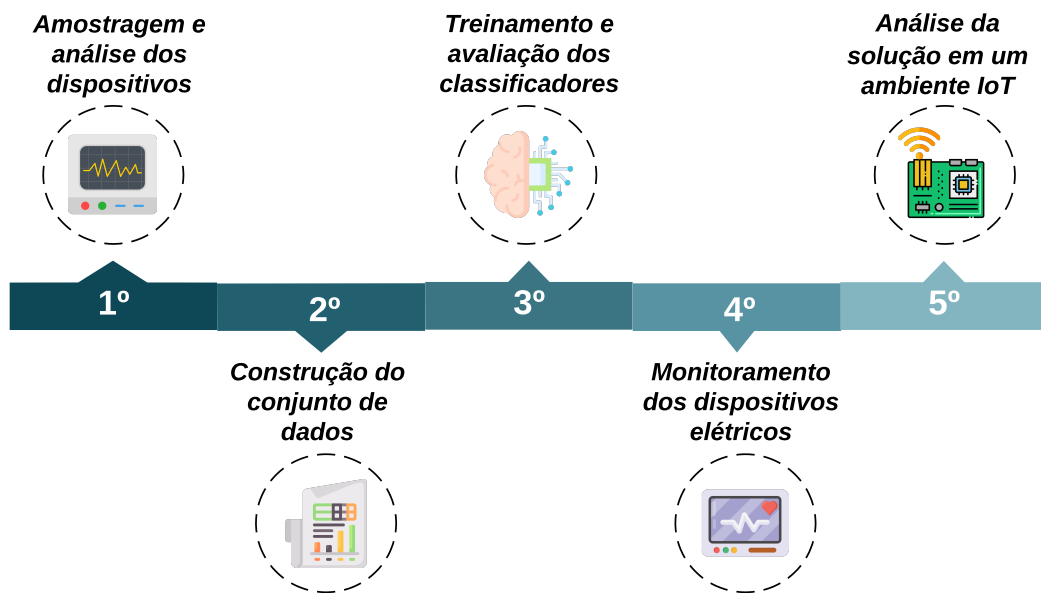
Fonte: Elaborada pelo próprio autor.

Na etapa de execução, cada tomada é monitorada por um dispositivo *IoT* equipado com um sensor não intrusivo. Quando um dispositivo elétrico é ligado à tomada, o dispositivo *IoT* já instalado captura os dados de corrente elétrica, os transforma para o domínio da frequência e envia esses dados para o servidor central via comunicação sem fio. O servidor central recebe as amostras e identifica o dispositivo elétrico usando os modelos de classificação pré-treinados. Além disso, o servidor avalia a saúde do dispositivo comparando os dados de corrente com as

especificações do fabricante. Por fim, essas informações são exibidas em uma aplicação.

O escopo deste trabalho compreende a identificação e o monitoramento de dispositivos elétricos por meio de modelos de aprendizado de máquina. Dessa forma, os esforços foram direcionados às etapas essenciais para atingir essa meta. Na Seção 4.1, é apresentado o procedimento de amostragem e análise dos sinais de corrente dos dispositivos elétricos. Na Seção 4.2, é detalhada a construção do conjunto de dados proposto, explicando as informações necessárias e sua organização. Na Seção 4.3, são fornecidas informações sobre o treinamento dos modelos de classificação, os classificadores utilizados e os detalhes da avaliação. Na Seção 4.4, é descrito o procedimento adotado para monitorar a condição operacional dos dispositivos elétricos. Por fim, na Seção 4.5, é realizada uma análise sobre a aplicação potencial deste estudo em um contexto de *IoT*, destacando suas implicações e considerações relevantes. A Figura 9 ilustra a sequência desses procedimentos.

Figura 9 – Sequência dos procedimentos metodológicos empregados neste trabalho.



Fonte: Elaborada pelo próprio autor.

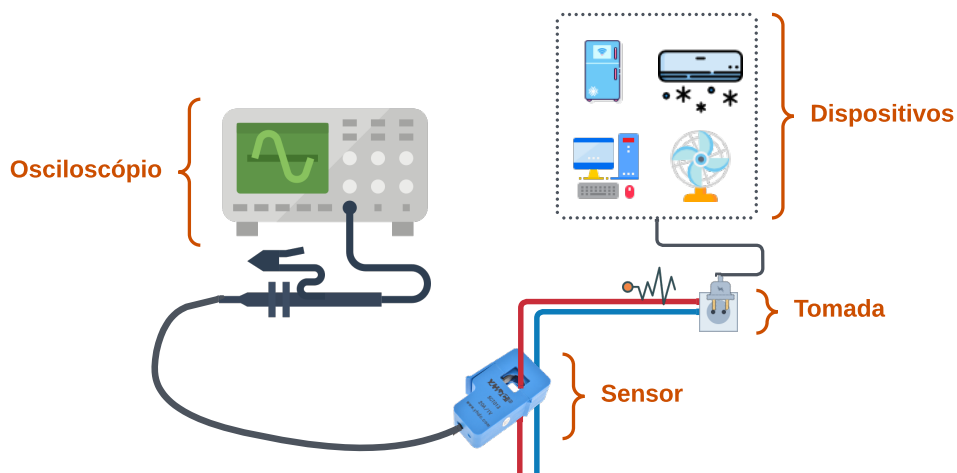
4.1 Amostragem e análise dos dispositivos

A etapa de amostragem desempenha um papel essencial na execução deste estudo, pois é fundamental para a análise dos sinais de corrente dos dispositivos e para a criação do conjunto de dados. Para esta finalidade, é vantajoso utilizar um sensor capaz de registrar amostras

do sinal de corrente consumida por cada dispositivo que será monitorado. Nesse sentido, o sensor *SCT-013-000* se destaca como uma escolha prática, dada sua capacidade de operação não intrusiva e sua eficiência na leitura de corrente, utilizando princípios do eletromagnetismo. Vale ressaltar que o sensor *SCT-013-000* é uma alternativa viável para essa aplicação. No entanto, podem ser utilizadas outras opções de sensores não intrusivos.

Adicionalmente, o sensor *SCT-013-000* integra-se facilmente com conversores analógico-digitais e microcontroladores, facilitando sua implementação em sistemas de monitoramento de corrente. No entanto, como a construção do dispositivo *IoT* não está dentro do escopo desta proposta, a leitura das amostras pode ser conduzida utilizando um osciloscópio conectado ao sensor não intrusivo. Dessa maneira, será possível analisar e registrar as amostras do sinal de corrente de cada dispositivo inspecionado. As conexões adotadas para realizar a amostragem são esquematizadas na Figura 10.

Figura 10 – Conexões adotadas para capturar amostras de cargas elétricas de dispositivos elétricos domésticos de forma não intrusiva.



Fonte: Elaborada pelo próprio autor.

Como ilustrado na Figura 10, o processo de coleta de amostras é realizado monitorando o sinal de corrente necessário da rede elétrica, à qual diversos dispositivos com diferentes características elétricas podem ser conectados individualmente¹. De acordo com essa abordagem, um sensor não intrusivo é colocado ao redor do cabo de fase para capturar o sinal de corrente elétrica. Posteriormente, um osciloscópio é utilizado para registrar e interpretar as informações fornecidas pelo sensor. Além disso, o osciloscópio oferece uma variedade de funcionalidades

¹ Todas as medições foram realizadas em uma tomada residencial de 10 A, conforme a norma NBR14136 da ABNT.

adicionais, incluindo ajuste da taxa de amostragem, realização da *FFT*, análise dos valores de amplitude e frequência, aplicação de filtros e a capacidade de salvar dados em diversos formatos de arquivo.

Ainda nesta etapa, é demonstrado o processo de identificação dos dispositivos elétricos, fundamentado nos padrões presentes em seus sinais de corrente elétrica. Essa abordagem possibilita a análise dos sinais por meio de seus padrões de onda e das componentes de frequência, utilizando a técnica de *FFT*.

4.1.1 Sensor SCT-013-000

Os sensores desempenham um papel crucial ao produzir respostas, como sinais elétricos, sempre que há alterações nas variáveis monitoradas do ambiente. Eles desempenham uma função essencial na coleta de dados em diversos cenários, capturando e processando valores para detectar mudanças físicas no ambiente. Além disso, os sensores podem ser categorizados com base em suas especificações, métodos de conversão, materiais utilizados, variáveis físicas detectadas e propriedades medidas (Sehrawat; Gill, 2019).

Na *IoT*, os sensores desempenham um papel fundamental no monitoramento de diversos fatores físicos e na detecção de valores relevantes. Além disso, convertem os dados capturados em sinais elétricos, mecânicos, ópticos e outros tipos de sinais (Solovev; Shadrin, 2017). Neste projeto, optou-se pelo uso do sensor *SCT-013-000* (*YHDC, Beijing, China*) no circuito de amostragem, com o objetivo de capturar amostras do sinal de corrente dos dispositivos elétricos.

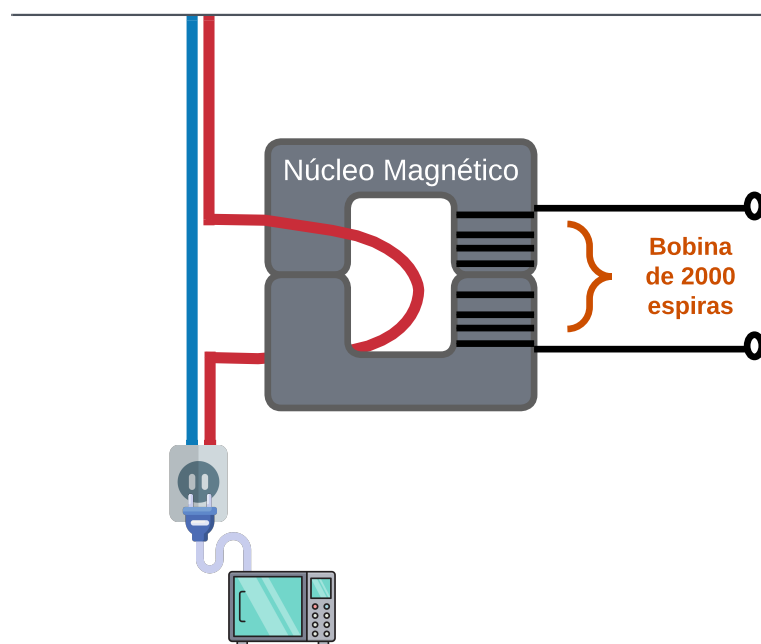
Os sensores do tipo *Split-core Current Transformer (SCT)* utilizam os princípios fundamentais do eletromagnetismo para realizar medições dos sinais de corrente elétrica. Em particular, o sensor *SCT-013-000* é composto por uma bobina contendo 2 000 espiras, juntamente com um núcleo magnético dividido ao meio. Essa característica permite que o sensor seja aberto e envolva o fio elétrico a ser monitorado. Essa propriedade confere a ele uma vantagem significativa, pois torna-o não intrusivo, ou seja, não é necessário interromper o sistema elétrico para realizar as medições.

Em suma, o funcionamento do sensor se baseia na Lei de Ampère, que estabelece que um fio condutor percorrido por corrente elétrica gera um campo magnético ao seu redor, sendo este campo magnético proporcional à corrente que o atravessa (Dias; Dias, 2015). Como o sensor monitora correntes alternadas, o campo magnético gerado varia no tempo. Além disso,

ele se baseia na Lei de Faraday, que afirma que um campo magnético variável no tempo induz uma força eletromotriz em uma espira condutora, resultando na geração de uma corrente elétrica proporcional à intensidade do campo magnético. Esses princípios físicos fundamentais são essenciais para a compreensão do funcionamento do sensor (Hessel *et al.*, 2015).

Com base nessas leis físicas, o sensor opera da seguinte forma: uma metade do núcleo magnético contém uma bobina de 2 000 espiras, enquanto a outra metade é envolvida pelo fio elétrico a ser monitorado, que funciona como uma única espira. Isso pode ser visualizado na Figura 11. Como resultado, os valores medidos pelo sensor serão inversamente proporcionais ao número de espiras da bobina secundária, que é de 2 000 espiras.

Figura 11 – Arquitetura detalhada dos principais componentes do sensor *SCT-013-000*



Fonte: Elaborada pelo próprio autor

De acordo com essa relação, o cálculo de proporção é expresso na equação 4.1, em que N_1 representa o número de espiras na bobina primária, que é uma única espira, e N_2 é o número de espiras na bobina secundária, que são 2 000 voltas. Assim, I_1 é o valor da corrente no fio e I_2 é o valor capturado pelo sensor. Com base nisso, o sensor *SCT-013-000* foi configurado para medir um valor máximo de corrente de 100 A. Portanto, de acordo com o cálculo de proporção, o valor máximo indicado pelo sensor é de 50 mA.

$$\frac{N_1}{N_2} = \frac{I_2}{I_1} \quad (4.1)$$

Em síntese, o sensor *SCT-013-000* desempenha um papel fundamental na captura de dados para este projeto, baseando-se nos princípios da Lei de Ampère e na Lei de Faraday. Sua capacidade de medir corrente elétrica sem a necessidade de interromper as conexões do circuito o torna uma escolha eficiente e conveniente. Além disso, sua concepção com uma bobina secundária de 2 000 espiras permite uma leitura precisa e proporcional da corrente elétrica, tornando-o adequado para uma variedade de aplicações. Ao projetar um circuito de amostragem em torno deste sensor, é possível obter valores exatos de corrente elétrica, fornecendo dados confiáveis para análise e identificação de dispositivos elétricos.

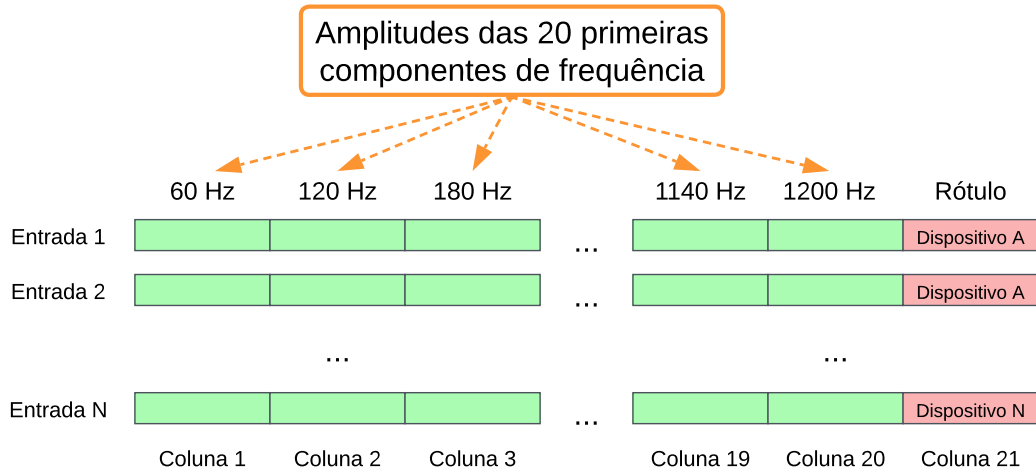
4.2 Construção do conjunto de dados

Nesta seção, é descrito o processo de construção do conjunto de dados utilizado neste estudo. A elaboração de um conjunto de dados de qualidade é fundamental para garantir a eficácia e a validade dos experimentos realizados. O conjunto de dados é a base sobre a qual os algoritmos de aprendizado de máquina serão treinados e testados, e, portanto, sua construção requer cuidado e atenção aos detalhes. É abordada a preparação e organização dos dados, que é fundamental para viabilizar análises posteriores.

Cada registro no conjunto de dados proposto contém as 20 primeiras componentes de frequência do sinal, que englobam a frequência principal e suas harmônicas, acompanhadas por um rótulo identificador do dispositivo. Dentro deste conjunto, estão organizadas 21 colunas, sendo a última designada para a identificação do dispositivo. Assim, múltiplas entradas representam cada dispositivo, exibindo as 20 primeiras componentes de frequência da corrente em momentos distintos. O número de entradas varia conforme o tamanho da amostra. A Figura 12 ilustra a estrutura do conjunto de dados.

A decisão de observar as 20 primeiras componentes de frequência no conjunto de dados é fundamentada em algumas razões apresentadas na sequência. Os principais harmônicos dos dispositivos elétricos se concentram próximos da frequência fundamental (Hu *et al.*, 2023). Isso possibilita uma análise detalhada das características de frequência dos dispositivos, aumentando as chances de sucesso da identificação. Além disso, ao abranger uma gama ampla de componentes de frequência, torna-se possível identificar padrões específicos nos sinais de

Figura 12 – Estrutura do conjunto de dados contendo as 20 primeiras componentes de frequência e os rótulos dos dispositivos.



Fonte: Elaborada pelo próprio autor

diferentes dispositivos, o que contribui para uma classificação mais eficaz e confiável.

Em suma, essa abordagem possibilita a identificação dos dispositivos analisando as amplitudes das diferentes componentes de frequência no sinal elétrico. Em contraste com uma análise temporal extensiva, que demandaria uma enorme quantidade de amostras para representar os diversos períodos do sinal, a análise de frequência oferece uma solução mais eficiente em termos de armazenamento e processamento de dados.

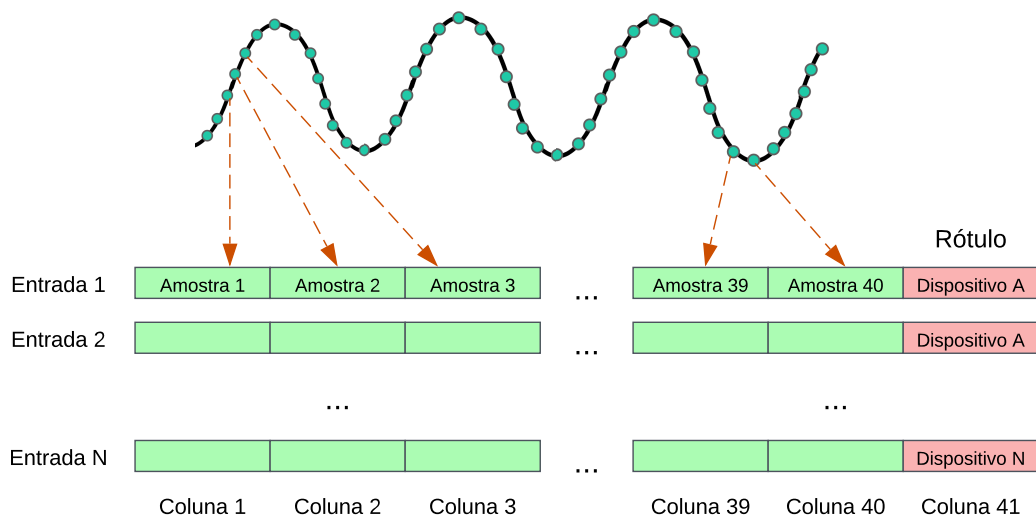
4.2.1 Dados no domínio do tempo

Para uma análise comparativa mais abrangente, é proposta a construção também de um conjunto de dados que incorpore amostras dos sinais de corrente no domínio do tempo. Este conjunto seria estruturado de forma a incluir amostras sequenciais do sinal de corrente. Assim, teríamos múltiplas colunas representando amostras consecutivas do sinal de corrente, o número de colunas sendo o suficiente para abranger o mesmo número de ciclos do sinal que o conjunto de dados no domínio da frequência englobou. A última coluna, por sua vez, conteria a etiqueta identificadora do dispositivo correspondente.

Portanto, o conjunto de dados contém várias entradas vinculadas a cada dispositivo presente. Cada uma dessas entradas representa múltiplas amostras do sinal de corrente consumida em momentos específicos no tempo. Esse formato possibilita a identificação de cada aparelho

elétrico com base em características singulares presentes em seu sinal de corrente, tais como os valores máximos, mínimos e a forma da onda. A Figura 13 ilustra a estrutura do conjunto de dados no domínio do tempo.

Figura 13 – Estrutura do conjunto de dados contendo amostras no domínio do tempo e os rótulos dos dispositivos.



