

ARTIGO ORIGINAL

Identificação Automática de Dispositivos Elétricos em Ambientes Domésticos

Automatic Identification of Electrical Appliances in Domestic Settings

Antonio D. T. Amurim ¹ and Elvis M. G. Stancanelli ¹

¹Campus Quixadá – Universidade Federal do Ceará (UFC), Quixadá – CE – Brasil

* davidamurim7@alu.ufc.br; elvis.stancanelli@ufc.br

Recebido: 23/09/2024. Revisado: 22/07/2025. Aceito: 31/07/2025.

Resumo

A Internet das Coisas (IoT) é fundamental para a automação, monitoramento e controle em ambientes modernos. No entanto, o gerenciamento eficaz de inúmeros dispositivos elétricos nesses cenários apresenta desafios significativos. Uma lacuna na compreensão do estado operacional e do consumo de energia desses dispositivos pode levar ao desperdício de energia e encurtar sua vida útil. Este artigo foca na identificação automática de dispositivos elétricos em ambientes IoT, utilizando características extraídas do domínio da frequência dos sinais de corrente elétrica. Coletamos medições de consumo de energia de eletrodomésticos e as aplicamos a algoritmos de aprendizado de máquina. Nossos resultados destacam a eficácia do classificador k-Nearest Neighbors, que alcançou um impressionante F1-score de 99% na identificação de dispositivos.

Palavras-Chave: Aprendizado de máquina; Classificação; Assinatura de carga; Internet das coisas.

Abstract

The Internet of Things (IoT) is pivotal for automation, monitoring, and control in modern environments. However, effectively managing numerous electrical devices within these settings presents significant challenges. A critical gap in understanding the operational status and power consumption of these devices often results in energy waste and can shorten their lifespan. This paper focuses on the automatic identification of electrical devices in IoT environments, leveraging features extracted from the frequency domain of electrical current signals. We collected power consumption measurements from household appliances and fed them to machine learning algorithms. Our results highlight the efficacy of the k-Nearest Neighbors classifier, which achieved an impressive F1-score of 99% in device identification.

Keywords: Machine learning; Classification; Load signature; Internet of things.

1 Introdução

Atualmente, assistimos a uma integração substancial e crescente da tecnologia da informação em nossas vidas cotidianas, o que está impactando significativamente o consumo de eletricidade (Swarnkar et al., 2023; Palomsba et al., 2022; Barron, 2021; Meena e Dubey, 2021). A ascen-

são da Internet das Coisas (IoT) tem levado a um aumento no consumo de eletricidade, particularmente em edifícios comerciais e residenciais (Gaikwad et al., 2025; Illakya et al., 2024; Yang et al., 2021; Rocha et al., 2020; Inibhunu e Carolyn McGregor, 2020). Nesse cenário, qualquer lapso na gestão de recursos energéticos pode culminar em desperdício e custos elevados.

Para mitigar esse problema, a identificação e o monitoramento de dispositivos elétricos, por meio de suas assinaturas de carga, emergem como uma solução promissora. Essas assinaturas de carga são essencialmente arquétipos ou padrões nos sinais elétricos (Shahid et al., 2024; Nuran et al., 2023; Chin e Zhang, 2021). Destarte, é possível identificar conexões inadequadas ou mesmo não autorizadas de dispositivos específicos à rede elétrica, assim como monitorar o estado operacional dos equipamentos, detectando prováveis anomalias ao comparar dados capturados em tempo real com os parâmetros especificados pelo fabricante do dispositivo e considerando o histórico operacional do equipamento. Como exemplo, o artigo Rebonatto et al. (2023) debruça-se sobre essa problemática dentro de um ambiente hospitalar.

Ao vislumbrar uma gama mais ampla de aplicações, acreditamos que algoritmos de aprendizado de máquina podem ser utilizados para categorizar dispositivos com base em suas assinaturas (Baharuddin et al., 2022; Jadhav et al., 2023; Kaloev e Krastev, 2023; Slijepcevic et al., 2023; Lu et al., 2022; Subramanian et al., 2021; Richardson et al., 2022). Este artigo traz um estudo com vistas a automatizar a identificação de eletrodomésticos por meio do reconhecimento de padrões no sinal de corrente elétrica dentro de um contexto IoT. Adotando uma abordagem comparativa, diversos algoritmos renomados foram postos a trabalhar no domínio da frequência, diminuindo-se assim o fluxo de dados. A diminuição no fluxo de dados tem grande valia em um ambiente com recursos computacionais e energéticos racionados. Os algoritmos de reconhecimento de padrões são aplicados e avaliados usando como métricas a acurácia e o F1-score. O propósito é auxiliar na gestão eficiente de recursos energéticos, reduzindo desperdícios e antevendo-se a anomalias no consumo de energia. Nossos experimentos alcançaram resultados mostrando acurácia e F1-score de cerca de 99% na identificação dos dispositivos. Além disso, foi constatada uma diminuição expressiva da quantidade de dados utilizados em comparação com abordagens que operam no domínio temporal.

O restante deste artigo está organizado da seguinte forma: na Seção 2, detalhamos os trabalhos relacionados em que são destacadas as análises dos modelos de classificação; na Seção 3, explicamos os procedimentos metodológicos necessários para a realização deste trabalho; na Seção 4, comparamos e analisamos os resultados de desempenho de cada algoritmo e, por fim, na Seção 5 trazemos as considerações finais em torno das principais descobertas.

2 Trabalhos Relacionados

A literatura científica já conta com estudos que propõem abordagens com o objetivo de reconhecer dispositivos elétricos com base nas assinaturas das cargas presentes nos sinais elétricos, como tensão, corrente e potência. Nesta seção, serão apresentados os trabalhos que abordam métodos de identificação de dispositivos elétricos por meio do uso de modelos de aprendizado de máquina. Além disso, esses estudos avaliam o desempenho dos algoritmos por meio de métricas como acurácia e F1-score.

Partamos do trabalho de Nuran et al. (2023). O estudo

propõe o uso do método NILM (*Nonintrusive load monitoring*) para desagregar o consumo total de energia elétrica e identificar o uso individual de cada carga com base nas assinaturas características dos dispositivos. Como principal contribuição, os autores implementam o algoritmo *Random Forest* para identificar os tipos de aparelhos e o comparam com outros algoritmos supervisionados, como *SVM* (*Support Vector Machine*), *MLP* (*Multilayer Perceptron*), *kNN* (*k-Nearest Neighbors*) e *Naive Bayes*. Os experimentos foram realizados com dados obtidos por meio de um sistema de medição monofásico aplicado a diferentes tipos de aparelhos elétricos, além da validação com o conjunto de dados público chamado *WHITED*. O método proposto alcançou mais de 90% de acurácia na identificação dos dispositivos em ambos os conjuntos de dados, superando os demais algoritmos avaliados. Embora o estudo apresente bons resultados, destaca-se que as variáveis utilizadas representam apenas valores agregados dos sinais, como o valor eficaz da corrente. Isso pode limitar a capacidade de discriminação em cenários mais complexos. Em contraste, este trabalho realiza a análise direta dos sinais no domínio da frequência, permitindo a extração de padrões harmônicos e oferecendo maior robustez na identificação dos dispositivos.

Shahid et al. (2024) propõem um sistema NILM para residências inteligentes baseado em medidores de energia. O sistema coleta informações elétricas essenciais – como tensão, corrente, potência ativa, potência reativa, potência aparente, fator de potência e frequência – tanto para dispositivos isolados quanto para suas combinações. A principal inovação reside na aplicação de técnicas de aprendizado de máquina, com ênfase em RNAs (Redes Neurais Artificiais) e RNP's (Redes Neurais Profundas), sendo esta última a que apresentou melhor desempenho em condições de uso em tempo real. Os modelos propostos alcançaram entre 80% e 85% de acurácia na identificação dos dispositivos. Em suma, o estudo demonstra a eficácia das técnicas de aprendizado profundo na identificação de cargas elétricas, especialmente em ambientes domésticos inteligentes. Em contrapartida, nosso trabalho adota uma abordagem voltada à eficiência computacional para contextos de IoT, explorando algoritmos de classificação mais leves e priorizando a análise dos sinais no domínio da frequência. Essa abordagem permite extrair informações relevantes mesmo com volumes de dados reduzidos.

Nos estudos de Chou et al. (2022), detalha-se a identificação de dispositivos elétricos por meio de uma assinatura de multirresolução conhecida como *TAR*. A assinatura *TAR* é calculada a partir das trajetórias tensão-corrente dos sinais em regiões de estado estacionário. A correção de fase de *Fourier* é empregada para melhorar o desempenho do classificador. O modelo de classificação *kNN* é então aplicado, obtendo 97,43% de acurácia com *k* igual a 1. No entanto, o processamento de imagens representadas por matrizes multidimensionais pode apresentar desafios em ambientes com escassez de recursos. Em contrapartida, nossa pesquisa aborda a redução de dados para implementação em ambientes IoT.

Wu et al. (2019) abordam a implementação de um método de classificação multirrotulo com foco na identificação de cargas não intrusivas, utilizando o algoritmo de aprendizado *Random Forest*. O estudo compara vários mo-

Tabela 1: Comparação entre os trabalhos relacionados e o proposto.

Trabalho	Classificadores	Variáveis	Acurácia	Outras métricas
Nuran et al. (2023)	Random Forest, SVM, MLP, kNN, Naive Bayes	Corrente elétrica em valor eficaz (RMS) e potência ativa	90% a 100%	Precisão, recall e F1-score
Shahid et al. (2024)	RNAs e RNPs	Tensão, corrente, potência ativa, potência reativa, potência aparente, fator de potência	80% a 85%	—
Chou et al. (2022)	kNN	Imagens 2D de trajetórias dos sinais de corrente e tensão	97%	Precisão, recall e F1-score
Wu et al. (2019)	Random Forest, MLP, SVM e kNN	Sinal de corrente no domínio do tempo	97% a 98%	F1-score
Massidda et al. (2020)	Várias RNAs	Imagens do sinal de potência no domínio do tempo	97% a 99%	Precisão, recall e F1-score
Este trabalho	kNN, SVM, Decision Tree, Random Forest, Gradient Boosting e MLP	Sinal de corrente no domínio da frequência	99%	F1-score

delos, como MLP, SVM e kNN, destacando a eficácia do *Random Forest*. A classificação multirrótulo é explorada para categorizar dados, identificando estados de operação de cargas a partir de sinais de corrente no domínio do tempo. Os experimentos começam em ambiente real e prosseguem com um conjunto de dados público. Os resultados alcançam 97% de acurácia e 98% de F1-score para o *Random Forest*. Nesse mesmo escopo, nosso estudo também avalia vários classificadores; contudo, aplica a classificação no domínio da frequência, com vistas à redução do tráfego de dados em ambientes de IoT.

O trabalho de Massidda et al. (2020) apresenta uma nova abordagem de monitoramento e desagregação de cargas elétricas usando NILM, com foco na identificação dos estados de ativação dos aparelhos. O método se baseia em redes neurais para emular a capacidade de reconhecimento de usuários experientes. Alcança altas taxas de acurácia (cerca de 99%) e F1-score (97%) com técnicas de aprendizado profundo adaptadas para baixa taxa de amostragem. Nosso trabalho prioriza modelos de classificação mais simples para uma possível aplicação em IoT e utiliza o domínio da frequência para reduzir o tempo de treinamento e avaliação.

A Tabela 1 compara este trabalho e os relacionados, incluindo os classificadores, variáveis utilizadas e métricas de avaliação. A acurácia alcançada varia de 90% a 100% em diferentes estudos, com destaque para este trabalho, atingindo 99% de acurácia. Os trabalhos empregam diversos classificadores, como kNN, *Random Forest*, MLP e outros tipos de RNAs (Redes Neurais Artificiais), além de abordagens variadas, incluindo análise de sinal de potência e corrente no domínio do tempo. Este trabalho aborda o domínio da frequência como principal diferencial, com vistas à diminuição da quantidade de dados. A avaliação é feita com métricas comuns, como precisão, recall e F1-score. No geral, esses estudos demonstram promissoras capacidades de classificação na identificação de dispositivos elétricos.

3 Metodologia

Nesta seção, abordamos as etapas necessárias para a realização deste estudo. A Fig. 1 apresenta uma visão geral do processo completo, incluindo processamento de sinais,

treinamento e execução de cada modelo de aprendizado. Na etapa de amostragem, foram coletadas amostras dos sinais de corrente de dispositivos elétricos domésticos, conectados individualmente a uma tomada. Utilizando um sensor não intrusivo e um osciloscópio, capturamos as amostras da corrente elétrica necessária para alimentar um certo equipamento doméstico. 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 ser usados para identificar os dispositivos elétricos durante a etapa de execução.

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.