Fonte: Elaborada pelo próprio autor

Destaca-se a importância de assegurar uma comparação justa, o que demanda que cada entrada nos conjuntos de dados, tanto no domínio da frequência quanto no domínio do tempo, seja adquirida com a mesma taxa de amostragem.

4.3 Treinamento e avaliação dos classificadores

Nesta etapa, são utilizados diversos modelos de aprendizado de máquina, como o *kNN*, *SVM*, *Árvore de Decisão*, *Random Forest*, *Gradient Boosting* e *MLP*, todos implementados em *Python* com o suporte das bibliotecas *Numpy*², *Pandas*³ e *Scikit-Learn*⁴. As bibliotecas *Numpy* e *Pandas* desempenham um papel crucial na manipulação de dados e foram aplicadas para essa finalidade.

² <https://numpy.org/>

³ <https://pandas.pydata.org/>

⁴ <https://scikit-learn.org/stable/>

A biblioteca *Scikit-Learn*, uma ferramenta de código aberto amplamente utilizada, oferece uma variedade de funcionalidades essenciais para aprendizado de máquina. Entre essas funcionalidades, destacam-se algoritmos para regressão e classificação, além de recursos para pré-processamento de dados. Além disso, o *Scikit-Learn* fornece métricas importantes para avaliação de modelos, como acurácia, precisão, revocação e F1-score, que são fundamentais para compreender o desempenho dos algoritmos em diferentes tarefas de aprendizado de máquina.

Dentre as métricas disponíveis, optou-se por utilizar o F1-score e a acurácia para avaliar a eficácia da identificação realizada. O F1-score é uma métrica que combina precisão e revocação em um único valor, buscando equilibrar a capacidade do modelo em evitar classificações equivocadas (precisão) e identificar corretamente todos os exemplos relevantes (revocação) na classe positiva. A precisão mede a habilidade do modelo de evitar classificar incorretamente exemplos negativos como positivos, enquanto a revocação avalia a capacidade do modelo de identificar corretamente todos os exemplos positivos.

Além disso, é importante mencionar que foi utilizado o *Google Colaboratory*⁵ como ambiente de desenvolvimento para a implementação dos modelos de aprendizado de máquina. O *Google Colab*, como é popularmente conhecido, é uma plataforma baseada na nuvem que oferece acesso gratuito a *GPUs* e *TPUs* para a execução de código *Python*.

Isso possibilita a execução rápida de algoritmos de aprendizado de máquina e aprendizado profundo, além de fornecer recursos de colaboração que permitem o compartilhamento e a edição colaborativa de *notebooks Jupyter*. Essa plataforma é amplamente utilizada pela comunidade de ciência de dados e aprendizado de máquina devido à sua praticidade, facilidade de uso e recursos computacionais disponíveis.

Assim, cada algoritmo de classificação foi ajustado meticulosamente para encontrar a combinação ideal, ou pelo menos satisfatória, de parâmetros, visando obter resultados positivos nas métricas selecionadas. Para isso, a biblioteca *Scikit-Learn* foi empregada para explorar diversas configurações de hiperparâmetros e realizar validação cruzada. Essa abordagem envolve a divisão do conjunto de dados em subconjuntos para treinamento e teste, permitindo a comparação do desempenho dos classificadores. O hiperparâmetro *k-fold* define em quantas partes os dados serão divididos para validação cruzada. Neste trabalho, optou-se por utilizar um *k-fold* de 10. O objetivo final é identificar o modelo mais eficaz para a tarefa de classificação em questão.

Além da validação cruzada, é importante ressaltar que, durante o processo de ajuste

⁵ <https://colab.research.google.com/>

de hiperparâmetros e treinamento dos modelos, apenas 60% dos dados foram utilizados, correspondentes ao conjunto de treinamento. Em outras palavras, durante o treinamento, nenhum modelo teve acesso aos 40% restantes dos dados, reservados para o conjunto de teste. Essa abordagem foi adotada para evitar resultados enviesados devido ao sobreajuste. Todos esses cuidados foram tomados ao longo de todos os experimentos realizados. A divisão entre os conjuntos de treinamento e teste seguiu uma proporção de 60/40, determinada pela quantidade de dados disponíveis, sendo realizada de forma aleatória em um conjunto de dados balanceado.

Além dos resultados obtidos na identificação dos dispositivos, outros experimentos foram conduzidos para explorar diferentes aspectos do problema. Estes incluíram a comparação entre os resultados obtidos no domínio da frequência e no domínio do tempo, a avaliação do impacto da presença simultânea de dispositivos na rede elétrica, a classificação de dispositivos desconhecidos e a análise da adição de novos dispositivos ao conjunto de dados. Cada um desses experimentos proporcionou percepções valiosas sobre a robustez e a eficácia dos classificadores em diversos cenários e condições. Os conjuntos de dados e os códigos-fonte⁶ criados para os experimentos estão disponíveis para a comunidade.

4.4 Monitoramento dos dispositivos elétricos

Para realizar o monitoramento da saúde dos dispositivos elétricos, foi adotada uma abordagem sistemática que envolveu várias etapas. Primeiramente, os dispositivos foram identificados por meio de técnicas de classificação utilizando dados de corrente elétrica no domínio da frequência. Uma vez identificados, foi iniciado o processo de monitoramento para verificar se os dispositivos estavam operando dentro dos padrões esperados.

Em seguida, a atenção é focalizada na análise da amplitude da componente de frequência fundamental de cada dispositivo. Esta análise foi realizada comparando a amplitude da frequência fundamental de cada dispositivo com a média calculada a partir do conjunto de dados. Qualquer desvio significativo dessa média poderia indicar potenciais problemas de funcionamento nos dispositivos.

Para definir um critério de anomalias, foi estabelecido um limiar de 30% em relação à média da amplitude da frequência fundamental. Caso a amplitude de um dispositivo ultrapassasse esse limiar, era considerada uma anomalia e o dispositivo era marcado como suspeito de apresentar defeitos. A escolha do limiar de 30% foi baseada em uma análise empírica dos

⁶ https://github.com/davidamurim7/codigos_dissertacao

valores disponíveis no conjunto de dados. Inicialmente, foi observado que, ao analisar os dados de corrente dos dispositivos, a amplitude da frequência fundamental não ultrapassava 30% em relação à média. Essa observação levou à definição inicial do limiar de 30%, assumindo que esses dispositivos estavam operando corretamente.

No entanto, é importante ressaltar que este limiar não é definitivo. Ele pode ser ajustado com base em especificações técnicas fornecidas pelos fabricantes dos dispositivos, bem como em estudos de referências ou padrões estabelecidos na literatura técnica. Nesse contexto, o limiar de 30% serve como um ponto de partida prático e baseado em observações, mas está sujeito a refinamentos futuros conforme novas informações e critérios mais específicos sejam considerados.

Após a identificação de dispositivos com potenciais problemas, é ressaltado que outras análises detalhadas podem ser realizadas para determinar a natureza das anomalias. Isso pode incluir a verificação de características adicionais dos sinais elétricos, como a presença de harmônicos indesejados, a estabilidade da frequência fundamental ao longo do tempo e a detecção de ruídos ou distorções. Essas informações complementares são cruciais para uma compreensão abrangente da condição de saúde dos dispositivos e para a tomada de medidas corretivas, se necessário.

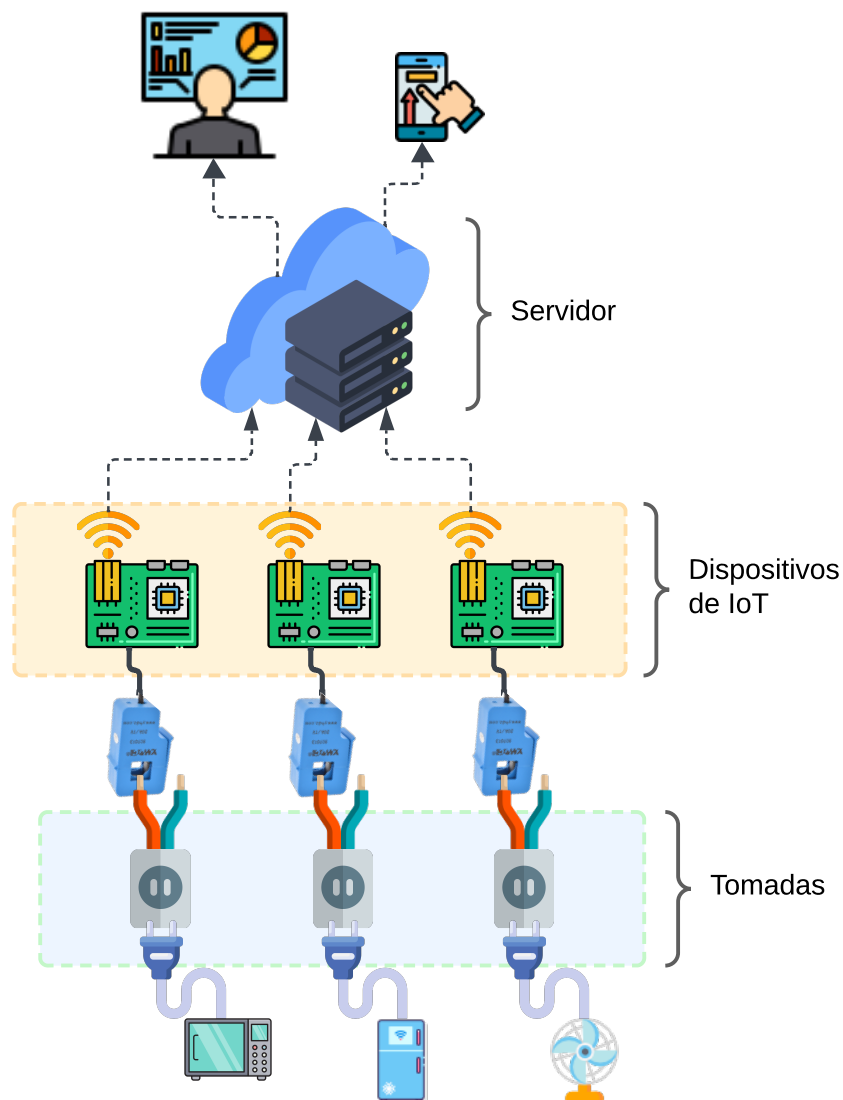
4.5 Análise da solução em um ambiente *IoT*

Em um contexto de *IoT*, enfrentam-se diversos desafios devido à limitação de recursos de processamento e memória. O volume de dados a serem processados ou transmitidos a partir dos sensores pode resultar em problemas como congestionamento de rede, alta latência e perda de pacotes. Nesse contexto, a abordagem de classificação no domínio da frequência se destaca, já que possibilita a transformação de várias amostras do sinal temporal em apenas 20 componentes de frequência por entrada. Isso reduz em até 5 vezes o tamanho dos pacotes a serem transmitidos pela rede, melhorando as métricas de comunicação. Além disso, ao adotar essa abordagem, o conjunto de dados utilizado no treinamento do classificador é consideravelmente reduzido, o que representa uma economia valiosa de recursos.

No entanto, surgem questões intrigantes sobre a localização do cálculo da *FFT* e quais recursos de *IoT* podem ser utilizados para reduzir os custos dessa operação. Nesta seção, será apresentado e analisado um modelo de comunicação, levando em consideração um ambiente *IoT*. Embora a identificação de dispositivos domésticos seja o foco principal deste trabalho, é

relevante considerar uma arquitetura de comunicação que viabilize a implementação prática desse sistema. Assim, ainda que a definição detalhada da infraestrutura de comunicação não esteja dentro do escopo imediato deste trabalho, é proposto um modelo conceitual para ilustrar como os dispositivos de *IoT* poderiam interagir e comunicar os dados relevantes para análise. Um esboço desse modelo de arquitetura de comunicação é ilustrado na Figura 14.

Figura 14 – Proposta de modelo de comunicação entre as diferentes partes do projeto de *IoT*.



Fonte: Elaborada pelo próprio autor.

Nesse modelo proposto, cada tomada doméstica seria equipada com um dispositivo de *IoT* dedicado, responsável por monitorar o terminal de fase da rede elétrica. Quando um

dispositivo doméstico é conectado à tomada, o dispositivo de *IoT* associado captura o sinal de corrente usando um sensor não intrusivo. Posteriormente, para extrair informações sobre as características do dispositivo conectado, é realizada uma *FFT* sobre os dados de corrente capturados.

Uma vez realizada a *FFT*, as amplitudes das 20 primeiras componentes de frequência são extraídas e transmitidas para um servidor centralizado. Este servidor tem a responsabilidade de receber e processar os dados enviados pelos dispositivos de *IoT*. Utilizando algoritmos de aprendizado de máquina previamente treinados, o servidor identifica os padrões associados aos dispositivos domésticos com base nas informações de frequência recebidas. Após a identificação, os resultados podem ser disponibilizados para aplicativos ou sistemas que possam utilizar esses dados para diferentes finalidades, como monitoramento de energia, controle de dispositivos, entre outros.

Essa arquitetura proposta exemplifica como a comunicação entre dispositivos de *IoT* e um servidor central pode facilitar a identificação e o controle de dispositivos domésticos, contribuindo para a automação residencial e a eficiência energética. Embora os detalhes de implementação estejam além do escopo deste estudo, a proposta desse modelo de comunicação oferece uma visão abrangente de como a integração de dispositivos e análise de dados pode ser realizada para alcançar os objetivos propostos neste trabalho.

4.5.1 Implementação da *FFT* em dispositivos de *IoT*

A implementação eficiente da *FFT* em dispositivos de *IoT* é uma questão que requer considerações específicas para garantir o desempenho adequado do sistema. Embora o algoritmo *FFT* esteja amplamente disponível em várias linguagens de programação e seja acessível para implementação em *software*, a obtenção de eficiência máxima para aplicações em tempo real, como as descritas neste contexto, pode demandar abordagens mais especializadas.

Uma alternativa viável é a utilização de circuitos dedicados para a *FFT*, citados na Seção 2.1, que oferecem vantagens significativas em termos de desempenho e eficiência energética. Autores como Arivamudhan e Santhi (2018) e Liu *et al.* (2022) destacam a eficácia desses circuitos para acelerar a execução da *FFT*. Na vasta literatura disponível, há uma variedade de abordagens para a implementação da *FFT*, cada uma com suas próprias características que exploram vantagens distintas, como desempenho, flexibilidade e eficiência energética, entre outros aspectos (Huy *et al.*, 2017; Yi *et al.*, 2023; Arivamudhan; Santhi, 2018; Liu *et al.*, 2022;

Sakashita *et al.*, 2015).

Ao aproveitar circuitos dedicados, é possível obter benefícios tangíveis, incluindo velocidades de processamento mais rápidas e redução significativa do consumo de energia, como observado em estudos como o de Sakashita *et al.* (2015). Portanto, para garantir a eficiência e a rapidez na execução da *FFT* em dispositivos de *IoT*, a consideração de implementações em hardware, como circuitos dedicados, emerge como uma estratégia promissora, capaz de atender às demandas de processamento em tempo real com eficiência energética. Esta abordagem oferece uma solução robusta para lidar com os desafios de processamento de sinais em ambientes de *IoT*, garantindo uma resposta eficaz ao monitoramento e análise dos dados de corrente elétrica.

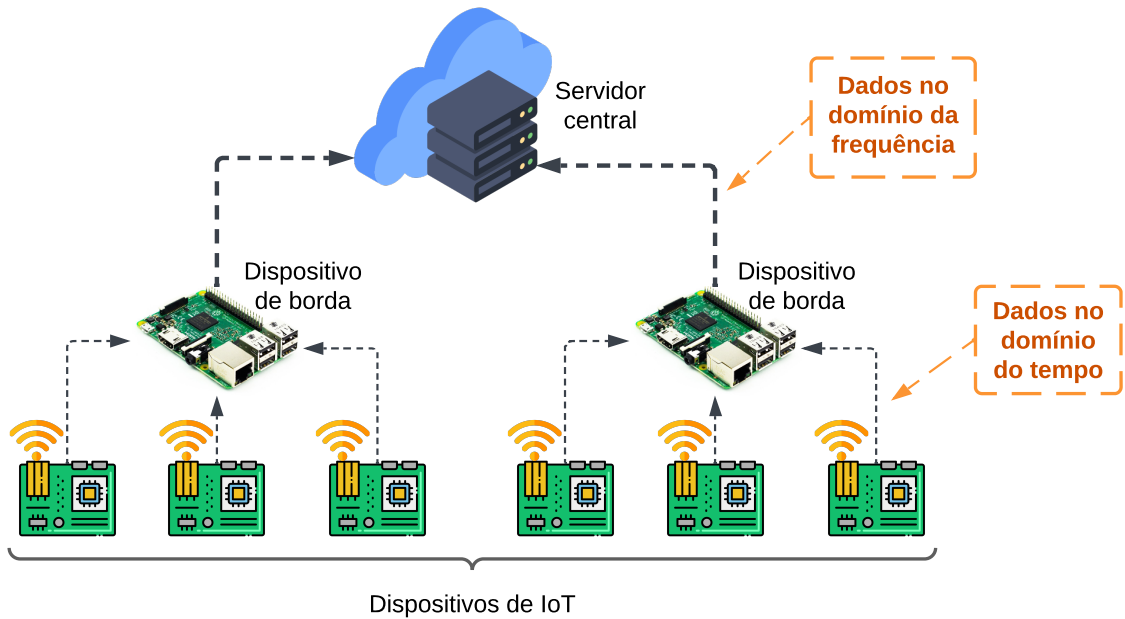
Uma alternativa é a utilização da computação de borda. Essa abordagem busca aproximar o poder de processamento e armazenamento de dados dos dispositivos de *IoT* e dos usuários finais, conforme documentado por autores como Rohith *et al.* (2021) e Ozcan *et al.* (2019). Essa proximidade permite reduzir a latência na troca de dados, minimizar o tráfego na rede e aumentar a eficiência na execução de determinadas operações.

No contexto proposto, a ideia de implementar a *FFT* usando computação de borda é relevante. Isso envolve colocar dispositivos intermediários, chamados de dispositivos de borda, entre os dispositivos de *IoT* e o servidor central. Esses dispositivos de borda são capazes de realizar a *FFT*, seja por meio de hardware dedicado ou de software otimizado, e estão estrategicamente posicionados para servir a um grupo específico de dispositivos de *IoT*, como mostra a Figura 15.

Ao adotar essa abordagem, várias vantagens se destacam. Primeiramente, a proximidade física entre os dispositivos de borda e os dispositivos de *IoT* reduz a latência na comunicação, permitindo respostas mais rápidas às operações realizadas. Além disso, a distribuição do processamento para os dispositivos de borda alivia a carga dos dispositivos de *IoT* e do servidor central, aumentando a escalabilidade do sistema como um todo.

De fato, os dados provenientes dos dispositivos de *IoT* e enviados para os dispositivos de borda permanecem no domínio do tempo até serem transformados em dados de frequência. No entanto, essa desvantagem na transferência de dados é atenuada pela proximidade física dos dispositivos de borda em relação aos dispositivos de *IoT*. Além disso, a utilização de circuitos especializados em *FFT* pode auxiliar no processamento intensivo necessário nos dispositivos de borda. No entanto, vale destacar que o consumo energético adicional permanece uma desvantagem.

Figura 15 – Modelo de comunicação com computação de borda adicionada para realizar o cálculo da *FFT*.



Fonte: Elaborada pelo próprio autor.

O papel dos dispositivos de borda nesse modelo de comunicação é crucial. Eles atuam como pontos de processamento intermediários, realizando tarefas específicas, como a *FFT*, antes de encaminhar os dados processados para o servidor central. Essa arquitetura descentralizada oferece maior flexibilidade e eficiência na operação do sistema de *IoT*, ao mesmo tempo em que mantém a robustez e a capacidade de análise do servidor central.

Em resumo, a integração da computação de borda no modelo de comunicação proposto proporciona benefícios significativos em termos de desempenho, eficiência e escalabilidade. Ao distribuir o processamento de dados de forma inteligente entre os dispositivos de borda e o servidor central, é possível otimizar a operação de sistemas de *IoT* e oferecer uma experiência mais ágil e eficaz para os usuários.