3.1 Amostragem e análise dos dispositivos

A etapa de coleta de amostras desempenha um papel fundamental no desenvolvimento deste estudo, uma vez que é essencial para a análise dos padrões de corrente elétrica exigida pelos dispositivos e para a formação do conjunto de dados. Nessa etapa, utilizamos um sensor não intrusivo capaz de registrar amostras do fluxo de corrente elétrica consumida por cada dispositivo a ser monitorado. Optamos pelo sensor SCT-013-000 (YHDC, Beijing, China) com capacidade máxima de leitura de 100 A, o qual oferece uma integração descomplicada com conversores analógico-digitaes e microcontroladores populares. A leitura do sensor foi obtida por meio de um osciloscópio, que permite a análise e armazenamento das amostras do sinal de corrente elé-

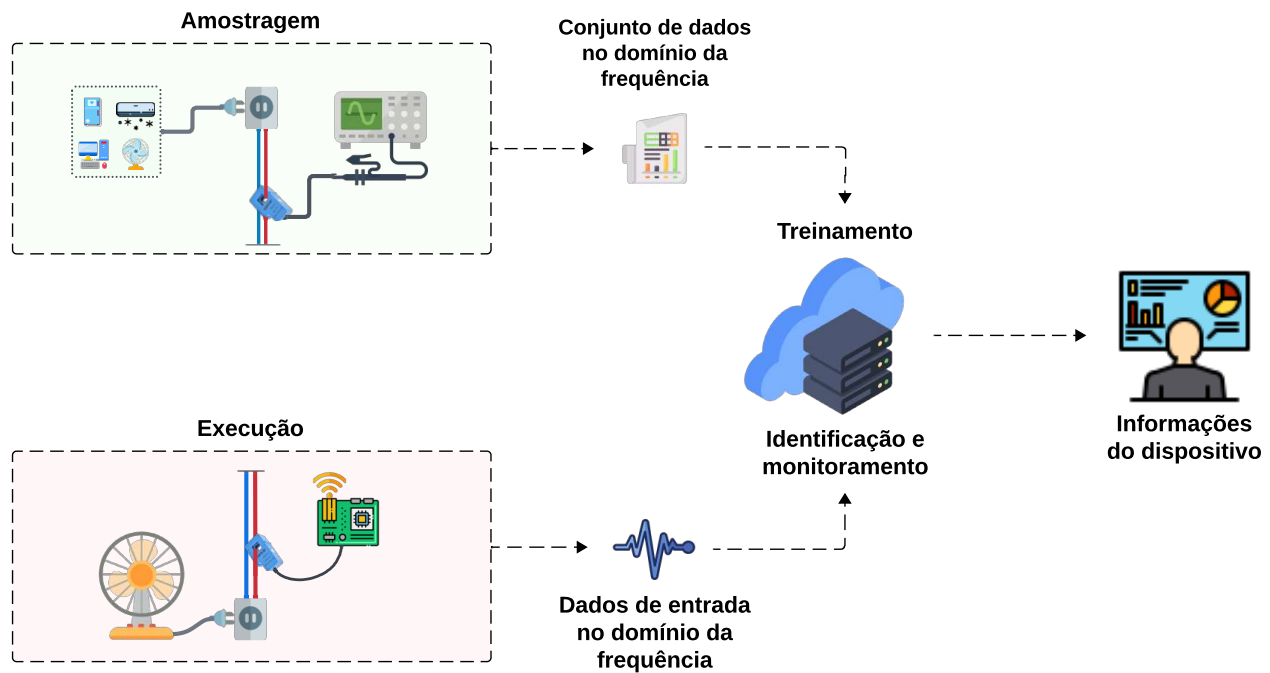


Figura 1: Visão geral do processo de monitoramento e identificação de dispositivos elétricos domésticos.

trica de cada dispositivo sob monitoramento. O esquema elétrico proposto para efetuar a amostragem está representado na fase de amostragem da Fig. 1.

Conforme ilustrado, o procedimento de coleta de amostras ocorre através do monitoramento do sinal de corrente requerido da rede de energia elétrica, à qual podem ser conectados diversos dispositivos com diferentes características elétricas¹. Um sensor não intrusivo é posicionado ao redor do cabo de fase para inferir o sinal de corrente elétrica, que é então lido através de um osciloscópio. O osciloscópio oferece recursos úteis adicionais, como ajuste da taxa de amostragem, cálculo da Transformada Rápida de Fourier (FFT), análise dos valores de amplitude e frequência, bem como a capacidade de salvar informações em diferentes formatos de arquivo.

Nesta etapa, efetuamos a identificação dos dispositivos elétricos, com base em seus sinais de corrente elétrica. Isso nos permite caracterizar a impedância a partir do padrão de onda e também explorar as componentes de frequência do sinal, por meio da análise de FFT.

3.2 Construção do conjunto de dados

O conjunto de dados proposto consiste em entradas individuais, cada uma contendo os 20 componentes iniciais de frequência do sinal, incluindo a frequência fundamental e seus harmônicos. Cada entrada também inclui uma eti-

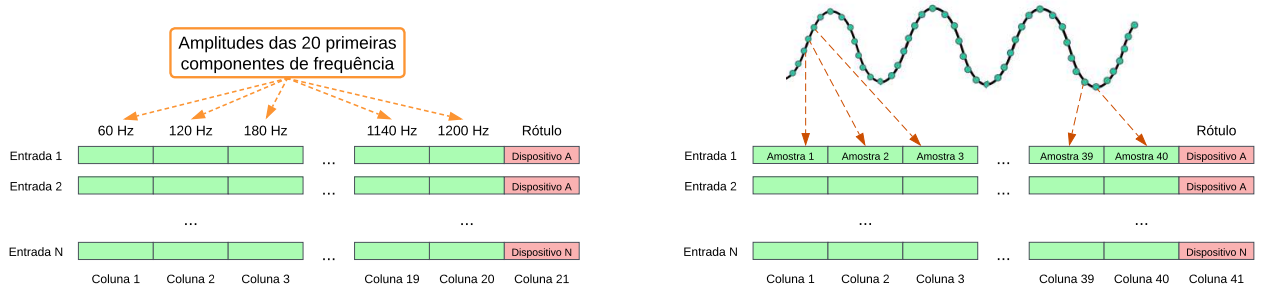
queta que identifica o dispositivo. O conjunto de dados é composto por 21 colunas, sendo a última coluna reservada para a etiqueta de identificação. Consequentemente, cada dispositivo é caracterizado por múltiplas entradas, cada uma representando as 20 primeiras componentes de frequência da corrente em diferentes instantes de tempo. A quantidade de entradas varia de acordo com o tamanho da amostra. A Fig. 2a ilustra a estrutura do conjunto de dados.

Esta abordagem permite a identificação do dispositivo com base nos níveis de amplitude dos componentes de frequência do sinal, resultando em um uso de armazenamento mais eficiente em comparação com técnicas calçadas no domínio do tempo, as quais requerem um grande volume de amostras para cobrir uma ampla faixa de períodos de sinal.

Para efeitos de comparação de resultados, sugerimos também a criação de um conjunto de dados que inclui amostras dos sinais de corrente no domínio do tempo. Esse conjunto consiste em amostras sequenciais do sinal de corrente no domínio do tempo, organizadas em 41 colunas. As 40 primeiras colunas representam amostras consecutivas do sinal de corrente no domínio do tempo, enquanto a última coluna contém a etiqueta que identifica o dispositivo correspondente. A Fig. 2b ilustra a estrutura do conjunto de dados no domínio do tempo.

Dessa forma, para cada dispositivo sob monitoramento, o conjunto de dados possui diversas entradas associadas a ele, sendo que cada entrada consiste em 40 amostras do sinal de corrente consumida em momentos específicos no tempo. Isso permite a identificação de cada dispositivo elétrico com base em características distintas presentes

¹Todas as medições foram feitas sobre uma tomada residencial de 10 A, em conformidade com a norma NBR14136 da ABNT.



(a) Componentes de frequência e os rótulos dos dispositivos. (b) Amostras no domínio do tempo e os rótulos dos dispositivos.

Figura 2: Estruturas do conjunto de dados nos domínios da frequência e do tempo.

em seu sinal de corrente, como valores máximos, mínimos e o formato da onda. A fim de promover uma comparação coerente, é mandatório que cada entrada nos conjuntos de dados no domínio da frequência e no domínio do tempo seja coletada com a mesma taxa de amostragem.

3.3 Treinamento e avaliação dos classificadores

Nesta fase são empregados diversos modelos de aprendizado de máquina, incluindo *kNN*, *SVM*, *Decision Tree*, *Random Forest*, *Gradient Boosting* e *MLP*, os quais são todos implementados em *Python* com o apoio das bibliotecas *Numpy*², *Pandas*³ e *Scikit-Learn*⁴. As bibliotecas *Numpy* e *Pandas* são comumente empregadas na manipulação de dados e foram utilizadas para esse fim. A biblioteca *Scikit-Learn*, de código aberto, disponibiliza funcionalidades úteis ao aprendizado de máquina, tais como regressão e classificação, bem como métricas incluindo acurácia, precisão, *recall* e F1-score.

Destas métricas, selecionamos o F1-score e a *acurácia* para a avaliação da eficácia da identificação realizada, que podem ser expressas, respectivamente, como Eq. (1) e Eq. (2), em que VP se refere ao número de “verdadeiros positivos” nas classificações, VN significa número de “verdadeiros negativos”, FP é o número de “falsos positivos” e FN de “falsos negativos”.

$$F1\text{-score} = \frac{VP}{VP + \frac{1}{2}(FN + FP)} \quad (1)$$

$$Acurácia = \frac{VP + VN}{VP + VN + FP + FN} \quad (2)$$

O F1-score combina precisão e *recall* em um único valor. A precisão refere-se à habilidade do modelo de evitar classificar equivocadamente exemplos negativos como po-

sitivos, enquanto o *recall* mensura a capacidade do modelo em identificar corretamente todos os exemplos relevantes na classe positiva.

Cada classificador foi ajustado de modo a encontrar a combinação ótima (ou, ao menos, subótima) de parâmetros, com o intuito de alcançar resultados satisfatórios nessas métricas. Com essa finalidade, a biblioteca *Scikit-Learn* foi utilizada para testar diversas configurações de hiperparâmetros e realizar validação cruzada, subdividindo o conjunto de dados em subconjuntos destinados ao treinamento e teste, permitindo, assim, a comparação dos desempenhos dos classificadores. O propósito é identificar o modelo mais eficiente no âmbito desta tarefa de classificação. Os conjuntos de dados e os códigos-fonte⁵ criados estão disponíveis à comunidade científica.

3.4 Monitoramento dos dispositivos elétricos

Para acompanhar o bom funcionamento dos dispositivos elétricos, adotamos uma abordagem sistemática que envolveu várias etapas. Primeiramente, identificamos os dispositivos por meio de técnicas de classificação utilizando dados de corrente elétrica no domínio da frequência. Uma vez identificados, iniciamos o processo de monitoramento para verificar se os dispositivos estavam operando dentro dos padrões esperados.

Em seguida, focalizamos nossa atenção 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 substancial desta média poderá indicar avarias nos dispositivos.

Para estabelecer um critério de anomalias, adotamos um limiar de 30% em relação à média da amplitude da frequência fundamental. Caso a amplitude de um dispositivo exceda esse limiar, considera-se haver uma anomalia e o dispositivo fica marcado como suspeito de estar defeituoso. A escolha do limiar de 30% foi baseada em uma

²<https://numpy.org/>

³<https://pandas.pydata.org/>

⁴<https://scikit-learn.org/stable/>

⁵https://github.com/davidamurim7/codigos_dissertacao

análise empírica dos valores disponíveis no conjunto de dados. Inicialmente, observamos 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 adoção inicial do limiar de 30%, sob o pressuposto de 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 referência 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 à medida que novas informações e critérios mais específicos são considerados.

Após a identificação de dispositivos com potenciais problemas, análises mais detalhadas podem ser conduzidas para compreender a natureza das anomalias. Isso pode envolver o exame 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 determinar quais ações corretivas se fazem necessárias.

4 Avaliação

Os procedimentos começaram com a etapa de coleta de amostras. Foram realizados testes iniciais nas salas e laboratórios no Campus de Quixadá da Universidade Federal do Ceará (UFC). Escolhemos aparelhos domésticos muito comuns, a saber: forno micro-ondas, geladeira, ar-condicionado, ventilador e monitor. Todos os aparelhos operam com uma fase da rede elétrica de 220 V e 60 Hz.

Primeiramente, o circuito de amostragem foi montado. Um osciloscópio da marca Instrutherm, modelo OD-265, com uma taxa máxima de amostragem de 10^9 amostras por segundo foi utilizado. Ele estava conectado a um sensor SCT-013-000 com capacidade máxima de leitura de 100 A. O sensor foi instalado no cabo de fase das tomadas onde os dispositivos elétricos estavam ligados.