5 RESULTADOS E DISCUSSÕES

Neste capítulo, apresenta-se uma análise detalhada dos resultados obtidos a partir dos experimentos e atividades realizadas. Na Seção 5.1, é descrito o processo de amostragem dos dispositivos. Na Seção 5.2, é feita uma análise sobre as amostras dos sinais capturados dos dispositivos. Na Seção 5.3, são explicadas em detalhes as técnicas empregadas no pré-processamento e na formação do conjunto de dados. Na Seção 5.4, é realizada uma avaliação dos principais parâmetros dos classificadores. Na Seção 5.5, são apresentados e discutidos os resultados da identificação dos dispositivos. Por fim, na Seção 5.6, é abordado o monitoramento do estado de funcionamento dos dispositivos.

5.1 Amostragem

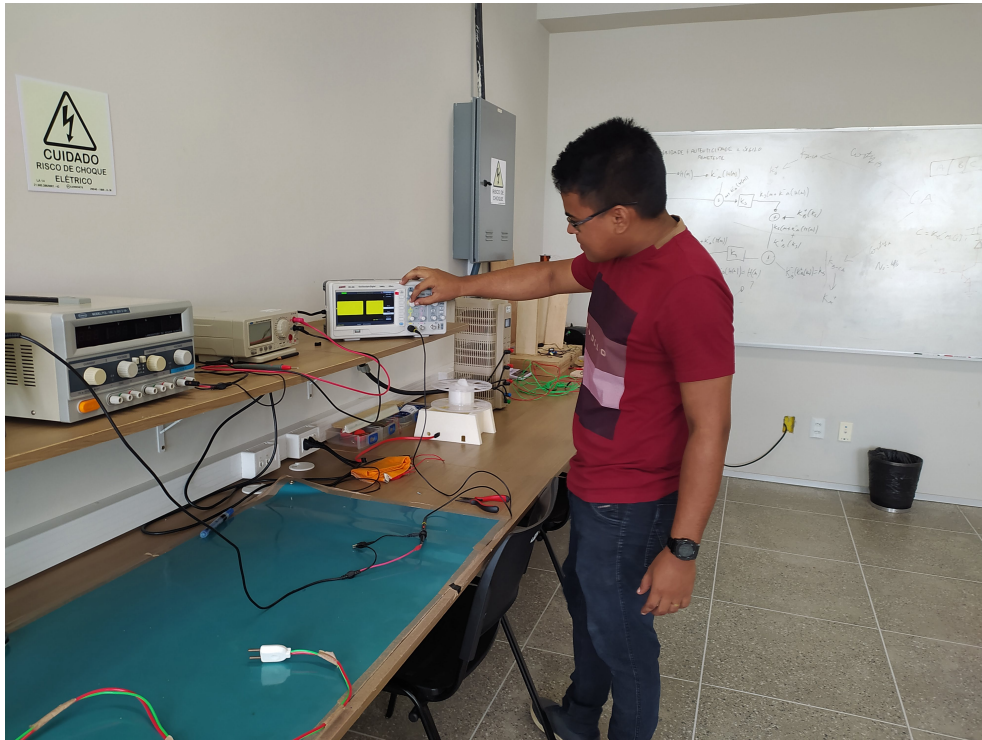
Os procedimentos começaram com a etapa de amostragem, que envolveu a realização de experimentos nas salas e laboratórios do Campus Quixadá da Universidade Federal do Ceará (UFC). Esses experimentos focalizaram dispositivos comuns, como micro-ondas, geladeira, ar condicionado, ventilador e monitor. Todos esses aparelhos são eletricamente alimentados através de uma forma de onda senoidal com frequência de 60 Hz, conforme o sistema elétrico brasileiro. Assim, procedeu-se com a coleta e análise do sinal de corrente elétrica de alimentação de cada um desses dispositivos.

Inicialmente, configurou-se o experimento utilizando os recursos disponíveis no laboratório do Campus. A Figura 16 ilustra o momento em que o experimento já montado está sendo ajustado. Utilizou-se um osciloscópio da marca *Instrutherm*¹, modelo *OD-265*, com uma taxa máxima de amostragem de 10^9 amostras por segundo. Esse osciloscópio foi conectado a um sensor *SCT-013-000*, com capacidade máxima de leitura de 100 A. Por sua vez, o sensor foi colocado em torno do cabo de fase de uma tomada residencial monofásica de 10 A, com cada dispositivo elétrico sendo conectado separadamente. No entanto, nos casos em que os dispositivos eram de difícil deslocamento, suas medições foram realizadas em seus locais de operação, conforme exemplificado na Figura 17, que mostra o micro-ondas e o ventilador que fizeram parte dos experimentos.

Seguindo essa abordagem, aproximadamente 40 000 amostras de corrente elétrica de cada dispositivo foram capturadas e armazenadas em um arquivo CSV, com uma taxa de

¹ <https://www.instrutherm.com.br>

Figura 16 – Imagem capturada durante a montagem do experimento de amostragem.



Fonte: Elaborada pelo próprio autor.

Figura 17 – Medição do ventilador e do micro-ondas em seu local de funcionamento.



Fonte: Elaborada pelo próprio autor.

amostragem de cerca de 2 500 amostras por segundo. É relevante destacar que, embora o osciloscópio possa operar com uma taxa de amostragem mais alta, optou-se por um valor

menor com o objetivo de capturar vários períodos do sinal, evitando a acumulação excessiva de dados. Isso porque uma quantidade excessiva de dados teria impacto direto no tempo de treinamento e no processo de classificação dos modelos de aprendizado. Nessa perspectiva, a taxa de 2 500 amostras por segundo foi selecionada para atender a esses requisitos, além de ser suficiente para capturar uma frequência máxima de, aproximadamente, 1 250 Hz, que compreende as 20 primeiras componentes de frequência dos dispositivos, conforme o Teorema de *Nyquist* (Zamaruiev, 2017).

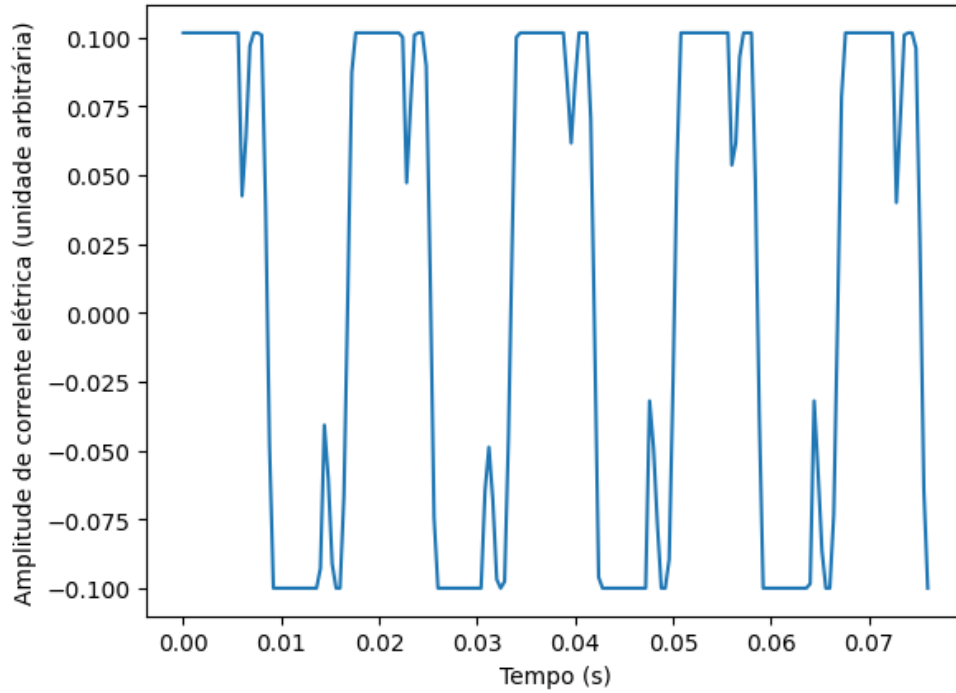
5.2 Análise dos dispositivos

Depois do processo de amostragem, os dados passaram por análise e processamento utilizando a linguagem *Python* dentro do ambiente virtual *Google Colaboratory*. Esse ambiente ofereceu 12 GB de memória *RAM* e 100 GB de armazenamento em disco, proporcionando os recursos necessários para as etapas de processamento. A Figura 18 mostra um exemplo da forma de onda da corrente do micro-ondas no domínio do tempo. É digno de nota ressaltar que o sensor *SCT-013-000* não fornece valores absolutos em ampères, mas sim valores proporcionais, e, portanto representada em uma unidade arbitrária qualquer. Por essa razão, foi decidido não converter os valores diretamente para amperagem real, a fim de evitar qualquer perda de acurácia devido a arredondamentos.

Como podemos observar, há diferenças significativas entre as formas de onda do sinal de corrente do micro-ondas e do sinal da geladeira, conforme mostrado na Figura 19, ambos no domínio do tempo. Os níveis de amplitude e o formato da onda exibem características distintas, as quais são geradas principalmente pela presença das frequências harmônicas. Portanto, é possível inferir que os sinais de corrente apresentam distinções discerníveis já no domínio do tempo, o que os torna suscetíveis a serem identificados por modelos de classificação baseados em aprendizado de máquina.

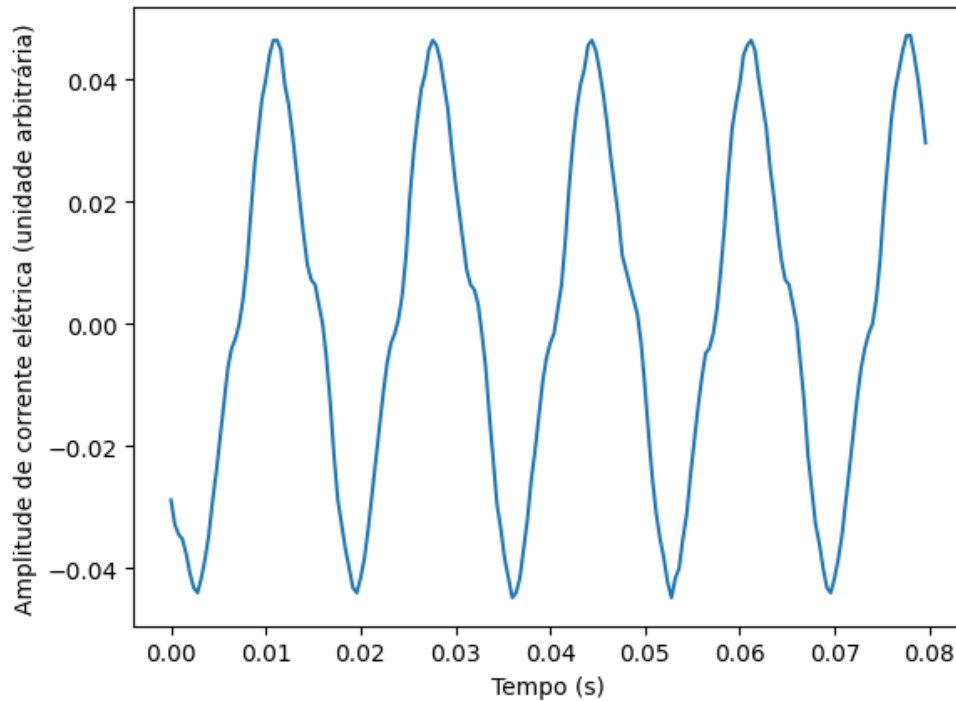
No entanto, é importante ressaltar que a representação dos sinais no domínio do tempo pode requerer um grande número de amostras, o que pode ser problemático, especialmente em ambientes com recursos limitados, como o da *IoT*. Este estudo propõe, alternativamente, a conversão desses sinais para o domínio da frequência, utilizando o cálculo da Transformada Rápida de *Fourier* (*FFT*). Dessa forma, um exemplo de um sinal com 100 amostras no domínio do tempo, abrangendo vários períodos do sinal, pode ser convertido em apenas 20 valores correspondentes às 20 primeiras componentes de frequência, resultando em uma redução de 5

Figura 18 – Sinal de corrente do micro-ondas no domínio do tempo construído a partir das amostras capturadas.



Fonte: Elaborada pelo próprio autor.

Figura 19 – Sinal de corrente da geladeira no domínio do tempo construído a partir das amostras capturadas.



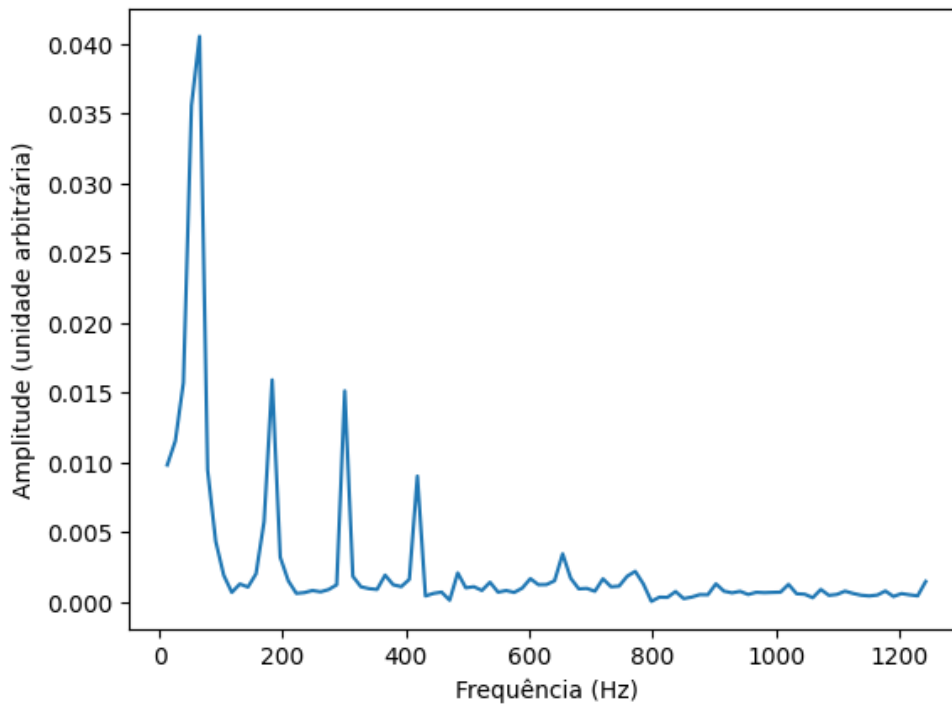
Fonte: Elaborada pelo próprio autor.

vezes na quantidade de dados.

A Figura 20 ilustra o sinal do micro-ondas convertido para o domínio da frequência. É

evidente que a frequência fundamental de 60 Hz se destaca com uma amplitude mais significativa. Além disso, é possível observar a presença de outras componentes de frequência com amplitudes menores, o que justifica a diferença no sinal no domínio do tempo em relação a uma senoide perfeita de 60 Hz.

Figura 20 – Sinal de corrente do micro-ondas no domínio da frequência construído utilizando o algoritmo *FFT*.

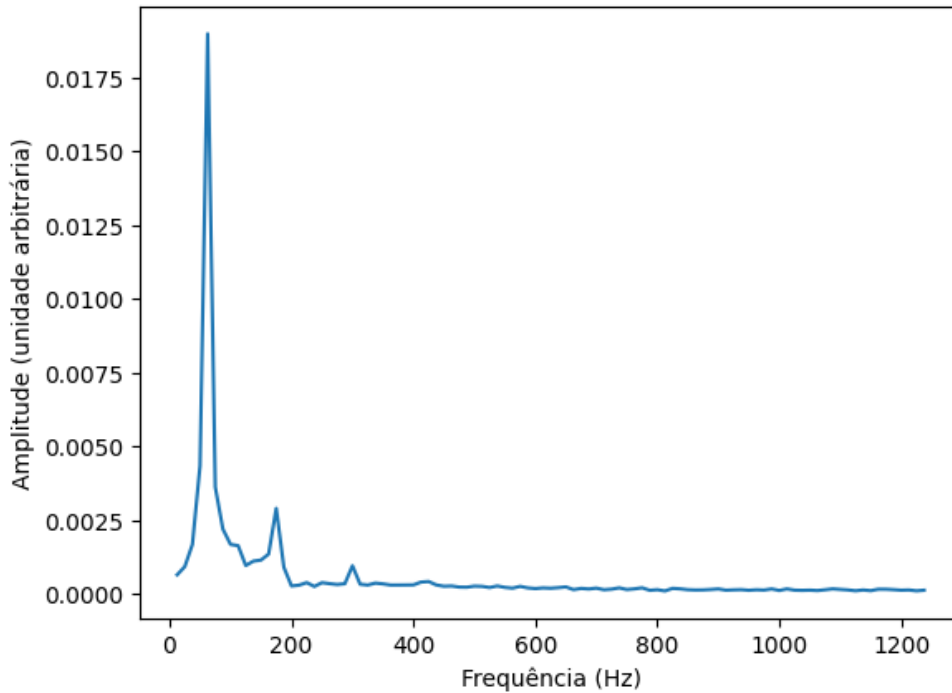


Fonte: Elaborada pelo próprio autor.

O espectro de frequência do micro-ondas ainda mostra características distintas quando comparado ao espectro da geladeira, como ilustrado na Figura 21. Conforme observado, o sinal da geladeira apresenta um número menor de frequências harmônicas, exibindo apenas três picos de menor magnitude, localizados relativamente próximos à frequência fundamental. Essa observação justifica-se pela forma de onda da geladeira se aproximar mais de uma senoide perfeita de 60 Hz no domínio do tempo.

Como visto, a presença de frequências harmônicas específicas e suas amplitudes são suficientes para identificar o dispositivo responsável pelo sinal elétrico de alimentação. Nesse contexto, a identificação dos dispositivos com base no sinal de corrente no domínio da frequência emerge como uma abordagem promissora, reduzindo a quantidade de dados necessária para treinamento e classificação. É importante ressaltar que os demais dispositivos avaliados nos experimentos foram submetidos a análises semelhantes, possibilitando também a diferenciação

Figura 21 – Sinal de corrente da geladeira no domínio da frequência construído utilizando o algoritmo *FFT*.



Fonte: Elaborada pelo próprio autor.

entre eles.

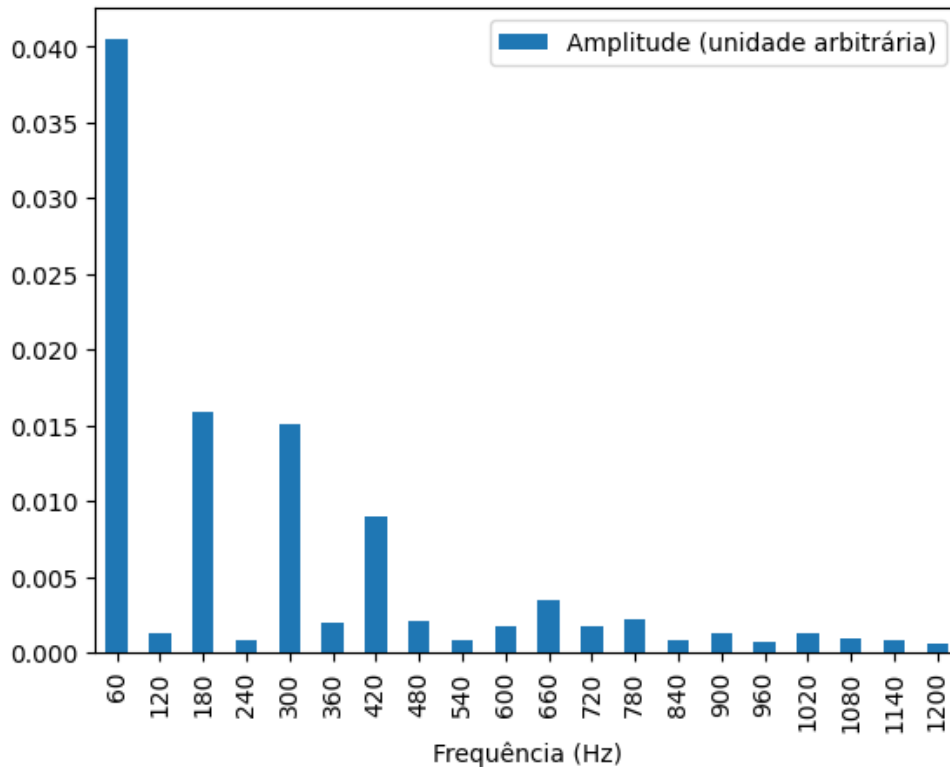
5.3 Construção do conjunto de dados

Depois de coletar e analisar as amostras, os valores medidos foram usados para formar o conjunto de dados. Para cada dispositivo, as 40 000 amostras foram divididas em blocos de 100 amostras consecutivas, criando 400 blocos no total. Em seguida, aplicou-se a *FFT* a cada bloco, ou seja, a cada conjunto de 100 amostras, transformando os dados para o domínio da frequência. Vale ressaltar que a *FFT* é mais eficiente para conjuntos de dados com tamanho múltiplo de potências de dois, mas o algoritmo pode lidar com outros tamanhos utilizando técnicas como o preenchimento com zeros, mantendo um desempenho superior ao da *FT* tradicional (Sheu *et al.*, 2024).

Depois, cada bloco foi discretizado, restando apenas as amplitudes das 20 primeiras componentes de frequência, que incluem a frequência fundamental e 19 harmônicas. A Figura 22 ilustra um bloco de amostras no domínio da frequência, em que é possível observar as amplitudes das componentes de frequência do sinal do micro-ondas em valores discretos.

Seguindo essa lógica, cada bloco se converteu em uma linha do conjunto de dados, onde as amplitudes das 20 primeiras componentes de frequência foram registradas nas colunas de

Figura 22 – Amplitudes discretizadas das 20 primeiras componentes de frequência do micro-ondas obtidas por meio da *FFT* das amostras capturadas.



Fonte: Elaborada pelo próprio autor.

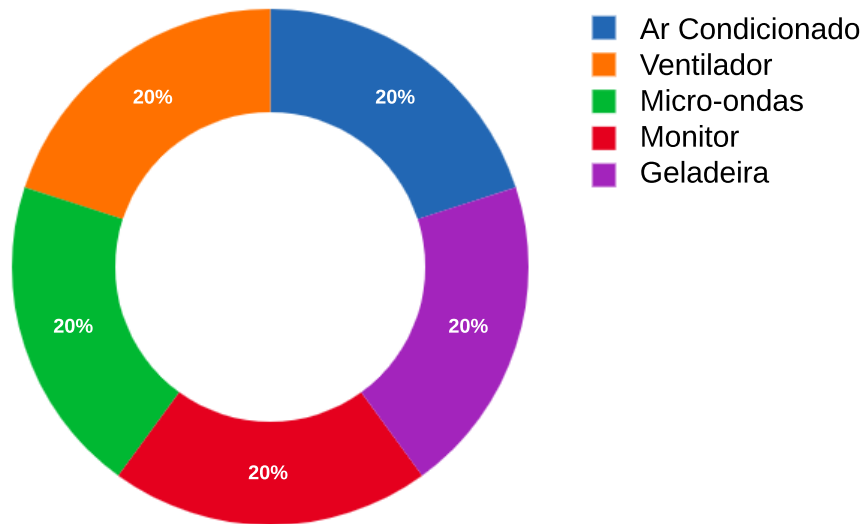
características (*features*), enquanto a coluna de rótulo recebeu o nome do dispositivo associado, resultando em um total de 21 colunas e 400 linhas para cada dispositivo elétrico. Desse modo, de um total de 40 000 amostras, resultou uma matriz com dimensões 400 x 21, o que corresponde a uma compactação de cerca de 5 vezes na informação em relação a representação temporal.

A Figura 23 exibe o percentual de distribuição de linhas relacionadas a cada dispositivo. Para esse estudo, foi criado um conjunto de dados único que integra informações de cinco dispositivos diferentes. Cada dispositivo contribuiu com 400 linhas, totalizando 2.000 linhas no conjunto de dados. Nota-se que o gráfico ilustra que cada dispositivo contribuiu com aproximadamente 20% do total de linhas, indicando um balanceamento adequado. Isso assegura que cada dispositivo seja treinado com o mesmo número de amostras, prevenindo viés no aprendizado do modelo em favor de algum dispositivo específico.

5.3.1 Dados no domínio do tempo

Para fins de comparação, também foi organizado um conjunto de dados no domínio do tempo. Nesse processo, as 40 000 amostras originais foram agrupadas em blocos de 40 amostras consecutivas, resultando em 1 000 blocos por dispositivo monitorado. Seguindo essa

Figura 23 – Percentual de distribuição de linhas associadas a cada dispositivo no conjunto de dados de frequência.



Fonte: Elaborada pelo próprio autor.

abordagem, cada bloco foi transformado em uma linha, onde as 40 amostras foram organizadas ao longo das colunas de características, enquanto a coluna de rótulos recebeu o nome do dispositivo associado. Portanto, temos um total de 41 colunas e 1 000 linhas associadas a cada dispositivo elétrico neste conjunto de dados no domínio do tempo.

A Figura 24 ilustra algumas linhas e colunas do conjunto de dados no domínio do tempo. Nota-se que as colunas de características são numeradas de t_1 a t_{40} , enquanto a coluna de rótulos é denominada *device*. Observa-se que o mesmo número de amostras resultou em um conjunto de dados significativamente maior, totalizando 5 000 linhas, aproximadamente 2,5 vezes mais do que o conjunto no domínio da frequência.

Além disso, o número de colunas é aproximadamente o dobro em comparação com o conjunto de dados de frequência. Em resumo, multiplicando o número de linhas pelo número de colunas, o conjunto de dados proposto no domínio da frequência é aproximadamente 5 vezes menor do que o conjunto com dados temporais. Essa disparidade se torna mais pronunciada à medida que a quantidade de dispositivos registrados no conjunto aumenta ou à medida que o número de componentes harmônicos em análise é menor.

5.4 Testes preliminares com os parâmetros fundamentais dos classificadores

Neste experimento preliminar, os parâmetros principais de cada classificador foram testados para determinar os intervalos de valores que serão considerados em testes subsequentes.

Figura 24 – Algumas amostras e características do conjunto de dados no domínio do tempo.

	t_1	t_2	t_3	...	t_39	t_40	device
0	-0.0072	-0.0064	-0.0088	...	-0.0064	-0.0096	fan
1	-8.0000	-8.0000	-8.0000	...	-0.0056	-0.0064	fan
2	-8.0000	-0.0072	-0.0088	...	-0.0032	-0.0048	fan
3	-0.0056	-0.0056	-8.0000	...	-0.0032	-0.0032	fan
4	-4.0000	-0.0056	-0.0064	...	0.0008	-0.0008	fan
...
4995	1016.0000	1016.0000	1016.0000	...	1016.0000	1016.0000	airConditioner
4996	1016.0000	1016.0000	1016.0000	...	-1.0000	-1.0000	airConditioner
4997	-1.0000	-1.0000	-1.0000	...	-1.0000	-1.0000	airConditioner
4998	-1.0000	-1.0000	-1.0000	...	1016.0000	1016.0000	airConditioner
4999	1016.0000	1016.0000	1016.0000	...	1016.0000	1016.0000	airConditioner

5000 rows × 41 columns

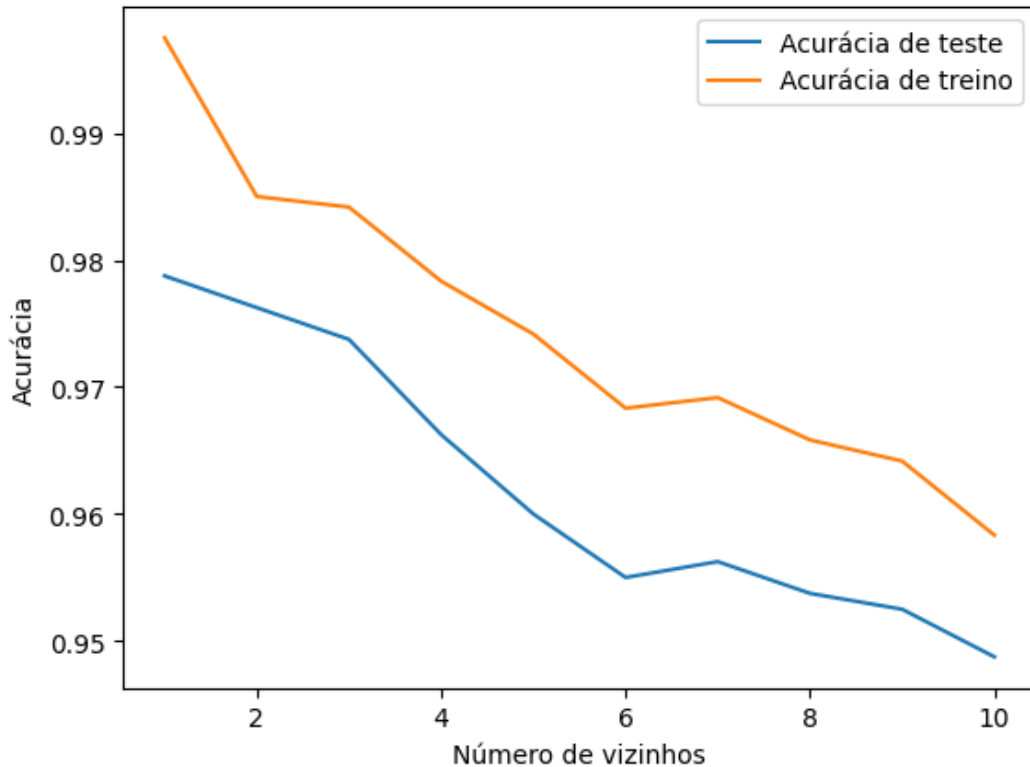
Fonte: Elaborada pelo próprio autor.

Inicialmente, o conjunto de dados de frequência foi normalizado e dividido em 60% para treinamento e 40% para teste dos modelos. Neste contexto, a normalização refere-se ao processo de ajustar os valores dos dados para que eles se situem em um intervalo específico, geralmente entre 0 e 1. Isso é feito para garantir que todas as colunas do conjunto de dados tenham a mesma escala, o que pode melhorar o desempenho dos modelos de aprendizado de máquina.

Em seguida, cada classificador foi avaliado com base em seu parâmetro fundamental específico. Por exemplo, no caso do *kNN*, variou-se o parâmetro *k*, que indica o número de vizinhos mais próximos considerados na classificação. Para cada valor de *k*, o modelo foi treinado e testado, registrando-se a acurácia resultante. A Figura 25 apresenta os resultados de acurácia do *kNN* com *k* variando de 1 a 10.

Conforme observado, o modelo foi avaliado tanto com o conjunto de treinamento quanto com o conjunto de teste. No entanto, nota-se que em ambos os conjuntos, a acurácia diminui à medida que o parâmetro *k* aumenta. Esse experimento inicial sugere que, para esse contexto de classificação, resultados de acurácia melhores são alcançados quando o número de vizinhos considerados pelo classificador é menor. Essas conclusões são levadas em conta nos experimentos subsequentes, indicando que é suficiente testar valores de *k* próximos de 1 no caso

Figura 25 – Avaliação do número de vizinhos do kNN utilizando o conjunto de dados no domínio da frequência.



Fonte: Elaborada pelo próprio autor.

do kNN .

Essa prática foi repetida com outros classificadores, variando as funções do *kernel* do *SVM*, o número de árvores de decisão para os modelos *Random Forest* e *Gradient Boosting*, e o número de neurônios e camadas ocultas no classificador *MLP*, com o objetivo de determinar o intervalo de valores mais adequado para esses parâmetros. Essa abordagem sistemática permite a seleção dos valores ótimos que resultam em um desempenho máximo do modelo para o problema específico em questão. Este procedimento foi executado novamente utilizando o conjunto de dados no domínio do tempo, com o intuito de identificar os valores ideais para fins de comparação em experimentos subsequentes.

5.5 Resultados de identificação dos dispositivos

Para este experimento, os dados foram separados aleatoriamente em conjuntos de treinamento e de teste, na proporção de 60% para treinamento e 40% para teste. Em seguida, foi utilizada a biblioteca *Scikit-Learn* para realizar uma análise detalhada no conjunto de treinamento, explorando diversas combinações de hiperparâmetros para cada classificador. Utilizou-se a

técnica de validação cruzada, dividindo o conjunto em S subconjuntos.

Cada combinação de hiperparâmetros é testada S vezes, onde a cada iteração, um dos subconjuntos é utilizado como conjunto de teste e os demais formam o conjunto de treinamento. O resultado final para uma combinação de hiperparâmetros específica é a média dos resultados das S iterações. O parâmetro que determina o número de subconjuntos é chamado de *k-fold*, e neste caso, adotamos um valor de 10. Além disso, optou-se por avaliar os resultados desses experimentos utilizando a métrica F1-score.

Ao término desse processo, foi possível identificar uma combinação otimizada de hiperparâmetros para cada classificador. O Quadro 3 apresenta os classificadores selecionados e seus hiperparâmetros escolhidos para os dados de frequência, juntamente com seus respectivos F1-scores de treinamento. É importante ressaltar que os hiperparâmetros não exibidos mantêm os valores padrão da biblioteca *Scikit Learn*. Conforme observado, os valores de F1-score estão todos acima de 90%, indicando uma perspectiva positiva. No entanto, os classificadores precisam ser avaliados com dados não vistos para uma validação mais completa.

Quadro 3 – Classificadores selecionados e seus F1-scores de treinamento após ajuste de hiperparâmetros e validação cruzada com *k-fold* igual a 10 utilizando o conjunto de dados de frequência.

Classificador	F1-score
KNeighborsClassifier(metric=manhattan, n_jobs=-1, n_neighbors=3, p=1)	0,981
SVC(C=10, gamma=1, random_state=42)	0,936
DecisionTreeClassifier(random_state=42)	0,925
RandomForestClassifier(max_depth=10, min_samples_leaf=2, min_samples_split=5, n_estimators=50, random_state=42)	0,933
GradientBoostingClassifier(learning_rate=0.2, max_depth=4, random_state=42)	0,940
MLPClassifier(hidden_layer_sizes=(400, 400, 400, 400, 400), random_state=42)	0,935

Fonte: Elaborado pelo próprio autor.

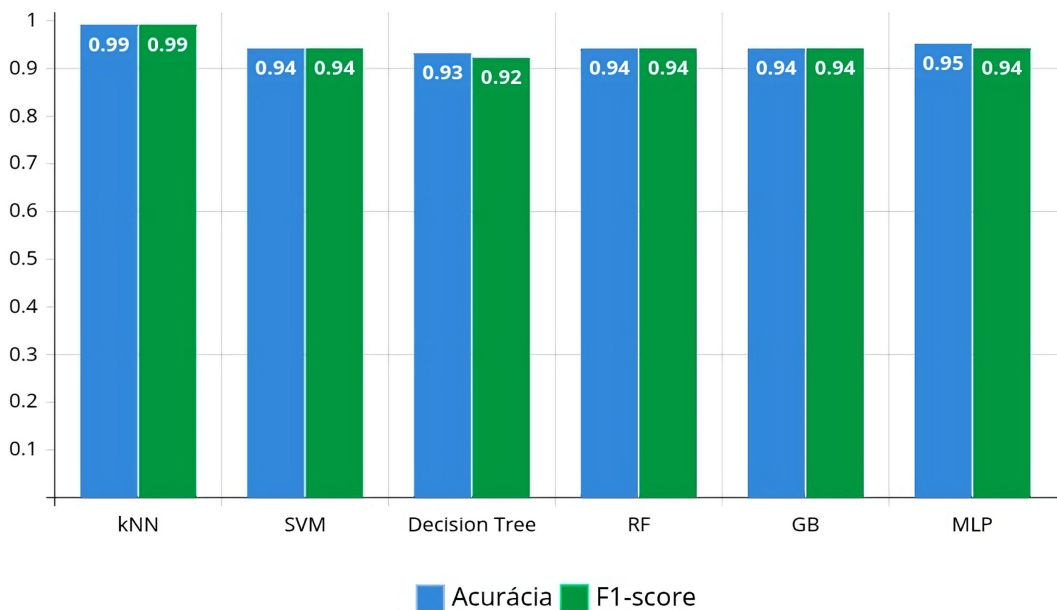
É crucial destacar que, nesse ponto, foram utilizados apenas 60% dos dados, referentes ao conjunto de treinamento, que foram obtidos na primeira divisão. Em outras palavras, durante o treinamento, nenhum modelo teve acesso aos 40% restantes dos dados, referentes ao conjunto de teste. Essa abordagem é adotada para evitar resultados tendenciosos causados por sobreajuste (*overfitting*). Essa sequência de passos foi seguida nas demais avaliações também.

Na sequência, cada classificador otimizado foi avaliado usando os restantes 40% dos dados, referentes ao conjunto de teste. Durante essa avaliação, foram aplicadas as métricas de acurácia e F1-score. Esses procedimentos foram conduzidos inicialmente no conjunto de dados principal, que opera no domínio da frequência. Posteriormente, essas etapas foram repetidas

no conjunto de dados que utiliza informações temporais, visando realizar uma comparação dos resultados.

A Figura 26 apresenta os resultados de acurácia e F1-score dos classificadores no contexto da análise de frequência. É notável que todos os classificadores superaram os 90% em ambas as métricas. Destaca-se, em particular, o *kNN*, que alcançou uma acurácia e F1-score impressionantes de 99%. Em seguida, o *MLP* registrou 95% de acurácia e 94% de F1-score. Para avaliar a eficácia do processo de otimização de hiperparâmetros, é relevante mencionar que o classificador que utiliza apenas um neurônio *Perceptron* da biblioteca *Scikit Learn* obteve apenas 69% de acurácia e 68% de F1-score em experimentos adicionais. Ademais, os classificadores *SVM*, *Random Forest (RF)* e *Gradient Boosting (GB)* atingiram uma marca de 94%, tanto em acurácia quanto em F1-score. Por outro lado, o classificador *Árvore de Decisão (Decision Tree)* obteve os resultados mais baixos, com 93% de acurácia e 92% de F1-score.

Figura 26 – Resultados de acurácia e F1-score dos classificadores utilizando o conjunto de dados no domínio da frequência.



Fonte: Elaborada pelo próprio autor.

Os resultados foram encorajadores, corroborando a viabilidade da classificação do sinal de corrente no domínio da frequência. Um projeto *IoT* enfrenta diversos desafios relacionados à escassez de recursos de processamento e memória; a quantidade de dados a serem processados ou lidos de um sensor pode resultar em gargalos na comunicação, alta latência e descarte de pacotes. Nesse contexto, a abordagem de classificação no domínio da frequência se destaca, pois converte várias amostras do sinal no domínio do tempo em 20 frequências. Isso

significa que, mesmo que o dispositivo *IoT* tenha que realizar o cálculo da *FFT*, os pacotes transmitidos na rede serão menores, o que, por sua vez, aprimora significativamente as métricas de comunicação. Outrossim, a utilização de conjuntos de dados no treinamento do classificador é substancialmente reduzida quando os dados estão no domínio da frequência.

Ademais, dispositivos de borda podem desempenhar um papel crucial na aplicação do cálculo da *FFT* e na implementação do modelo de classificação. Assim, embora o escopo desta dissertação não abranja o modelo de comunicação do projeto de *IoT*, essas soluções são relevantes para viabilizar a implementação deste estudo. Em resumo, lidar com uma quantidade reduzida de dados oferece vantagens em diversos aspectos de um sistema *IoT*.

5.5.1 Comparação com domínio do tempo

Semelhantemente, os passos foram repetidos com os dados temporais visando a comparação dos resultados. O Quadro 4 apresenta os classificadores selecionados e seus hiperparâmetros escolhidos para os dados no domínio do tempo, juntamente com seus respectivos F1-scores de treinamento, após o processo de ajuste dos hiperparâmetros. Vale salientar que os hiperparâmetros não exibidos mantêm os valores padrão da biblioteca *Scikit Learn*.