Aproximadamente 40 000 amostras de cada dispositivo foram registradas e armazenadas em um arquivo CSV, a uma taxa de cerca de 2 500 amostras por segundo. Essa taxa de amostragem é suficiente para capturar as 20 primeiras componentes de frequência, de acordo com o Teorema de Nyquist (Zeng et al., 2024).

Em seguida, os dados foram processados e analisados no ambiente virtual Google Colaboratory. A Fig. 3 exemplifica as formas de onda da corrente do forno micro-ondas e da geladeira no domínio do tempo.

É importante destacar que o sensor SCT-013-000 não fornece valores em ampères, mas sim uma amplitude proporcional. As formas de onda apresentam características distintas, principalmente devido à presença de frequências harmônicas. Portanto, é possível inferir que os sinais de corrente são distintos no domínio do tempo, o que os torna passíveis de identificação por meio de modelos de classificação baseados em aprendizado de máquina.

No entanto, ressaltamos que a representação dos sinais no domínio do tempo pode exigir um número excessivo de amostras, tornando-se um problema em um ambiente com recursos limitados, como o da IoT. Nossa abordagem propõe um procedimento alternativo, realizando a conversão desses sinais para o domínio da frequência e explorando suas principais componentes.

A Fig. 4a mostra o sinal do forno micro-ondas convertido para o domínio da frequência por meio da FFT. É notável que a frequência fundamental de 60 Hz se destaca com a amplitude mais significativa. Também é possível observar a presença de outras componentes de frequência com amplitudes menores, o que explica a diferença no sinal no domínio do tempo em relação a uma senoide perfeita de 60 Hz.

O espectro de frequência do forno micro-ondas ainda apresenta características distintas quando comparado ao espectro da geladeira, como mostrado na Fig. 4b. Conforme observado, o sinal da geladeira exibe um número menor de frequências harmônicas, mostrando apenas dois picos de menor magnitude, localizados relativamente próximos à frequência fundamental.

Como observado, a presença de frequências harmônicas específicas e suas amplitudes é suficiente para identificar o dispositivo responsável pelo sinal. Nesse contexto, a identificação dos dispositivos com base no sinal de corrente no domínio da frequência se apresenta como uma abordagem promissora, reduzindo a quantidade de dados necessária para treinamento e classificação. É relevante destacar que os demais dispositivos avaliados nos experimentos passaram por análises semelhantes, permitindo também a diferenciação entre eles.

Após a fase de amostragem, os valores medidos foram usados para construir o conjunto de dados. Para cada dispositivo, as 40 000 amostras coletadas foram divididas em blocos de 100 amostras consecutivas, resultando em um total de 400 blocos. Em seguida, aplicou-se a FFT a cada bloco, ou seja, a cada segmento composto por 100 amostras. Dessa forma, cada bloco foi convertido 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 Transformada de Fourier tradicional (Sheu et al., 2024).

Posteriormente, cada espectro de frequência foi discretizado, mantendo apenas as amplitudes das 20 primeiras componentes de frequência, que incluem a frequência fundamental e 19 harmônicas, como ilustrado na Fig. 5. Nela, é possível visualizar as amplitudes das componentes de frequência presentes no sinal do forno micro-ondas em valores discretos.

Consequentemente, as amplitudes das 20 primeiras componentes de frequência constam nas colunas de recursos e a coluna de rótulo contém o nome do dispositivo, resultando em um total de 21 colunas e 400 linhas para cada dispositivo elétrico. Dessa forma, o conjunto de dados armazena um bloco de amostras, que pode ser consideravelmente grande, mas foi compactado em apenas 20 valores no domínio da frequência.

A título de comparação, também foi construído um conjunto de dados no domínio do tempo. As 40 000 amostras

foram divididas em blocos de 40 amostras consecutivas, resultando em um total de 1 000 blocos por dispositivo monitorado. Cada bloco de 40 amostras foi disposto ao longo das colunas de recursos; com a coluna de rótulo contendo o nome do dispositivo, temos um total de 41 colunas e 1 000 linhas associadas a cada dispositivo elétrico nesse conjunto de dados no domínio do tempo.

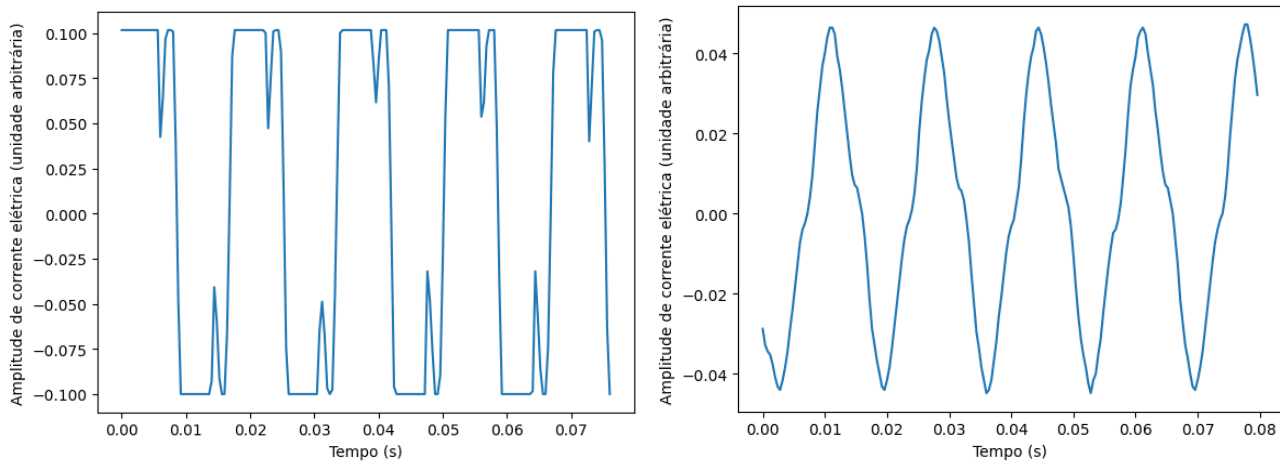
4.1 Resultados de identificação dos dispositivos

Inicialmente, os dados foram separados em conjuntos de treinamento e de teste, na proporção de 60% para treinamento e 40% para teste. Em seguida, utilizamos a biblioteca *Scikit-Learn* para realizar uma análise detalhada no conjunto de treinamento, explorando diversas combinações de hiperparâmetros para cada classificador. Utilizamos a técnica de validação cruzada, dividindo o conjunto em k subconjuntos. Cada combinação de hiperparâmetros

é testada k vezes; 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 certa combinação de hiperparâmetros é a média dos resultados das k iterações.