Quadro 4 – Classificadores selecionados e seus F1-scores de treinamento após ajuste de hiperparâmetros e validação cruzada com *k-fold* igual a 10 utilizando o conjunto de dados temporais.

Classificador	F1-score
KNeighborsClassifier(metric=manhattan, n_jobs=-1, n_neighbors=3, p=1, weights=distance)	0,803
SVC(C=10, gamma=1, random_state=42)	0,663
DecisionTreeClassifier(random_state=42)	0,956
RandomForestClassifier(n_estimators=200, random_state=42)	0,995
GradientBoostingClassifier(max_depth=4, n_estimators=200, random_state=42)	0,995
MLPClassifier(hidden_layer_sizes=(400, 400, 400, 400, 400), random_state=42)	0,809

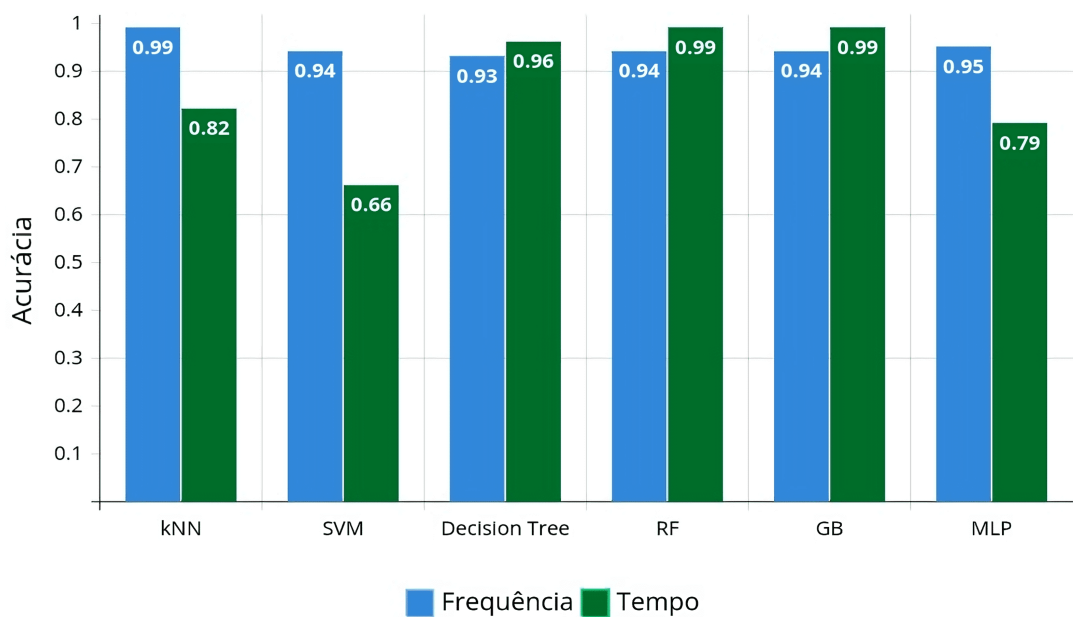
Fonte: Elaborado pelo próprio autor.

Conforme observado, alguns classificadores não apresentaram resultados tão satisfatórios quando comparados aos obtidos no domínio da frequência. Por exemplo, o *SVM* obteve aproximadamente 66%, o *kNN* cerca de 80% e o *MLP* também em torno de 80%. No entanto, os classificadores *Árvore de Decisão*, *Random Forest* e *Gradient Boosting* alcançaram valores superiores a 95%, chegando até surpreendentes 99%. No entanto, é necessário submetê-los ao conjunto de teste para analisar os resultados em comparação com os obtidos com dados de

frequência. Assim, após o treinamento realizado com o conjunto correspondente a 60% dos dados, os modelos treinados foram avaliados com os 40% restantes dos dados, referentes ao conjunto de teste.

Como relatado, o conjunto de dados no domínio da frequência condensou 40 000 amostras em 21 colunas e 400 linhas para cada dispositivo rotulado. Em contraste, o conjunto de dados no domínio do tempo reorganizou essas mesmas amostras em 41 colunas e 1 000 linhas para cada dispositivo. No entanto, apesar do maior volume de dados, os resultados no domínio do tempo não se destacaram em comparação com os resultados no domínio da frequência, conforme ilustrado nas Figuras 27 e 28 que apresentam uma análise comparativa com os resultados de acurácia e F1-score entre os domínios da frequência e do tempo.

Figura 27 – Resultados de acurácia dos classificadores ao empregar os conjuntos de dados nos domínios da frequência e do tempo.

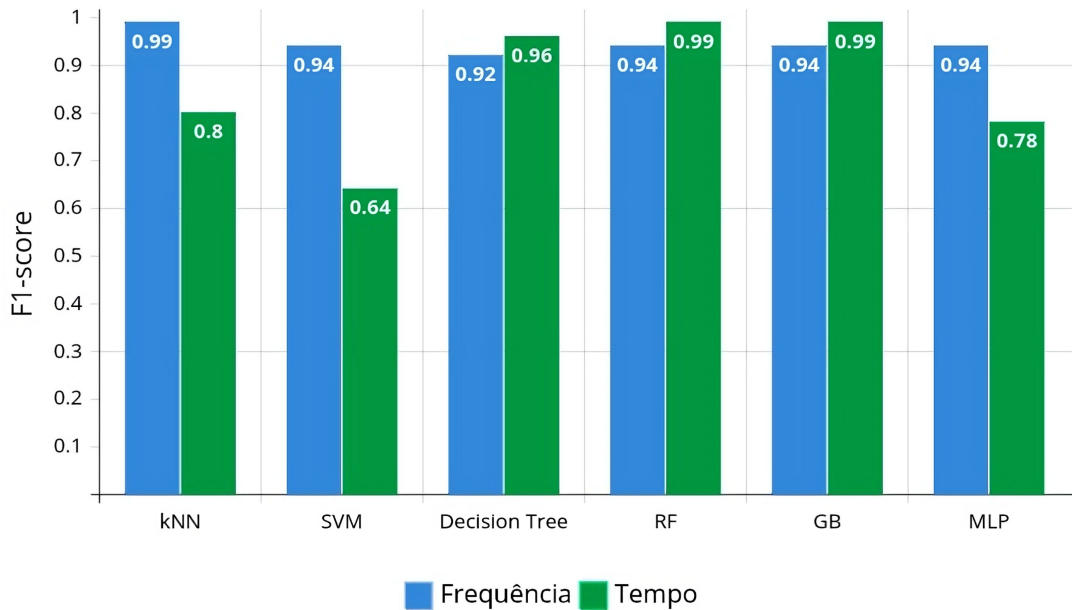


Fonte: Elaborada pelo próprio autor.

Conforme destacado na Figura 27, os resultados de acurácia para os classificadores *kNN*, *SVM* e *MLP* revelaram uma superioridade no domínio da frequência, com diferenças notáveis de 17, 28 e 16 pontos percentuais (pp), respectivamente. No entanto, para os demais classificadores, observou-se acurácias mais altas no domínio do tempo. Essas discrepâncias, entretanto, foram mais modestas, totalizando apenas 5 pontos percentuais nos classificadores *Random Forest* e *Gradient Boosting*, e 3 pontos percentuais no classificador *Árvore de Decisão*.

Conforme a Figura 28, os resultados do F1-score para os classificadores *kNN*, *SVM* e *MLP* também foram superiores no domínio da frequência, apresentando diferenças significativas

Figura 28 – Pontuações F1-score dos classificadores ao empregar os conjuntos de dados nos domínios da frequência e do tempo.



Fonte: Elaborada pelo próprio autor.

de 19, 30 e 16 pp, respectivamente. Em contrapartida, para os demais classificadores, as pontuações mais elevadas do F1-score foram observadas no domínio do tempo. Contudo, as discrepâncias permaneceram modestas, sendo de apenas 5 pp nos classificadores *Random Forest* e *Gradient Boosting*, e de 4 pp no classificador *Árvore de Decisão*.

Em síntese, ao analisar os resultados nos domínios da frequência e do tempo, observa-se que em metade dos classificadores os valores de acurácia e F1-score foram significativamente superiores no domínio da frequência em comparação com o tempo. Esses resultados destacam a vantagem do uso do domínio da frequência, uma vez que, com uma quantidade menor de dados, foi possível alcançar resultados superiores em termos de desempenho preditivo. Nos casos em que os algoritmos apresentaram resultados superiores no domínio do tempo, a diferença em relação ao domínio da frequência foi relativamente pequena.

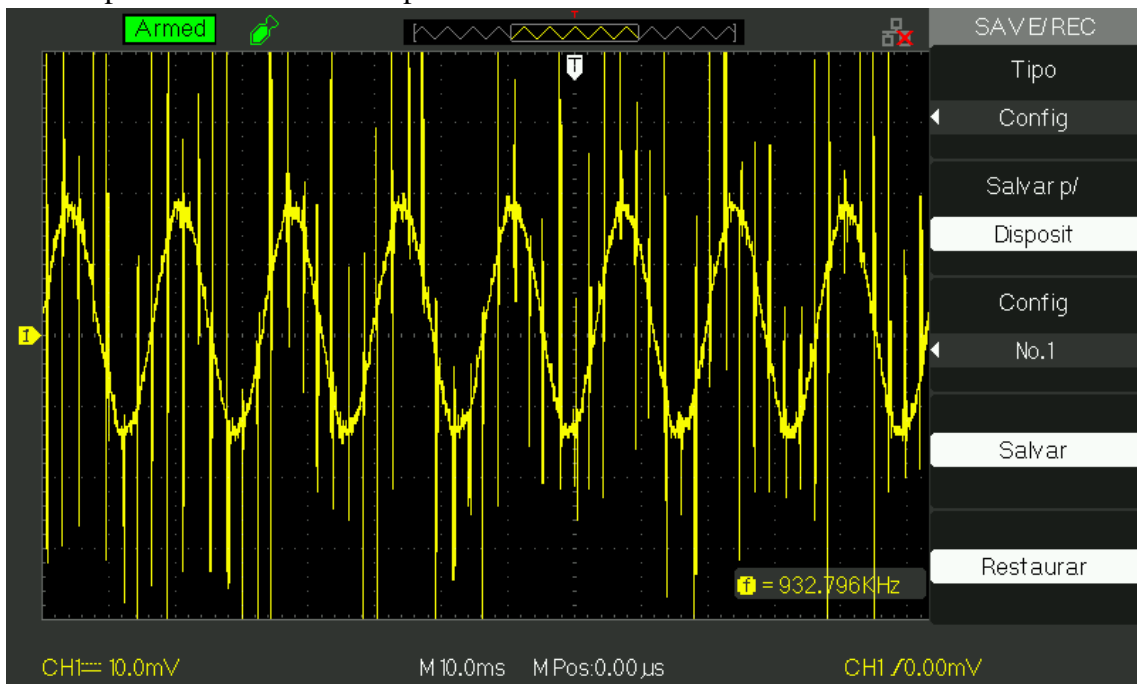
5.5.2 Impacto da presença de dispositivos simultâneos na rede: Avaliação dos classificadores em ambientes com ruído

O ruído na rede elétrica pode ser atribuído a uma variedade de fontes, como motores, transformadores e lâmpadas fluorescentes, que geram ondas eletromagnéticas durante o seu funcionamento (Wang *et al.*, 2017). Por exemplo, motores elétricos produzem campos magnéticos em torno de si, enquanto transformadores podem introduzir ruídos devido às variações

na corrente elétrica. Além disso, lâmpadas fluorescentes, que utilizam descargas elétricas para produzir luz, podem gerar ruído devido às suas características de funcionamento. Esse ruído pode se propagar pela rede elétrica e afetar dispositivos conectados a ela, causando distorções nos sinais elétricos (Michałowska *et al.*, 2019).

Dispositivos com baixa potência são especialmente suscetíveis a esses ruídos devido à sua menor capacidade de lidar com flutuações na rede elétrica. Por exemplo, uma lâmpada incandescente pode apresentar oscilações na intensidade da luz ou até mesmo falhas de funcionamento quando exposta a ruídos. A Figura 29 mostra uma visualização através de osciloscópio do sinal de corrente de uma lâmpada incandescente operando em conjunto com outros dispositivos. Devido à natureza resistiva da lâmpada incandescente, era esperado um sinal próximo a uma forma de onda senoidal. No entanto, o sinal exibido é distorcido e repleto de ruídos, devido à influência da operação simultânea com outros dispositivos na rede elétrica.

Figura 29 – Sinal de corrente distorcido de uma lâmpada incandescente operando com outros dispositivos.



Fonte: Elaborada pelo próprio autor.

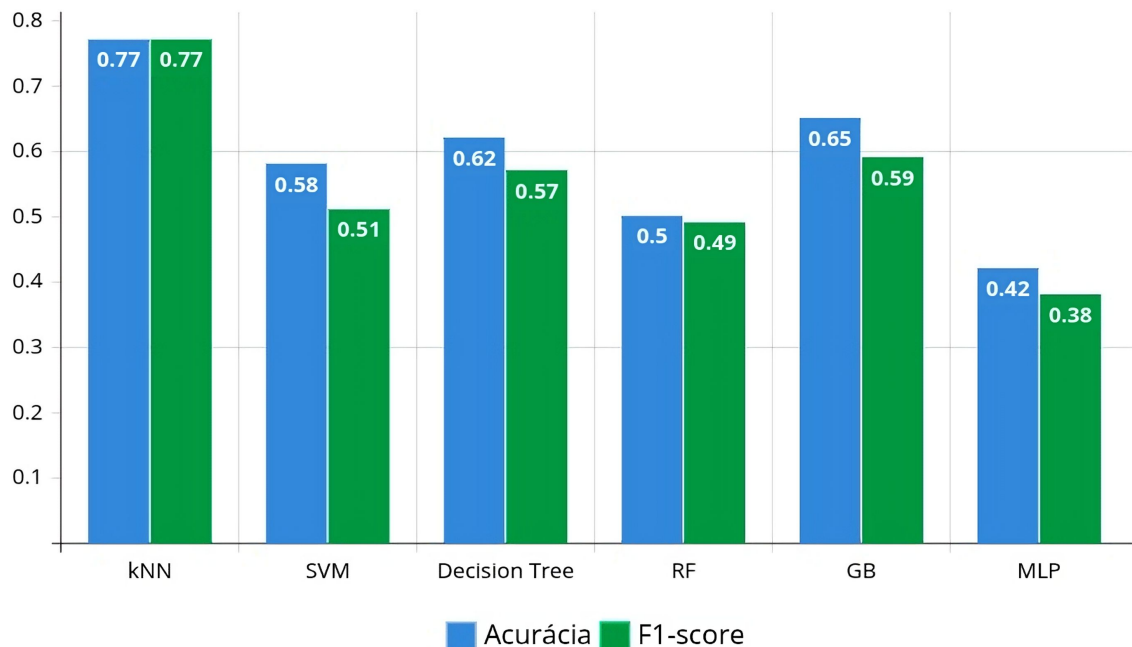
Neste novo experimento, foi montado um ambiente mais realista ao introduzir a presença simultânea de outros dispositivos na rede elétrica, o que resultou em dados de entrada mais ruidosos para os modelos de classificação. Destaca-se que este estudo contempla um ambiente de *IoT* no qual múltiplas tomadas podem ser monitoradas simultaneamente. Inicialmente, os modelos foram treinados com um conjunto de dados isolados, que não continham ruído. Esses

dados foram coletados dos dispositivos, como aparelho de ar-condicionado, monitor, lâmpada incandescente e micro-ondas.

Após o treinamento inicial dos modelos, foi realizado o teste dos modelos utilizando um novo conjunto de dados. Esse conjunto soma os sinais de corrente dos dispositivos mencionados anteriormente, os quais estavam operando simultaneamente com outros dispositivos na rede elétrica. Esse cenário resultou em um ambiente caracterizado pela presença significativa de ruído nos dados de entrada dos modelos.

Para este experimento, foram selecionados novamente os classificadores *kNN*, *SVM*, *Árvore de Decisão*, *Random Forest*, *Gradient Boosting* e *MLP*. Cada classificador passou por testes de validação cruzada com *k-fold* igual a 10 para otimizar seus hiperparâmetros usando o conjunto de dados sem ruído. Após determinar a configuração ideal dos hiperparâmetros, os modelos foram treinados com esse conjunto de dados. Posteriormente, cada modelo foi avaliado em termos de acurácia e F1-score utilizando entradas do conjunto de dados com ruído. Os resultados obtidos estão representados no gráfico da Figura 30.

Figura 30 – Desempenho dos modelos na classificação de dados de dispositivos operando simultaneamente com outros em ambientes ruidosos.



Fonte: Elaborada pelo próprio autor.

Foram alcançados resultados bastante reveladores. O classificador *kNN* alcançou 77% de acurácia e F1-score, números significativamente inferiores em comparação com os resultados anteriores, sugerindo que a presença de ruído impactou negativamente seu desempenho. O *SVM*

também apresentou uma queda substancial na acurácia, atingindo apenas 58%, e um F1-score ainda menor de 51%, indicando uma maior dificuldade em lidar com a presença de ruído nos dados de entrada.

Os classificadores *Árvore de Decisão*, *Random Forest* e *Gradient Boosting*, também demonstraram uma redução em relação aos resultados anteriores. Especificamente, o *Random Forest* apresentou uma acurácia de apenas 50%, indicando uma grande sensibilidade ao ruído presente nos dados. O *MLP*, que obteve os piores resultados nesta configuração, registrou apenas 42% de acurácia e 38% de F1-score, sugerindo uma significativa dificuldade em generalizar os padrões presentes nos dados com a presença de ruído.

5.5.2.1 *Empregando filtros para diminuir os ruídos*

Uma lâmpada de LED, posicionada de forma estratégica, foi a maior responsável pela geração de ruído. A lâmpada de LED foi deliberadamente ligada simultaneamente aos dispositivos a serem classificados com o intuito de criar uma fonte de ruído na rede elétrica. O motivo dessa escolha reside no fato de que o reator presente na lâmpada de LED é conhecido por gerar ruídos, afetando outros dispositivos conectados à rede elétrica. Essa ação foi tomada como parte do experimento para avaliar a sensibilidade dos dispositivos em relação a ruídos.

É importante ressaltar que, além da lâmpada de LED, outros dispositivos elétricos adicionais podem gerar ruído, tanto branco quanto colorido, afetando os resultados do experimento. O ruído colorido é caracterizado por uma densidade espectral de potência desigual, com algumas frequências mais intensas do que outras, em contraste com o ruído branco, que tem uma distribuição uniforme de densidade de potência ao longo de todo o espectro. Dispositivos como eletrodomésticos e motores elétricos podem introduzir esses ruídos em frequências específicas, influenciando a análise dos dispositivos testados e a precisão dos resultados.

Deste modo, os ruídos provenientes de dispositivos adicionais podem impactar significativamente o desempenho dos modelos, tornando crucial a adoção de estratégias eficientes de pré-processamento de dados. A presença de ruídos pode ser representada matematicamente como a adição de sinais indesejados ao sinal desejado, conforme ilustrado na Equação 5.1. Nessa equação, B_i são as amplitudes dos ruídos em diversas frequências, f_i são as frequências dos ruídos, que podem incluir harmônicas e outras frequências, e ψ_i são as fases dos ruídos.

$$i_{\text{ruído}}(t) = \sum_{k=1}^N B_k \cdot \sin(2\pi \cdot f_k \cdot t + \psi_k) \quad (5.1)$$

O sinal total observado é a combinação do sinal de corrente com as 20 primeiras componentes de frequência e os ruídos, conforme ilustram as Equações 5.2 e 5.3. A soma de sinais é viável graças ao princípio da sobreposição, que afirma que os efeitos de múltiplas fontes de sinal podem ser combinados somando-se suas contribuições individuais (Barbi, 2021). Nesse contexto, pode-se interpretar duas fontes distintas: o sinal desejado, composto por uma componente fundamental e suas harmônicas, e o ruído, representado por sinais indesejados de diversas frequências e amplitudes. Ao aplicar a sobreposição, somaram-se os efeitos de ambas as fontes para obter o sinal total observado.

$$i_{\text{total}}(t) = i_A(t) + i_{\text{ruído}}(t) \quad (5.2)$$

$$i_A(t) = A_0 \cdot \sin(2\pi \cdot 60 \cdot t) + \sum_{k=2}^{20} A_k \cdot \sin(2\pi \cdot k \cdot 60 \cdot t + \phi_k) \quad (5.3)$$

Uma alternativa para lidar com o problema de ruídos nas frequências desejadas é utilizar filtros passa-baixa. Esses filtros são projetados para permitir a passagem de frequências até um ponto de corte específico, atenuando frequências mais altas. Para capturar as 20 primeiras componentes de frequência, pode-se utilizar um filtro passa-baixa com uma frequência de corte de 1200 Hz (correspondente à 20ª harmônica). A função de transferência do filtro é representada como $H_{\text{passa-baixa}}(f)$. A aplicação do filtro ao sinal total é ilustrada na Equação 5.4.

$$i_{\text{filtrado}}(t) = H_{\text{passa-baixa}}(f) \cdot i_{\text{total}}(t) \quad (5.4)$$

Esse filtro passa-baixa permite que as 20 primeiras componentes de frequência passem quase sem alterações, ao mesmo tempo em que atenua as frequências mais altas, reduzindo os ruídos nos sinais. Ao empregar o filtro passa-baixa dessa forma, é possível isolar as 20 primeiras componentes de frequência do sinal desejado e minimizar o impacto dos ruídos de frequências mais altas, resultando em um sinal mais limpo e preciso.

No entanto, ruídos na rede elétrica podem ser provocados não apenas em frequências altas, mas também nas 20 primeiras componentes de frequência do sinal de corrente. Isso ocorre

porque as fontes de ruído podem gerar sinais que coincidem com essas frequências. A presença de ruídos nas 20 primeiras componentes de frequência pode ser explicada novamente como a presença de sinais indesejados somados ao sinal desejado.

Por exemplo, se um dispositivo conectado à rede elétrica gerar ruído nas frequências das harmônicas do sinal desejado, essas frequências indesejadas podem ser expressas conforme a Equação 5.5. Nessa equação, B_k representam as amplitudes dos ruídos em cada uma das 20 primeiras componentes de frequência, enquanto ψ_k são as fases desses ruídos nas mesmas frequências.

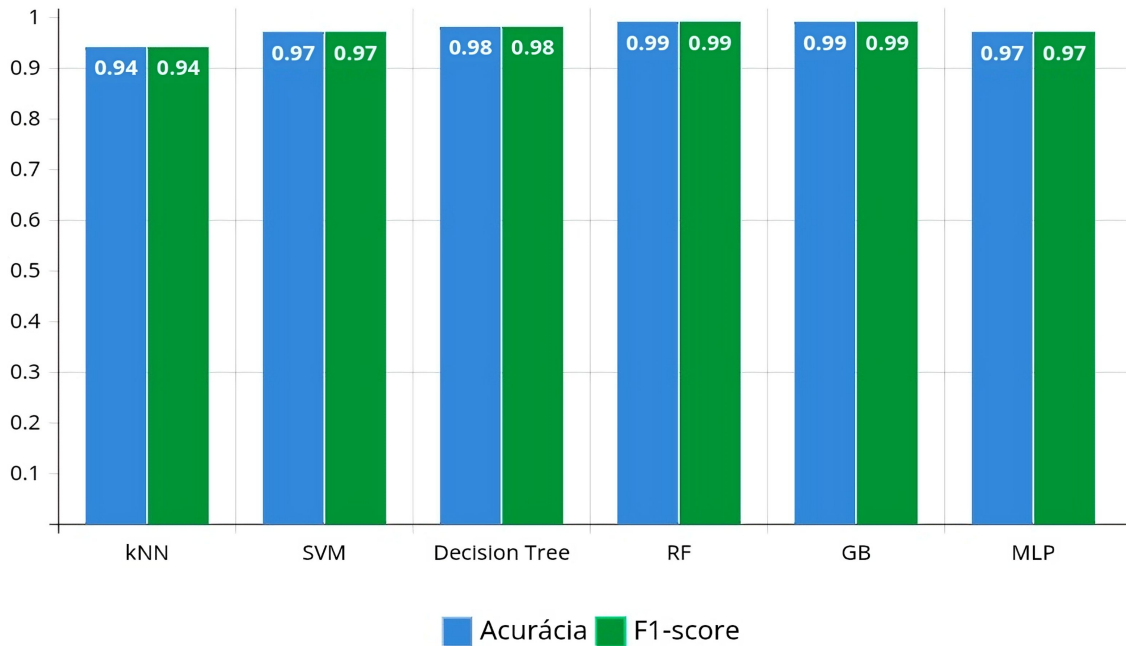
$$i_{\text{ruído}}^k(t) = B_k \cdot \sin(2\pi \cdot k \cdot 60 \cdot t + \psi_k) \quad (5.5)$$

Esses ruídos podem se somar ao sinal desejado, influenciando a amplitude das harmônicas. Contudo, embora os ruídos nas 20 primeiras componentes de frequência sejam uma preocupação, atenuar as frequências mais altas com um filtro passa-baixa ainda é uma grande ajuda. Isso ocorre porque a presença de frequências altas pode afetar a clareza do sinal total, contribuindo para a distorção do sinal e tornando-o mais difícil de analisar.

A aplicação de um filtro passa-baixa com uma frequência de corte em 1200 Hz pode eliminar as frequências acima da 20ª harmônica, resultando em um sinal total mais limpo e próximo ao sinal desejado. Mesmo havendo ruído residual nas 20 primeiras componentes de frequência, a remoção das frequências mais altas facilita a detecção e a análise das harmônicas desejadas. Portanto, o uso de um filtro passa-baixa para atenuar as frequências acima de 1200 Hz contribui significativamente para melhorar a qualidade do sinal total.

Diante disso, optou-se por empregar um filtro passa-baixa de primeira ordem, integrado ao próprio osciloscópio, durante os testes. Considerando a precisão da frequência de corte de 5%, foi utilizada uma frequência de corte de 1300 Hz. Após experimentação, foi identificado que esse filtro mostrou-se eficaz na atenuação dos ruídos, ao mesmo tempo em que preservava as 20 primeiras componentes de frequência dos sinais de entrada. Nesse contexto, para avaliar o desempenho dos modelos diante da presença simultânea de dispositivos na rede elétrica, foi criado um novo conjunto de dados de teste. No entanto, diferentemente dos experimentos com dados ruidosos, utilizou-se o filtro passa baixa embutido no osciloscópio para processar os dados. A Figura 31 ilustra o desempenho dos modelos na classificação utilizando dados de entrada filtrados.

Figura 31 – Desempenho dos modelos na classificação com dados filtrados de dispositivos operando simultaneamente com outros em ambientes ruidosos.



Fonte: Elaborada pelo próprio autor.

Por exemplo, o classificador *kNN* registrou uma acurácia e um F1-score de 94%, evidenciando uma melhoria substancial em relação aos resultados sem filtro. O mesmo padrão foi observado para o *SVM*, *Árvore de Decisão* e *MLP*, com todos esses modelos atingindo uma acurácia e F1-score de 97% ou superior. Destaca-se ainda que os modelos *Random Forest* e *Gradient Boosting* alcançaram uma acurácia e F1-score de 99%, demonstrando um desempenho excepcional após a aplicação do filtro.

Além disso, é notável que os resultados obtidos com dados filtrados estão muito próximos dos resultados anteriores com dados isolados, indicando que o filtro passa baixa foi eficaz em preservar as características importantes dos dados enquanto eliminava o ruído indesejado. Isso sugere que a utilização de técnicas de pré-processamento adequadas pode garantir a robustez e a confiabilidade dos modelos de classificação, mesmo em ambientes ruidosos e complexos.

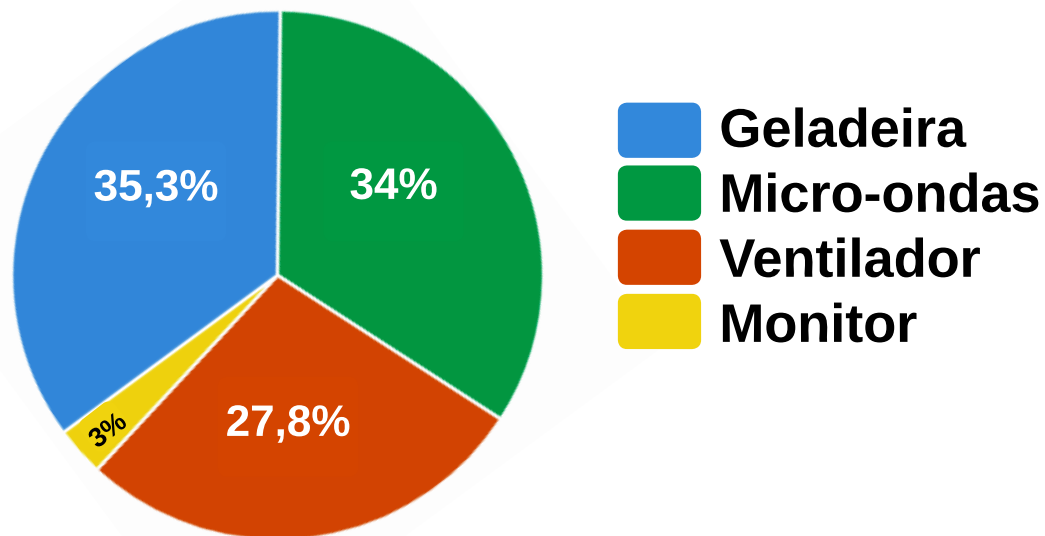
5.5.3 Classificação de dispositivos desconhecidos

Uma questão intrigante diz respeito ao comportamento dos modelos ao receberem informações de um dispositivo desconhecido, ou seja, sem um treinamento prévio do referido dispositivo. Nesse contexto, foi conduzido um experimento com o objetivo de mapear o desempenho. Para isso, foram excluídas todas as linhas relacionadas ao dispositivo de ar condicionado

do conjunto de dados de frequência. Em seguida, foi realizado o treinamento do modelo apenas com os dispositivos que restaram, a saber: geladeira, micro-ondas, ventilador e monitor. Ou seja, a partir de agora os modelos desconhecem o ar condicionado.

O algoritmo selecionado para este experimento foi o *kNN*, devido aos bons resultados obtidos nos testes anteriores. Assim, o classificador foi mais uma vez submetido a testes de validação cruzada, a fim de otimizar seus hiperparâmetros com o conjunto de dados modificado. Após definir a configuração ideal dos hiperparâmetros, foi realizado o treinamento do modelo. Logo em seguida, o modelo foi empregado para classificar cada uma das 400 linhas contendo as componentes de frequência do dispositivo de ar condicionado. Os resultados são apresentados no gráfico ilustrado na Figura 32.

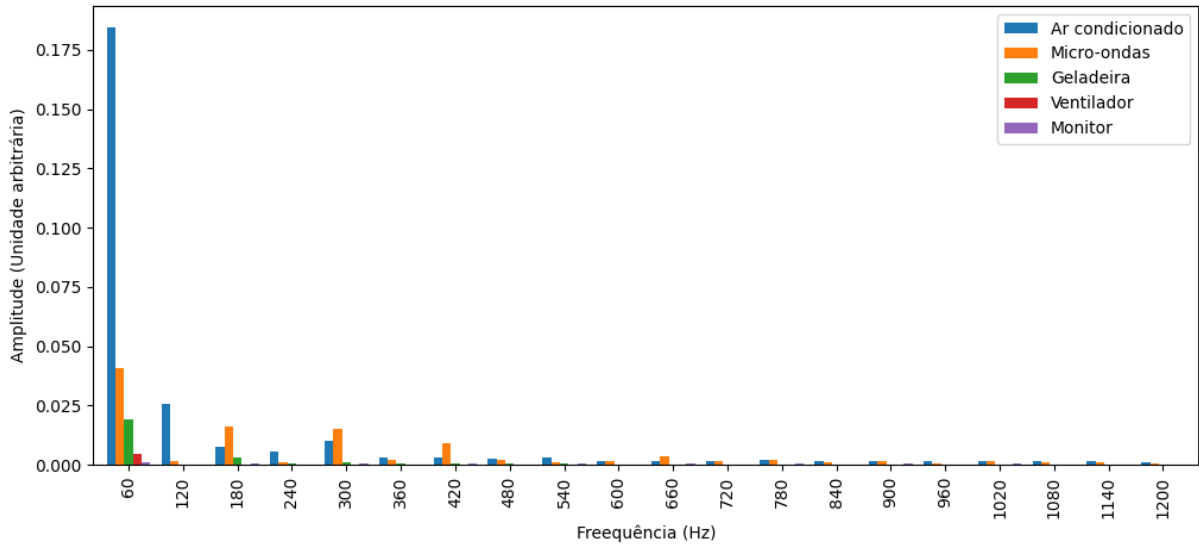
Figura 32 – Desempenho do modelo *kNN* ao tentar classificar o dispositivo de ar condicionado sem possuir conhecimento sobre o aparelho durante o treinamento.



Fonte: Elaborada pelo próprio autor.

Conforme observado, o modelo tendeu a classificar o dispositivo de ar condicionado como outros dispositivos que apresentam algumas semelhanças em suas componentes de frequência: em 35,3% dos casos, o *kNN* o classificou como uma geladeira; em 34% como um micro-ondas; e em 27,8% dos casos, como um ventilador. A classificação menos frequente foi como um monitor, ocorrendo em apenas 3% dos casos. A Figura 33 traz um comparativo das amplitudes das componentes de frequência dos dispositivos abordados.

Figura 33 – Comparativo das amplitudes das componentes de frequência dos dispositivos.



Fonte: Elaborada pelo próprio autor.

Como pode ser observado, a amplitude da frequência fundamental de 60 Hz provou ser crucial para a classificação do modelo. Isso ocorre porque as amplitudes da frequência de 60 Hz do micro-ondas e da geladeira são as que mais se aproximam da amplitude do dispositivo de ar condicionado. Em seguida, o ventilador mostra-se como o dispositivo com a próxima semelhança, seguido de forma mais distante pelo monitor. Nas outras componentes de frequência, os valores de amplitude do dispositivo de ar condicionado se assemelham aos da geladeira e do micro-ondas, justificando assim os resultados de classificação deste experimento.

Em resumo, pode-se concluir que quando o modelo recebe dados de um dispositivo desconhecido, ele tende a classificá-lo como um dispositivo com componentes de frequência semelhantes presentes no conjunto de dados. Seria ainda mais valioso identificar que o dispositivo é desconhecido. Entretanto, essa funcionalidade está além do escopo deste trabalho. Nesse contexto, é recomendável que o máximo de dispositivos geralmente encontrados no ambiente físico em questão sejam adicionados ao conjunto de dados, com o objetivo de reduzir o erro causado por dispositivos desconhecidos.

5.5.4 Inclusão de novos dispositivos ao conjunto de dados

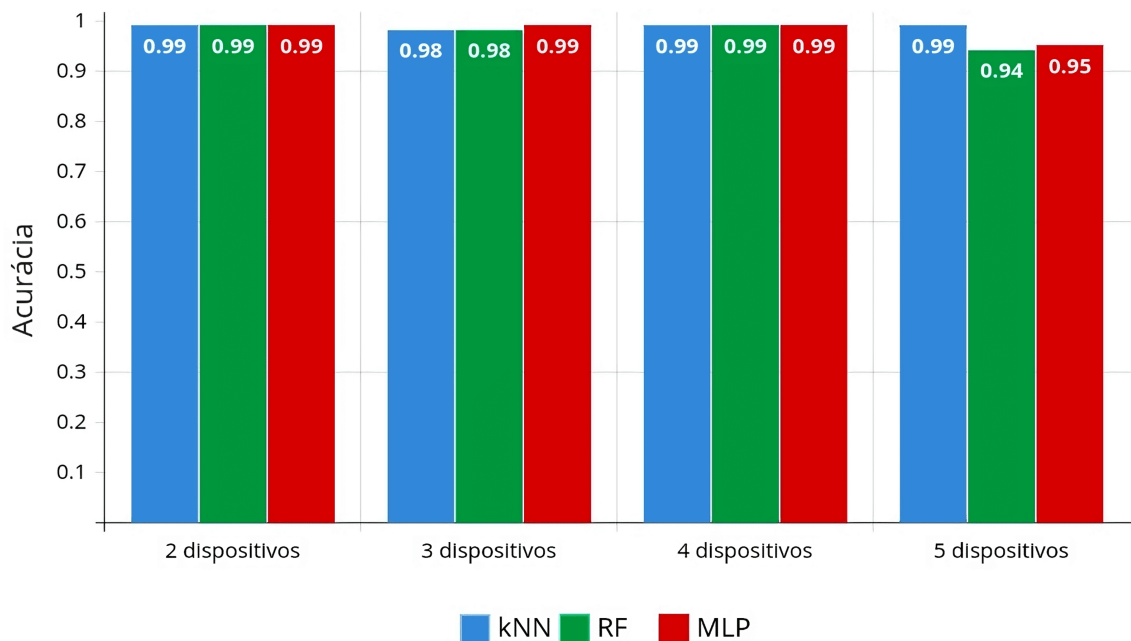
Outro aspecto de grande importância está ligado ao desempenho dos modelos quando amostras de novos dispositivos são introduzidas no conjunto de dados. Nesse contexto, foi realizado um experimento para analisar os resultados de acurácia e F1-score à medida que novos dispositivos são incorporados ao conjunto. Inicialmente, apenas as entradas correspondentes a 2

dispositivos foram deixadas no conjunto de dados, nomeadamente o micro-ondas e a geladeira. Em seguida, foram adicionadas as entradas relacionadas ao aparelho de ar condicionado. Posteriormente, foram incluídas as entradas associadas ao monitor e, por último, o ventilador. Em cada etapa, os classificadores foram avaliados em relação aos seus resultados de acurácia e F1-score.

Para este experimento, foram selecionados os classificadores *kNN*, *Random Forest* e *MLP* devido às suas abordagens distintas de classificação, o que pode afetar os resultados à medida que novos dispositivos são incluídos no conjunto de dados. Em cada configuração do conjunto de dados, ou seja, ao adicionar um novo dispositivo, todos os classificadores passaram por testes de validação cruzada com *k-fold* igual a 10, visando otimizar seus hiperparâmetros de acordo com o número atualizado de dispositivos no conjunto. Após determinar a configuração ideal dos hiperparâmetros, foi realizado o treinamento dos modelos.

É importante salientar que, neste ponto, foram utilizados apenas 60% dos dados, correspondentes ao conjunto de treinamento. Durante o treinamento, nenhum modelo teve acesso aos 40% restantes dos dados, referentes ao conjunto de teste. Essa abordagem é adotada para evitar resultados enviesados causados por sobreajuste (*overfitting*). Em seguida, cada modelo foi testado e avaliado em termos de sua acurácia e F1-score. Os resultados de acurácia estão apresentados no gráfico mostrado na Figura 34.

Figura 34 – Resultados de acurácia dos classificadores com a inclusão de novos dispositivos no conjunto de dados.