O parâmetro que determina o número de subconjuntos é chamado de k -fold, que neste caso adotamos o valor 10. Além disso, optamos por avaliar os resultados desses experimentos através da métrica F1-score.

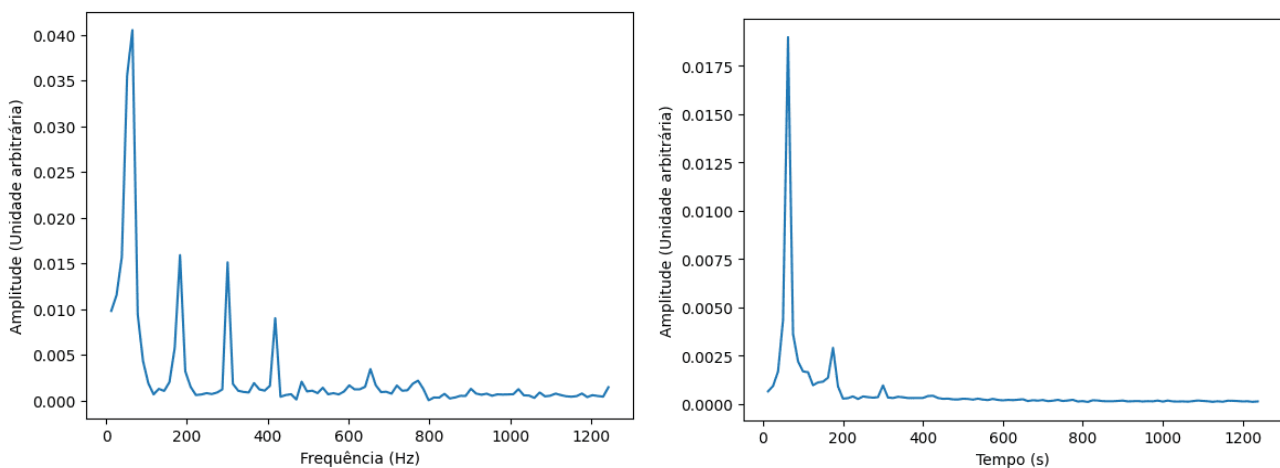
Ao término do experimento, conseguimos identificar uma combinação otimizada de hiperparâmetros para cada classificador. É crucial destacar que, nesse ponto, utilizamos apenas 60% dos dados, referentes ao conjunto de treinamento, que foram obtidos na primeira divisão. Em seguida, cada classificador otimizado foi avaliado com os 40% restantes dos dados, correspondentes ao conjunto de teste. Durante essa avaliação, aplicamos as métricas de acurácia e F1-score. Esses procedimentos foram conduzi-



(a) forno micro-ondas.

(b) geladeira.

Figura 3: Sinais de corrente no domínio do tempo construídos a partir das amostras capturadas.



(a) forno micro-ondas.

(b) geladeira.

Figura 4: Sinais de corrente no domínio da frequência construídos aplicando FFT sobre as amostras capturadas.

dos inicialmente no conjunto de dados principal, que opera no domínio da frequência. Posteriormente, repetimos essas etapas no conjunto de dados que utiliza informações temporais, visando a uma comparação dos resultados.

A Fig. 6 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%. Destaca-se, em particular, o *kNN*, que alcançou uma acurácia e um 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 *Decision Tree* obteve os resultados mais baixos, com 93% de acurácia e 92% de F1-score.

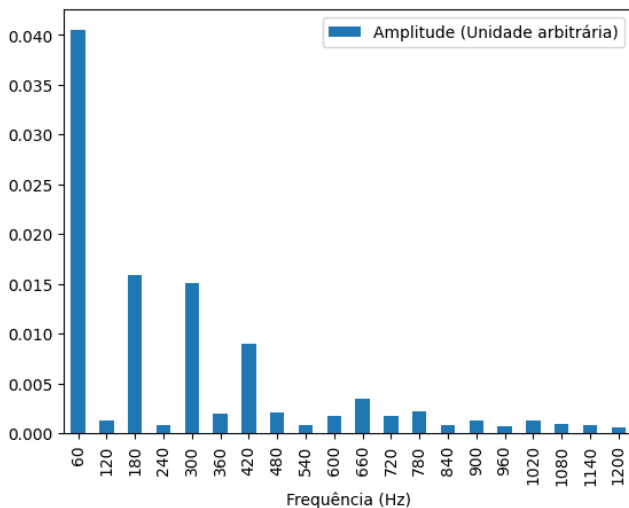


Figura 5: Amplitudes das 20 primeiras frequências do forno micro-ondas mediante a *FFT* das amostras capturadas.

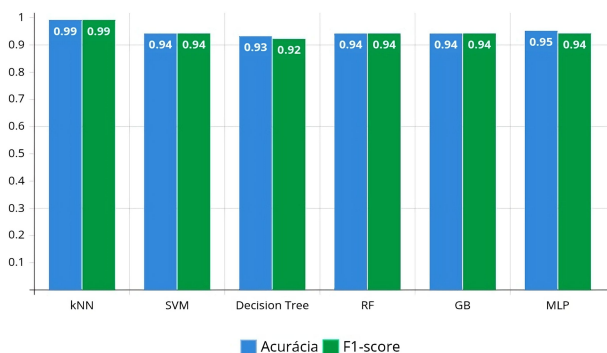


Figura 6: Resultados de acurácia e F1-score dos classificadores utilizando o conjunto de dados no domínio da frequência.

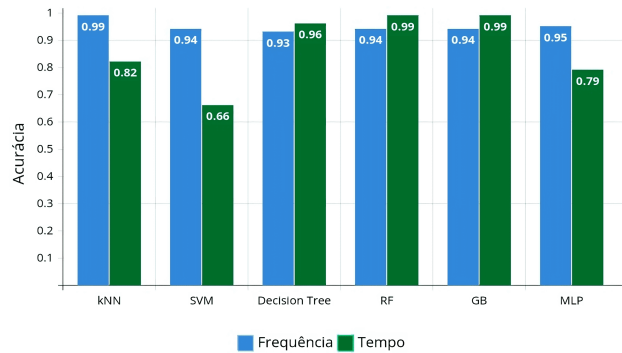


Figura 7: Resultados de acurácia dos classificadores ao empregar os conjuntos de dados nos domínios da frequência e do tempo.

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 componentes de 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 deste artigo não abranja a arquitetura de comunicação IoT, essas soluções são relevantes para viabilizar a implementação da solução. Em resumo, lidar com uma quantidade reduzida de dados oferece vantagens em diversos aspectos de um sistema IoT.