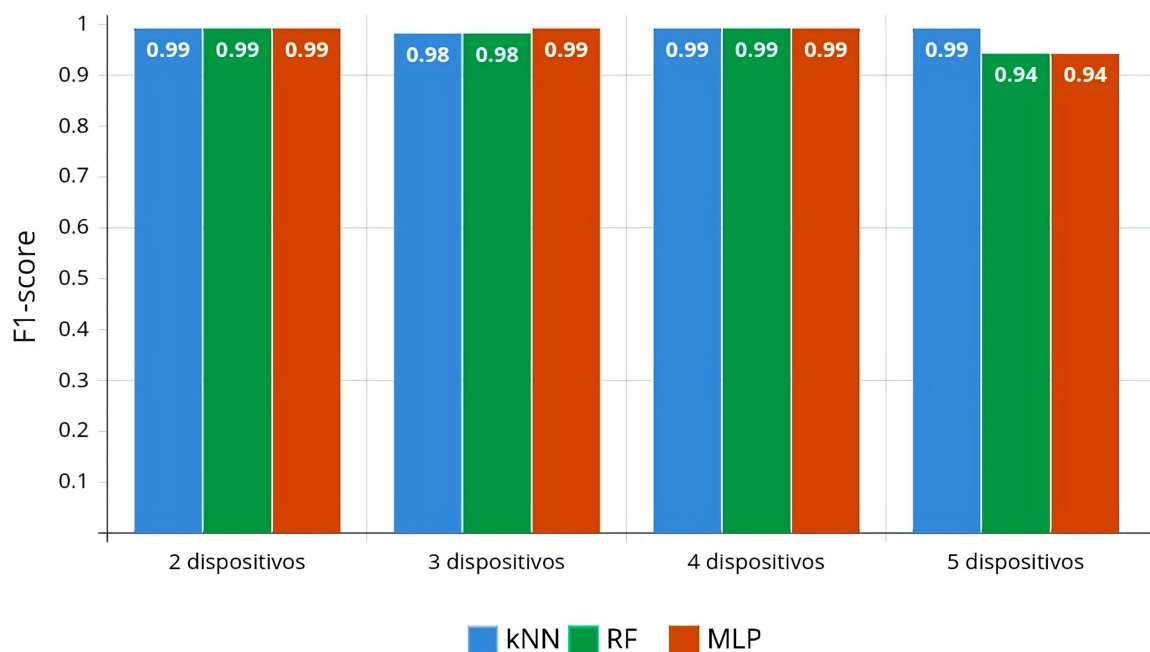


Fonte: Elaborada pelo próprio autor.

Como observado, todos os classificadores alcançaram aproximadamente 99% de acurácia quando o conjunto de dados continha apenas 2 dispositivos. Após a primeira inclusão, com o conjunto composto por 3 dispositivos, tanto o *kNN* quanto o *Random Forest* atingiram 98% de acurácia, enquanto o *MLP* permaneceu com 99%. Com a inclusão de mais um dispositivo, totalizando amostras de 4 dispositivos diferentes, todos os classificadores novamente alcançaram cerca de 99% de acurácia. Nessa perspectiva, nota-se que após duas inclusões de amostras de novos dispositivos, os classificadores não apresentaram variações significativas em seus resultados de acurácia.

No entanto, ao comparar os resultados com 4 dispositivos aos resultados com 5 dispositivos, observa-se uma leve degradação nos resultados do *Random Forest* para 94% e do *MLP* para cerca de 95%, enquanto o *kNN* permaneceu em torno de 99%. Resultados semelhantes foram obtidos utilizando a métrica F1-score, mostrados no gráfico da Figura 35. Em resumo, não houve oscilações significativas nos resultados de acurácia e F1-score dos classificadores. Nessa perspectiva, conclui-se que a inclusão de novos dispositivos tem um impacto relativamente pequeno nos resultados de classificação, sugerindo que os modelos são capazes de distinguir bem as características de frequência de cada dispositivo.

Figura 35 – Resultados de F1-score dos classificadores com a inclusão de novos dispositivos no conjunto de dados.



Fonte: Elaborada pelo próprio autor.

Entretanto, ao analisar os resultados deste experimento e os resultados do experi-

mento com dispositivos desconhecidos relatados na subseção 5.5.3, é possível destacar que a inclusão de dispositivos com componentes de frequência semelhantes pode ter um impacto mais significativo nos resultados de classificação. Por exemplo, dado que o conjunto de dados atual inclui amostras de um monitor, podemos inferir que a inclusão de amostras de uma TV teria um impacto considerável, dada a similaridade entre esses dispositivos. No entanto, é necessário questionar a real necessidade de distinguir dispositivos tão similares. Neste trabalho, optou-se por considerar esses casos como pertencentes ao mesmo dispositivo, não vendo a necessidade de uma classificação tão específica.

5.6 Monitoramento dos dispositivos elétricos

Monitorar dispositivos elétricos após sua identificação é crucial para garantir seu funcionamento adequado e detectar eventuais anomalias. Uma vez que um dispositivo é identificado, é possível analisar diversos parâmetros para verificar se está operando dentro do padrão esperado. Uma abordagem comum é verificar a amplitude da componente de frequência fundamental do sinal elétrico gerado pelo dispositivo.

A amplitude da componente de frequência fundamental é um indicador importante da saúde e do desempenho do dispositivo. Em dispositivos usados em ambientes residenciais, a amplitude de cada componente de frequência indica a contribuição dessa frequência particular para a potência total gerada pelo dispositivo. Essa potência deve se manter dentro de uma faixa específica para garantir que o dispositivo funcione corretamente. Valores muito baixos podem indicar falhas no funcionamento do dispositivo, como problemas no circuito de potência ou desgaste dos componentes internos. Por outro lado, amplitudes muito altas podem indicar superaquecimento ou sobrecarga do dispositivo, o que também pode ser prejudicial.

Para realizar essa análise, é necessário comparar a amplitude da componente de frequência fundamental do dispositivo em questão com os padrões estabelecidos para aquele tipo específico de dispositivo. Isso pode ser feito utilizando técnicas de análise de dados, como a comparação com dados históricos de dispositivos semelhantes ou a definição de limites aceitáveis de amplitude com base em especificações técnicas. Para ilustrar este conceito, foi realizado um experimento prático utilizando dados reais provenientes de dispositivos elétricos, os quais estão armazenados no conjunto de dados criado. No decorrer deste experimento, foi feita uma análise na amplitude da componente de frequência fundamental como um indicador fundamental do desempenho e da saúde desses dispositivos.

Para conduzir o experimento, primeiramente foi determinada a média da amplitude da frequência fundamental para os dispositivos de micro-ondas e geladeira presentes no conjunto de dados. Essa média serviu como referência para avaliar o desempenho individual de cada dispositivo. Em seguida, foi criado um conjunto de dados de teste com entradas dos dispositivos de micro-ondas e geladeira. Após a identificação de cada entrada do conjunto de teste, foi calculado o erro da amplitude da frequência fundamental desse dispositivo em relação à média do conjunto de dados. Um limiar de 30% foi estabelecido como critério para determinar se o dispositivo estava operando dentro dos padrões esperados. Se o erro fosse superior a esse limite, aciona-se um alerta indicando a possibilidade de um defeito no dispositivo.

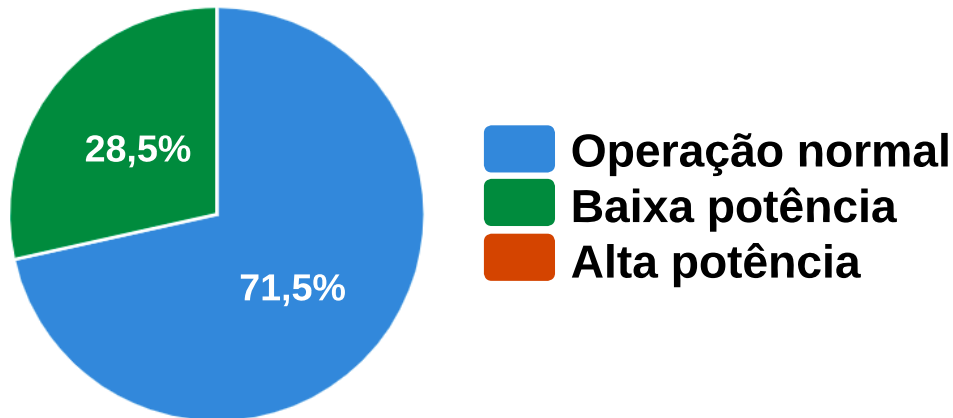
Além disso, se a amplitude da frequência fundamental do dispositivo fosse maior que a média, indicava um potencial de operação em alta potência. Por outro lado, se fosse menor que a média, sugeria uma operação com baixa potência. Essa abordagem possibilitou identificar possíveis problemas que podem ocorrer durante o funcionamento do dispositivo, fornecendo alertas automáticos em caso de desvios significativos em relação aos padrões estabelecidos. Essa análise sistemática permite uma rápida intervenção para corrigir problemas, garantindo assim a eficiência e a segurança operacional dos dispositivos elétricos.

A Figura 36 apresenta um gráfico dos resultados deste experimento. Após a análise das entradas dos dispositivos no conjunto de teste, observa-se que a maioria das amostras, totalizando 71,5%, estava dentro da faixa considerada como operação normal. Entretanto, aproximadamente 28,5% das entradas foram identificadas com baixa potência, indicando um possível desvio em relação à média estabelecida. É importante ressaltar que, embora nenhuma amostra com alta potência tenha sido encontrada, a presença de um número significativo de entradas com baixa potência é um indicativo que aciona um alerta para esses dispositivos.

É crucial ressaltar que outros parâmetros e variáveis também devem ser verificados para uma avaliação completa, porém, esse resultado representa um exemplo de sinais detectados com baixa potência, considerando o limiar estabelecido. Vale destacar que o limiar escolhido para detectar anomalias na amplitude da frequência fundamental pode ser facilmente ajustado com base nas informações fornecidas pelo fabricante do dispositivo. Além disso, após a identificação do dispositivo, é possível coletar e analisar várias outras informações para verificar sua saúde e desempenho, contribuindo para a manutenção preventiva e o diagnóstico de possíveis problemas.

Além da amplitude da componente de frequência fundamental, outros parâmetros podem ser monitorados para avaliar a saúde e o desempenho do dispositivo, como a presença

Figura 36 – Percentual de entradas classificadas como operação normal, baixa potência e alta potência.



Fonte: Elaborada pelo próprio autor.

de harmônicos indesejados, a estabilidade da frequência fundamental ao longo do tempo e a presença de ruídos ou distorções no sinal elétrico.

Em resumo, o monitoramento contínuo dos dispositivos elétricos após sua identificação é essencial para garantir seu funcionamento adequado e detectar problemas precocemente. A análise da amplitude da componente de frequência fundamental é uma parte importante desse processo, pois fornece percepções valiosas sobre a saúde e o desempenho do dispositivo. Ao manter os parâmetros elétricos dentro dos limites esperados, é possível prolongar a vida útil do dispositivo, evitar falhas e garantir a segurança e eficiência operacional.

6 CONCLUSÕES

No presente trabalho, foi explorado o conceito de classificação de dispositivos elétricos utilizando classificadores baseados em aprendizado de máquina. Ao revisar o estado da arte, foram encontradas várias abordagens que se propõem a resolver o mesmo problema que o enfrentado neste trabalho. Este estudo, por sua vez, focou na classificação com base no sinal de corrente no domínio da frequência, empregando técnicas de aprendizado de máquina.

A identificação visual de dispositivos elétricos é uma prática comum no cotidiano, exigindo frequentemente intervenção física para determinar quais dispositivos estão conectados e em funcionamento. No entanto, essa abordagem é muitas vezes ineficiente, especialmente em ambientes com grande número de dispositivos ou em locais de difícil acesso. Em contraste, a abordagem automatizada utilizada neste trabalho busca simplificar esse processo por meio da análise do sinal de corrente elétrica, eliminando a necessidade de intervenção manual e oferecendo potencial para uma gestão mais eficiente e monitoramento dos dispositivos elétricos em uma variedade de ambientes.

Além das limitações inerentes à identificação visual, existem diversos desafios adicionais que tornam a automação dessa tarefa ainda mais crucial e complexa em determinados contextos. Em ambientes hospitalares, por exemplo, onde o fornecimento ininterrupto de energia com qualidade é vital para o funcionamento dos equipamentos médicos, a rápida e precisa identificação dos dispositivos ligados na rede elétrica é fundamental, especialmente em situações de emergência ou quando há dependência de geradores de energia (Rebonatto *et al.*, 2023).

O roubo de energia é outra preocupação significativa, especialmente em áreas residenciais. No Brasil, não há atualmente formas de identificar dispositivos clandestinos conectados a rede elétrica de uma residência. A identificação automatizada de dispositivos elétricos pode ajudar a detectar padrões de consumo suspeitos e identificar atividades ilegais de desvio de energia, auxiliando as autoridades na aplicação da lei e na proteção da integridade das redes elétricas.

Além disso, o uso indevido de energia pública, como o carregamento não autorizado de baterias em espaços públicos, pode ser mitigado com a automatização da identificação de dispositivos elétricos. Isso permite um monitoramento mais eficaz do uso da energia pública e a implementação de medidas para garantir seu uso justo e equitativo, beneficiando toda a comunidade.

Em resumo, muitos são os problemas enfrentados pelo método atual, que não é capaz

de identificar quais dispositivos estão conectados a rede elétrica. Nesse contexto, ao enfrentar esses desafios e comparar este estudo com a linha de base tradicional, procura-se não apenas melhorar a eficiência e a precisão da identificação de dispositivos elétricos, mas também mitigar riscos potenciais e promover o uso responsável e seguro da energia elétrica em diversos contextos e cenários.

6.1 Considerações finais

Então, este trabalho começa construindo um conjunto de dados que incorpora as amplitudes das 20 primeiras componentes de frequência do sinal de corrente, obtidas a partir de dispositivos domésticos reais. Essa abordagem permitiu representar de forma eficiente as características essenciais dos dispositivos elétricos, enquanto o volume de dados foi reduzido em cinco vezes em comparação com conjuntos de dados baseados no domínio do tempo. Essa redução foi crucial para garantir a viabilidade computacional e a eficácia do método de classificação utilizado.

Para validar a metodologia, foram realizados diversos experimentos. Primeiramente, comparando os resultados obtidos no domínio da frequência com os do domínio do tempo, a fim de avaliar a eficácia da abordagem empregada. Em seguida, foram conduzidos experimentos adicionais, nos quais foram introduzidos ruídos gerados por outros dispositivos ligados simultaneamente, visando entender como tais interferências afetam o processo de classificação. Também foi analisado o impacto da adição de novos dispositivos ao conjunto de dados e investigado a capacidade de classificação frente a dispositivos desconhecidos.

Foram alcançados resultados promissores, com uma taxa de acerto de aproximadamente 99% ao empregar o modelo *kNN*. O requisito de um número reduzido de dados para treinamento e classificação é crucial em ambientes *IoT*, sobretudo pela escassez ou racionamento de recursos energéticos. Portanto, a abordagem no domínio da frequência desempenha um papel significativo nesse contexto. Este trabalho valida resultados de classificação promissores utilizando uma abordagem que requer menos dados em comparação com os estudos mencionados na Capítulo 3.

Além disso, também foi realizado o monitoramento dos dispositivos. Foi analisada inicialmente a frequência fundamental do sinal de corrente como indicador da saúde de funcionamento dos dispositivos. Essa abordagem permitiu detectar anomalias e identificar problemas potenciais de forma eficiente. Ademais, outras características do sinal de corrente também podem

ser exploradas para uma avaliação mais abrangente da condição dos dispositivos.

Em resumo, este trabalho alcançou os objetivos específicos propostos. Foi desenvolvido um banco de dados robusto que relaciona dispositivos elétricos às suas componentes de frequência no sinal de corrente, permitindo a avaliação de classificadores baseados em aprendizado de máquina. Os experimentos realizados demonstraram a eficácia desses classificadores na identificação de dispositivos elétricos domésticos, com uma taxa de acerto próxima de 99%. O monitoramento dos dispositivos permitiu identificar anomalias e irregularidades no consumo de energia. Além disso, este trabalho conseguiu reduzir o volume de dados em até cinco vezes em comparação com abordagens baseadas no domínio do tempo, validando assim a viabilidade da solução para ambientes de *IoT* com recursos computacionais e energéticos limitados.

Em relação às perguntas norteadoras apresentadas no Capítulo 1, foi mostrado que o gerenciamento eficiente do consumo de energia em ambientes com muitos dispositivos conectados pode ser alcançado por meio da análise das componentes de frequência no sinal de corrente elétrica, permitindo a identificação dos dispositivos em operação. Com essa identificação, é possível determinar quais dispositivos estão conectados de forma indevida e detectar qualquer anomalia no consumo de energia ao comparar suas características reais com as especificações fornecidas pelo fabricante.

Em relação aos desafios enfrentados na identificação e monitoramento de dispositivos em grandes edificações, onde o risco de desperdício de energia é significativo, esta pesquisa lidou com várias dificuldades. Entre os principais desafios estavam os ruídos na rede elétrica e o grande número de dispositivos conectados. Para superar esses obstáculos, foram desenvolvidos métodos para filtrar os ruídos e reduzir a quantidade de dados necessários para o processo de monitoramento, levando em conta as restrições de recursos em ambientes de *IoT*. Essas abordagens foram cruciais para assegurar uma identificação e monitoramento eficientes, mesmo em cenários complexos e com recursos limitados.

6.2 Trabalhos futuros

Em síntese, este estudo traz contribuições significativas para a gestão eficiente de recursos energéticos em ambientes de *IoT* em residências de médio e grande porte. Os próximos passos nessa linha de pesquisa envolvem a investigação da desagregação de cargas elétricas, permitindo a identificação de sinais de corrente provenientes de múltiplos dispositivos. Adicionalmente, este trabalho busca aprimorar a identificação de dispositivos desconhecidos,

visando determinar se eles fazem parte do conjunto de dados ou não.

Em complemento, este trabalho também pretende expandir os experimentos realizados por meio da aplicação de outros classificadores e técnicas de aprendizado profundo, visando aprimorar a precisão dos modelos. Além disso, planeja-se incluir novos dispositivos no conjunto de dados, ampliando a diversidade e a representatividade dos sinais capturados, o que contribuirá para uma análise mais robusta e abrangente.

REFERÊNCIAS

- ABEYKOON, V.; KANKANAMDURAGE, N.; SENEVIRATHNA, A.; RANAWEERA, P.; UDAWALPOLA, R. Real time identification of electrical devices through power consumption pattern detection. **Pervasive Comput**, v. 10, n. 1, p. 40–48, 2016.
- ALBASIOUNY, E. R.; HELIEL, A.-F. A.; ABDELMUNIM, H. E.; ABBAS, H. M. Multilayer perceptron generative model via adversarial learning for robust visual tracking. **IEEE Access**, v. 10, p. 121230–121248, 2022.
- ALBERTI, A. M.; SANTOS, M. A. S.; SOUZA, R.; SILVA, H. D. L. D.; CARNEIRO, J. R.; FIGUEIREDO, V. A. C.; RODRIGUES, J. J. P. C. Platforms for smart environments and future internet design: A survey. **IEEE Access**, v. 7, p. 165748–165778, 2019.
- ALMEIDA, J. S.; MARINHO, L. B.; SOUZA, J. M.; ASSIS, E.; FILHO, P. R. Localization system for autonomous mobile robots using machine learning methods and omnidirectional sonar. **IEEE Latin America Transactions**, IEEE, v. 16, n. 2, p. 368–374, 2018.
- ARIVAMUDHAN, M.; SANTHI, S. Fpga realization of fft and dwt algorithms for the detection of transformer winding deformation. In: **2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)**. [S. l.: s. n.], 2018. p. 210–214.
- ATLAM, H. F.; WALTERS, R. J.; WILLS, G. B. Fog computing and the internet of things: A review. **Big Data and Cognitive Computing**, v. 2, n. 2, 2018. ISSN 2504-2289.
- BAHARUDDIN, R. R.; NISWAR, M.; ILHAM, A. A.; KASHIHARA, S. Crab molting identification using machine learning classifiers. In: **2021 International Seminar on Machine Learning, Optimization, and Data Science (ISMODE)**. [S. l.: s. n.], 2022. p. 295–300.
- BARBI, I. A theorem on power superposition in resistive networks. **IEEE Transactions on Circuits and Systems II: Express Briefs**, v. 68, n. 7, p. 2362–2363, 2021.
- BARKER, S.; MUSTHAG, M.; IRWIN, D.; SHENOY, P. Non-intrusive load identification for smart outlets. In: **2014 IEEE International Conference on Smart Grid Communications (SmartGridComm)**. [S. l.: s. n.], 2014. p. 548–553.
- BARRON, L. The road to a smarter future: The smart city, connected cars and autonomous mobility. In: **2021 26th International Conference on Automation and Computing (ICAC)**. [S. l.: s. n.], 2021. p. 1–6.
- BOURSIANIS, A. D.; PAPADOPOULOU, M. S.; DIAMANTOULAKIS, P.; LIOPATSAKALIDI, A.; BAROUCHAS, P.; SALAHAS, G.; KARAGIANNIDIS, G.; WAN, S.; GOUDOS, S. K. Internet of things (iot) and agricultural unmanned aerial vehicles (uavs) in smart farming: A comprehensive review. **Internet of Things**, v. 18, p. 100187, 2022. ISSN 2542-6605.
- CHICCO, D.; JURMAN, G. The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation. **BMC genomics**, Springer, v. 21, n. 1, p. 1–13, 2020.
- CHIN, C. S.; ZHANG, J. Wavelet scattering transform for multiclass support vector machines in audio devices classification system. In: **2021 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)**. [S. l.: s. n.], 2021. p. 735–740.

- CHOU, M.; THOURN, K.; CHEA, R. Multi-scale electrical appliance load signature for non-intrusive load monitoring classification. In: **2022 14th International Conference on Software, Knowledge, Information Management and Applications (SKIMA)**. [S. l.: s. n.], 2022. p. 79–83.
- DEY, S. K.; HOSSAIN, A.; RAHMAN, M. M. Implementation of a web application to predict diabetes disease: an approach using machine learning algorithm. In: IEEE. **2018 21st international conference of computer and information technology (ICCIIT)**. [S. l.], 2018. p. 1–5.
- DHANKA, S.; MAINI, S. Random forest for heart disease detection: A classification approach. In: **2021 IEEE 2nd International Conference On Electrical Power and Energy Systems (ICEPES)**. [S. l.: s. n.], 2021. p. 1–3.
- DIAS, V. H. A.; DIAS, P. M. C. Escrevendo o “livro da natureza” na linguagem da matemática: A lei de ampère. **Revista Brasileira de Ensino de Física**, SciELO Brasil, v. 37, p. 4601–1, 2015.
- DOU, H.; WANG, G. A dynamic multitarget detection algorithm in front of vehicle based on embedded system and internet of things. **Scientific Programming**, Hindawi, v. 2022, 2022.
- ECEIZA, M.; FLORES, J. L.; ITURBE, M. Fuzzing the internet of things: A review on the techniques and challenges for efficient vulnerability discovery in embedded systems. **IEEE Internet of Things Journal**, v. 8, n. 13, p. 10390–10411, 2021.
- FARAHZADI, A.; SHAMS, P.; REZAZADEH, J.; FARAHBAKHSR, R. Middleware technologies for cloud of things: a survey. **Digital Communications and Networks**, v. 4, n. 3, p. 176–188, 2018. ISSN 2352-8648.
- FAYZRAKHMANTOV, R.; KULIKOV, A.; REPP, P. The difference between precision-recall and roc curves for evaluating the performance of credit card fraud detection models. In: ANHALT UNIVERSITY OF APPLIED SCIENCES. **Proceedings of International Conference on Applied Innovation in IT**. [S. l.], 2018. v. 6, n. 1, p. 17–22.
- FEITOSA, S. O. *et al.* **Desenvolvimento e aplicabilidade de um multimedidor de grandezas elétricas utilizando plataforma arduíno**. [S. l.]: Pesqueira, 2021.
- FENG, Y.; JIA, L.; ZHANG, J.; CHEN, J. Ffswinnet: Cnn-transformer combined network with fft for shale core sem image segmentation. **IEEE Access**, p. 1–1, 2024.
- FERREIRA, J. C.; AFONSO, J. A.; MONTEIRO, V.; AFONSO, J. L. An energy management platform for public buildings. **Electronics**, MDPI, v. 7, n. 11, p. 294, 2018.
- GUO, Y.; ZHOU, Y.; HU, X.; CHENG, W. Research on recommendation of insurance products based on random forest. In: IEEE. **2019 International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI)**. [S. l.], 2019. p. 308–311.
- HESSEL, R.; FRESCHI, A. A.; SANTOS, F. J. d. Lei de indução de faraday: Uma verificação experimental. **Revista Brasileira de Ensino de Física**, SciELO Brasil, v. 37, 2015.
- HU, Z.; HAN, Y.; ZALHAF, A. S.; ZHOU, S.; ZHAO, E.; YANG, P. Harmonic sources modeling and characterization in modern power systems: A comprehensive overview. **Electric Power Systems Research**, v. 218, p. 109234, 2023. ISSN 0378-7796.

HUSEIEN, G. F.; SHAH, K. W. A review on 5g technology for smart energy management and smart buildings in singapore. **Energy and AI**, v. 7, p. 100116, 2022. ISSN 2666-5468.

HUY, N. L.; THANH, P. C.; LINH, T. D.; STOJCEVSKI, A. Design and simulation of a novel clockless fast fourier transform (fft) circuit. In: **2017 Seventh International Conference on Information Science and Technology (ICIST)**. [S. l.: s. n.], 2017. p. 304–310.

INIBHUNU, C.; MCGREGOR, A. C. A privacy preserving framework for smart cities utilising iot, smart buildings and big data. In: **2020 IEEE 22nd International Conference on High Performance Computing and Communications; IEEE 18th International Conference on Smart City; IEEE 6th International Conference on Data Science and Systems (HPCC/SmartCity/DSS)**. [S. l.: s. n.], 2020. p. 1096–1103.

JADHAV, T.; TANISHQ; JAGADISH, S.; GAIKWAD, M.; WAGLE, S.; JAYASWAL, R.; PATIL, S. Predicting urban land cover using classification: A machine learning approach. In: **2023 IEEE 11th Region 10 Humanitarian Technology Conference (R10-HTC)**. [S. l.: s. n.], 2023. p. 450–454.

JAIN, A.; SHAH, K.; CHATURVEDI, P.; TAMBE, A. Prediction and analysis of student performance using hybrid model of multilayer perceptron and random forest. In: **2018 International Conference on Advanced Computation and Telecommunication (ICACAT)**. [S. l.: s. n.], 2018. p. 1–7.

KALOEV, M.; KRASSTEV, G. Tailored learning rates for reinforcement learning: A visual exploration and guideline formulation. In: **2023 7th International Symposium on Innovative Approaches in Smart Technologies (ISAS)**. [S. l.: s. n.], 2023. p. 1–7.

KATE, C.; KALPANA, C.; SHARMA, A.; YADAV, A. S.; KUMAR, A.; KUMAR, S. S. Investigation of machine learning algorithms for pattern recognition in image processing. In: **2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA)**. [S. l.: s. n.], 2023. p. 898–904.

KHAN, M. M. R.; SIDDIQUE, M. A. B.; SAKIB, S. Non-intrusive electrical appliances monitoring and classification using k-nearest neighbors. In: **2019 2nd International Conference on Innovation in Engineering and Technology (ICIET)**. [S. l.: s. n.], 2019. p. 1–5.

LI, J.-F.; LU, S.-K.; HWANG, S.-A.; WU, C.-W. Easily testable and fault-tolerant fft butterfly networks. **IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing**, v. 47, n. 9, p. 919–929, 2000.

LIANG, J.; NG, S. K.; KENDALL, G.; CHENG, J. W. Load signature study—part i: Basic concept, structure, and methodology. **IEEE transactions on power Delivery**, IEEE, v. 25, n. 2, p. 551–560, 2009.

LIN, J.; YU, W.; ZHANG, N.; YANG, X.; ZHANG, H.; ZHAO, W. A survey on internet of things: Architecture, enabling technologies, security and privacy, and applications. **IEEE Internet of Things Journal**, v. 4, n. 5, p. 1125–1142, 2017.

LIU, W.; LIAO, Q.; QIAO, F.; XIA, W.; WANG, C.; LOMBARDI, F. Approximate designs for fast fourier transform (fft) with application to speech recognition. **IEEE Transactions on Circuits and Systems I: Regular Papers**, v. 66, n. 12, p. 4727–4739, 2019.

- LIU, Y.; WANG, C.; SUN, J.; DU, S.; HONG, Q. One-step calculation circuit of fft and its application. **IEEE Transactions on Circuits and Systems I: Regular Papers**, v. 69, n. 7, p. 2781–2793, 2022.
- LU, Y.; YE, T.; ZHENG, J. Decision tree algorithm in machine learning. In: **2022 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA)**. [S. l.: s. n.], 2022. p. 1014–1017.
- MAHESH, B. Machine learning algorithms-a review. **International Journal of Science and Research (IJSR)**. [Internet]. [S. l.], v. 9, p. 381–386, 2020.
- MAHMOUD, R.; YOUSUF, T.; ALOUL, F.; ZUALKERNAN, I. Internet of things (iot) security: Current status, challenges and prospective measures. In: **2015 10th International Conference for Internet Technology and Secured Transactions (ICITST)**. [S. l.: s. n.], 2015. p. 336–341.
- MASSIDDA, L.; MARROCU, M.; MANCA, S. Non-intrusive load disaggregation by convolutional neural network and multilabel classification. **Applied Sciences**, v. 10, n. 4, 2020. ISSN 2076-3417.
- MEENA, R.; DUBEY, S. Smart houses with the application of energy management system & smart grid. In: **2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)**. [S. l.: s. n.], 2021. p. 947–952.
- MIAH, M. O.; KHAN, S. S.; SHATABDA, S.; FARID, D. M. Improving detection accuracy for imbalanced network intrusion classification using cluster-based under-sampling with random forests. In: **2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)**. [S. l.: s. n.], 2019. p. 1–5.
- MICHAŁOWSKA, J.; TOFIL, A.; JÓZWIK, J. Analysis of the radiated interferences emission of small household appliances. In: **2019 Applications of Electromagnetics in Modern Engineering and Medicine (PTZE)**. [S. l.: s. n.], 2019. p. 117–120.
- MISHRA, K. M.; HUHTALA, K. J. Fault detection of elevator systems using multilayer perceptron neural network. In: **2019 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)**. [S. l.: s. n.], 2019. p. 904–909.
- NIU, W.; SONG, X.; WANG, H.; CHEN, J. Forecasting the load flow of engine driven pump based on light gradient boosting machine model. In: **2021 IEEE International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI)**. [S. l.: s. n.], 2021. p. 771–774.
- OPPENHEIM, A. V.; WILLSKY, A. S. **Sinai e Sistemas**. [S. l.]: Massachusetts Institute of Technology, 2010. ISBN 978-85-4301-380-0.
- OSISANWO, F.; AKINSOLA, J.; AWODELE, O.; HINMIKAIYE, J.; OLAKANMI, O.; AKINJOBI, J. Supervised machine learning algorithms: classification and comparison. **International Journal of Computer Trends and Technology (IJCTT)**. [S. l.], v. 48, n. 3, p. 128–138, 2017.
- OZCAN, M. O.; ODACI, F.; ARI, I. Remote debugging for containerized applications in edge computing environments. In: **2019 IEEE International Conference on Edge Computing (EDGE)**. [S. l.: s. n.], 2019. p. 30–32.

PALOMSBA, V.; BRUNACCINI, G.; ALOISIO, D.; DINO, G. E.; LUNA, M.; TONA, G. L.; PIAZZA, M. C. D.; FRAZZICA, A.; SERGI, F. A smart thermal management system for high efficiency data center cooling in microgrid environment. In: **2022 AEIT International Annual Conference (AEIT)**. [S. l.: s. n.], 2022. p. 1–6.

PANORI, A.; KAKDERI, C.; KOMNINOS, N.; FELLNHOFER, K.; REID, A.; MORA, L. Smart systems of innovation for smart places: Challenges in deploying digital platforms for co-creation and data-intelligence. **Land Use Policy**, v. 111, p. 104631, 2021. ISSN 0264-8377.

PHILIPPOU, N.; AJOODHA, R.; JADHAV, A. Using machine learning techniques and matric grades to predict the success of first year university students. In: IEEE. **2020 2nd International Multidisciplinary Information Technology and Engineering Conference (IMITEC)**. [S. l.], 2020. p. 1–5.

PRIHATNO, A. T.; NURCAHYANTO, H.; JANG, Y. M. Predictive maintenance of relative humidity using random forest method. In: IEEE. **2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC)**. [S. l.], 2021. p. 497–499.

RAMTEKE, N.; MAIDAMWAR, P. Cardiac patient data classification using ensemble machine learning technique. In: **2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)**. [S. l.: s. n.], 2023. p. 1–6.

REBONATTO, M. T.; PAIXÃO, O. J. P. da; SPALDING, L. E. S. Detecção e alerta de equipamentos não permitidos em quartos hospitalares por meio da supervisão da corrente elétrica. **Revista Brasileira de Computação Aplicada**, v. 15, n. 2, p. 73–87, jul. 2023.

RICHARDSON, C.; SHAH, S.; YUAN, X. Semi-supervised machine learning for analyzing covid-19 related twitter data for asian hate speech. In: **2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA)**. [S. l.: s. n.], 2022. p. 1643–1648.

ROCHA, F.; DANTAS, L. C.; SANTOS, L. F.; FERREIRA, S.; SOARES, B.; FERNANDES, A.; CAVALCANTE, E.; BATISTA, T. Energy efficiency in smart buildings: An iot-based air conditioning control system. In: CASACA, A.; KATKOORI, S.; RAY, S.; STROUS, L. (Ed.). **Internet of Things. A Confluence of Many Disciplines**. Cham: Springer International Publishing, 2020. p. 21–35. ISBN 978-3-030-43605-6.

ROHITH, M.; SUNIL, A.; MOHANA. Comparative analysis of edge computing and edge devices: Key technology in iot and computer vision applications. In: **2021 International Conference on Recent Trends on Electronics, Information, Communication Technology (RTEICT)**. [S. l.: s. n.], 2021. p. 722–727.

SAKASHITA, Y.; YAMANASHI, Y.; YOSHIKAWA, N. High-speed operation of an sfq butterfly processing circuit for fft processors using the 10 ka/cm² nb process. **IEEE Transactions on Applied Superconductivity**, v. 25, n. 3, p. 1–5, 2015.

SEHRAWAT, D.; GILL, N. S. Smart sensors: Analysis of different types of iot sensors. In: IEEE. **2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)**. [S. l.], 2019. p. 523–528.

SHEU, M.-H.; YANG, E.-C.; WANG, S.-T.; JUANG, W.-H.; LAI, S.-C. High precision, low complexity, and fast calculation based on hybrid recursive dft and fft algorithms for

electrochemical impedance spectroscopy system. **IEEE Transactions on Instrumentation and Measurement**, p. 1–1, 2024.

SINGH, A.; THAKUR, N.; SHARMA, A. A review of supervised machine learning algorithms. In: IEEE. **2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)**. [S. l.], 2016. p. 1310–1315.

SINHA, B. B.; DHANALAKSHMI, R. Recent advancements and challenges of internet of things in smart agriculture: A survey. **Future Generation Computer Systems**, v. 126, p. 169–184, 2022. ISSN 0167-739X.

SLIJEPCEVIC, D.; ZEPPELZAUER, M.; UNGLAUBE, F.; KRANZL, A.; BREITENEDER, C.; HORSACK, B. Explainable machine learning in human gait analysis: A study on children with cerebral palsy. **IEEE Access**, p. 1–1, 2023.

SOLOVEV, D. B.; SHADRIN, A. S. Instrument current transducer for measurements in asymmetrical conditions in three-phase circuits with upper harmonics. **International Journal of Electrical Power & Energy Systems**, Elsevier, v. 84, p. 195–201, 2017.

SU, Y. Prediction of air quality based on gradient boosting machine method. In: **2020 International Conference on Big Data and Informatization Education (ICBDIE)**. [S. l.: s. n.], 2020. p. 395–397.

SUBRAMANIAN, S.; TSENG, B.; BARBIERI, R.; BROWN, E. N. Unsupervised machine learning methods for artifact removal in electrodermal activity. In: **2021 43rd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)**. [S. l.: s. n.], 2021. p. 399–402.

SWARNKAR, R.; HARIKRISHNAN, R.; THAKUR, P.; SINGH, G. Electric vehicle lithium-ion battery ageing analysis under dynamic condition: A machine learning approach. **SAIEE Africa Research Journal**, v. 114, n. 1, p. 4–13, 2023.

TANG, J.; DENG, C.; HUANG, G.-B. Extreme learning machine for multilayer perceptron. **IEEE Transactions on Neural Networks and Learning Systems**, v. 27, n. 4, p. 809–821, 2016.

TEKLER, Z. D.; LOW, R.; YUEN, C.; BLESSING, L. Plug-mate: An iot-based occupancy-driven plug load management system in smart buildings. **Building and Environment**, v. 223, p. 109472, 2022. ISSN 0360-1323.

THEPADE, S. D.; KALBHOR, M. M. Ensemble of machine learning classifiers for improved image category prediction using fractional coefficients of hartley and sine transforms. In: **2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)**. [S. l.: s. n.], 2018. p. 1–5.

TOUZANI, S.; GRANDERSON, J.; FERNANDES, S. Gradient boosting machine for modeling the energy consumption of commercial buildings. **Energy and Buildings**, Elsevier, v. 158, p. 1533–1543, 2018.

TRIPATHY, R. K.; DASH, D. K.; GHOSH, S. K.; PACHORI, R. B. Detection of different stages of anxiety from single channel wearable ecg sensor signal using fourier-bessel domain adaptive wavelet transform. **IEEE Sensors Letters**, p. 1–4, 2023.

TUNDIS, A.; FAIZAN, A.; MÜHLHÄUSER, M. A feature-based model for the identification of electrical devices in smart environments. **Sensors**, MDPI, v. 19, n. 11, p. 2611, 2019.

WANG, K.; WANG, Y.; SUN, Y.; GUO, S.; WU, J. Green industrial internet of things architecture: An energy-efficient perspective. **IEEE Communications Magazine**, v. 54, n. 12, p. 48–54, 2016.

WANG, Y.; WEN, H.; HOU, X.; TANG, H.; SUN, H.; ZHENG, K.; LI, S. Comparison of differential-mode and mixed-mode conducted emission for household appliances in power-line communication system. **IEEE Transactions on Electromagnetic Compatibility**, v. 59, n. 6, p. 2023–2028, 2017.

WU, X.; GAO, Y.; JIAO, D. Multi-label classification based on random forest algorithm for non-intrusive load monitoring system. **Processes**, v. 7, n. 6, 2019. ISSN 2227-9717.

YAN, Z.; WEN, H. Comparative study of electricity-theft detection based on gradient boosting machine. In: IEEE. **2021 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)**. [S. l.], 2021. p. 1–6.

YANG, J.; KWON, Y.; KIM, D. Regional smart city development focus: The south korean national strategic smart city program. **IEEE Access**, v. 9, p. 7193–7210, 2021.

YI, Y.; GU, J.; LUO, L.; WANG, Z.; HAN, B.; HE, H.; YIN, S. Efficient and flexible implementation of fft application for cgra processor. In: **2023 8th International Conference on Intelligent Computing and Signal Processing (ICSP)**. [S. l.: s. n.], 2023. p. 957–960.

ZAMARUIEV, V. V. The use of kotelnikov-nyquist-shannon sampling theorem for designing of digital control system for a power converter. In: **2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON)**. [S. l.: s. n.], 2017. p. 522–527.

ZHANG, S.; LI, X.; ZONG, M.; ZHU, X.; CHENG, D. Learning k for knn classification. **ACM Transactions on Intelligent Systems and Technology (TIST)**, ACM New York, NY, USA, v. 8, n. 3, p. 1–19, 2017.

ZHANG, Z. Decision trees for objective house price prediction. In: **2021 3rd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI)**. [S. l.: s. n.], 2021. p. 280–283.

ÖZCAN, U.; ARSLAN, A.; İLKYZAZ, M.; KARAARSLAN, E. An augmented reality application for smart campus urbanization: Msku campus prototype. In: **2017 5th International Istanbul Smart Grid and Cities Congress and Fair (ICSG)**. [S. l.: s. n.], 2017. p. 100–104.