4.1.1 Comparação com domínio do tempo

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 Fig. 7 e Fig. 8 que apresentam uma análise comparativa com os resultados de acurácia e F1-score entre os domínios da frequência e do tempo.

Conforme destacado na Fig. 7, 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 (p.p.), respectivamente. No entanto, para os demais classificadores,

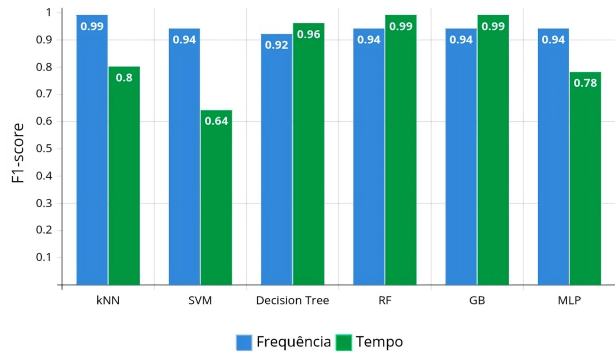


Figura 8: Pontuações F1-score dos classificadores ao empregar os conjuntos de dados nos domínios da frequência e do tempo.

observou-se acurácias mais altas no domínio do tempo. Essas discrepâncias, entretanto, foram mais modestas, totalizando apenas 5 p.p. nos classificadores *Random Forest* e *Gradient Boosting*, e 3 p.p. no classificador *Decision Tree*.

Conforme a Fig. 8, os resultados do F1-score para os classificadores *kNN*, *SVM* e *MLP* também foram superiores no domínio da frequência, apresentando diferenças significativas de 19, 30 e 16 p.p., 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 p.p. nos classificadores *Random Forest* e *Gradient Boosting*, e de 4 p.p. no classificador *Decision Tree*.

Em síntese, ao analisar os resultados nos domínios da frequência e do tempo, observamos 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.

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

Neste novo experimento, buscamos montar 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, treinamos os modelos 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 forno micro-ondas.

Após o treinamento inicial dos modelos, procedemos com o teste dos modelos utilizando um novo conjunto de dados. Esse conjunto soma os sinais de corrente dos dis-

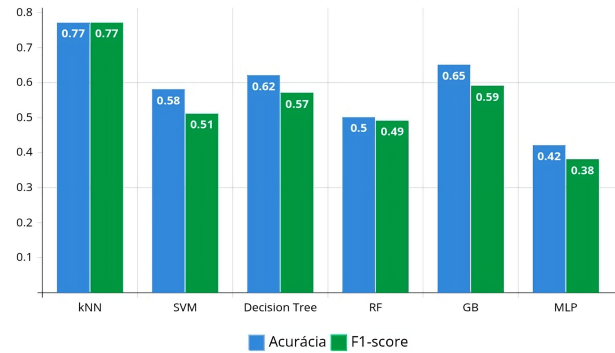


Figura 9: Desempenho dos modelos na classificação de dados de dispositivos operando simultaneamente com outros em ambientes ruidosos.

positivos 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, selecionamos novamente os classificadores *kNN*, *SVM*, *Decision Tree*, *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 Fig. 9.

Alcançamos 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.

O *Decision Tree*, *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.

Uma lâmpada de LED (*Light Emitting Diode*), 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 desigual, com algumas frequências mais intensas do que outras, em contraste com o ruído branco, que tem uma distribuição uniforme de energia 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 soma de sinais indesejados ao sinal desejado, conforme ilustrado na Eq. (3). 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.

$$i_{\text{ruído}}(t) = \sum_{k=1}^N B_k \sin(2\pi f_k t + \psi_k) \quad (3)$$

O sinal total observado é a combinação do sinal de corrente com as 20 primeiras componentes de frequência e os ruídos, conforme ilustram as Eq. (4) e Eq. (5). A soma de sinais é viável graças ao Princípio da Superposição, que afirma que os efeitos de múltiplas fontes de sinal podem ser combinados somando-se suas contribuições individuais (Barbi, 2021). Nesse contexto, podemos 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 aplicarmos a superposição, somamos os efeitos de ambas as fontes para obter o sinal total observado.

$$i_{\text{total}}(t) = i_A(t) + i_{\text{ruído}}(t) \quad (4)$$

$$i_A(t) = A_0 \sin(120\pi t) + \sum_{k=2}^{20} A_k \sin(120\pi k t + \phi_k) \quad (5)$$

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 Eq. (6).

$$i_{\text{filtrado}}(f) = H_{\text{passa-baixa}}(f) \cdot i_{\text{total}}(f) \quad (6)$$

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 causar ruído nas frequências das harmônicas do sinal desejado, essas frequências indesejadas podem ser expressas conforme a Eq. (7). 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.

$$i_{\text{ruído}}^k(t) = B_k \sin(120\pi k t + \psi_k) \quad (7)$$

Esses ruídos podem se somar ao sinal desejado, influenciando a amplitude das harmônicas. No entanto, 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, optamos 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%, utilizamos uma frequência de corte de 1300 Hz. Após experimentação, identificamos 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, utilizamos

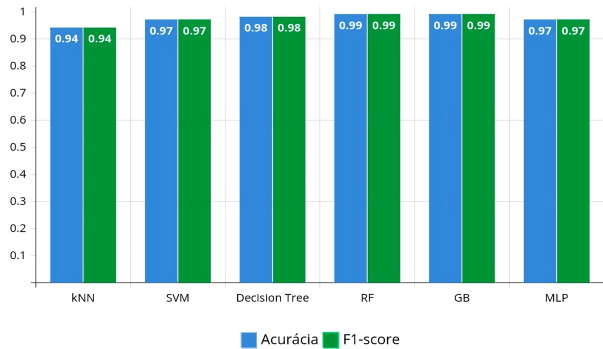


Figura 10: Desempenho dos modelos na classificação com dados filtrados de dispositivos operando simultaneamente com outros em ambientes ruidosos.

o filtro passa-baixa embutido no osciloscópio para processar os dados. A Fig. 10 ilustra o desempenho dos modelos na classificação utilizando dados de entrada filtrados.

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*, *Decision Tree* 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.

4.1.3 Classificação de dispositivos desconhecidos

Uma questão intrigante diz respeito ao comportamento dos modelos ao receberem informações de um dispositivo desconhecido, ou seja, sem treinamento prévio do referido dispositivo. Nesse contexto, conduzimos um experimento com o objetivo de mapear o desempenho. Para isso, excluímos todas as linhas relacionadas ao dispositivo de ar-condicionado do conjunto de dados de frequência. Em seguida, realizamos o treinamento do modelo apenas com os dispositivos restantes, a saber: geladeira, forno micro-ondas, ventilador e monitor. Ou seja, a partir de agora, nossos 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, procedemos ao 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

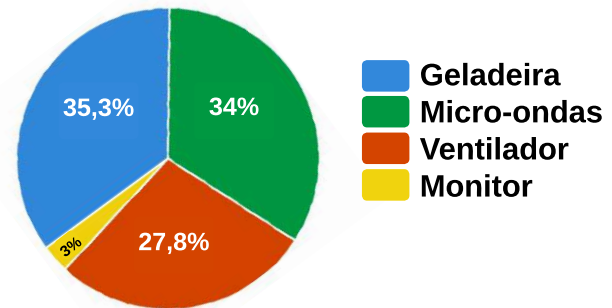


Figura 11: Desempenho do modelo *kNN* ao tentar reconhecer o dispositivo de ar-condicionado sem possuir conhecimento prévio sobre o referido aparelho.

no gráfico ilustrado na Fig. 11.

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* classificou como uma geladeira; em 34% como um forno 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 Fig. 12 traz um comparativo das amplitudes das componentes de frequência dos dispositivos abordados.

Como pode ser observado, a amplitude da frequência fundamental de 60 Hz foi crucial para a classificação do modelo. Isso ocorre porque as amplitudes da frequência de 60 Hz do forno 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 forno micro-ondas, justificando assim os resultados de classificação deste experimento.

Em resumo, podemos 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 seja adicionado ao conjunto de dados, com o objetivo de reduzir o erro causado por dispositivos desconhecidos.

4.1.4 Inclusão de novos dispositivos ao conjunto de dados

Outro aspecto muito importante é como os modelos se saem quando adicionamos amostras de novos dispositivos aos dados. Nesse contexto, realizamos um experimento para analisar os resultados de acurácia e F1-score à medida que novos dispositivos são incorporados ao conjunto.

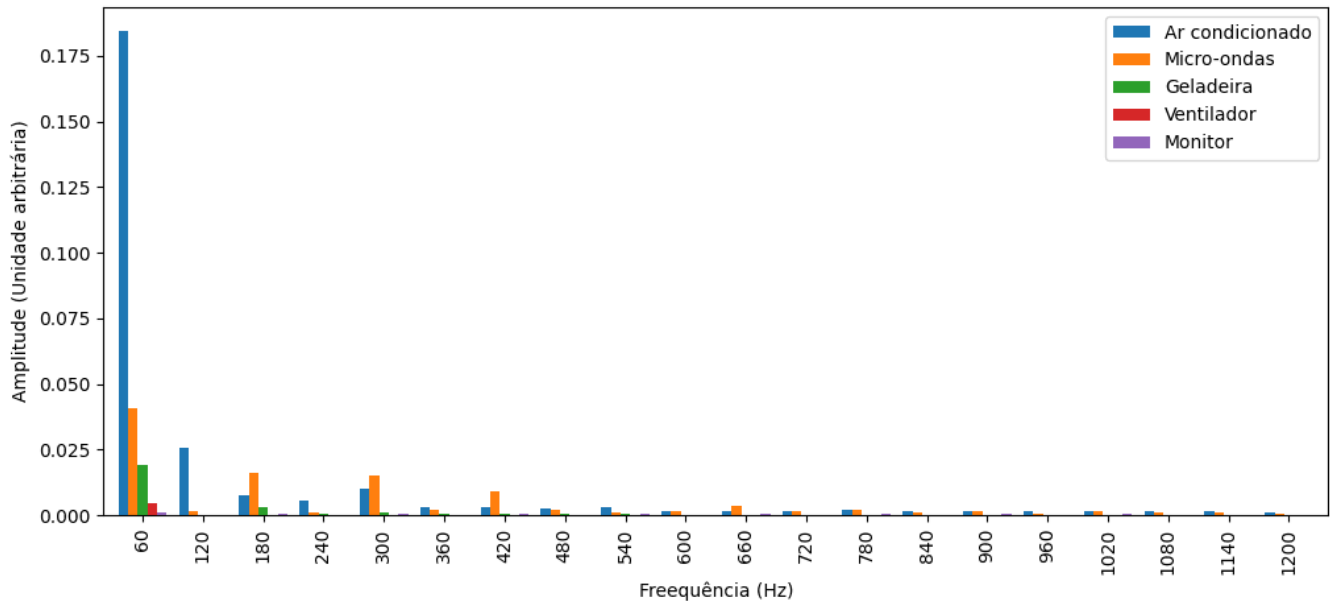


Figura 12: Comparativo das amplitudes das componentes de frequência dos dispositivos.

Inicialmente, apenas as entradas correspondentes a 2 dispositivos foram deixadas no conjunto de dados, nomeadamente o forno micro-ondas e a geladeira. Em seguida, foram adicionadas as entradas relacionadas ao aparelho de ar-condicionado. Posteriormente, incluímos 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 selecionamos os classificadores *kNN*, *Random Forest* e *MLP* devido às suas abordagens distintas de classificação, o que pode afetar os resultados à medida que incluímos novos dispositivos no conjunto de dados. A cada inclusão de um novo dispositivo, todos os classificadores passaram por testes de validação cruzada com *k-fold* igual a 10, visando ao reajuste de seus hiperparâmetros. Após encontrar a configuração ideal dos hiperparâmetros, procedemos ao treinamento dos modelos.

Como observado, todos os classificadores alcançaram

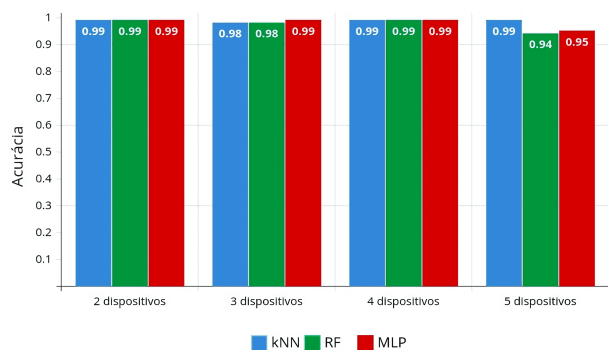


Figura 13: Resultados de acurácia dos classificadores com a inclusão de novos dispositivos no conjunto de dados.

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, observa-se que, após duas inclusões de amostras de novos dispositivos, os classificadores não apresentaram variações significativas nos resultados de acurácia.

No entanto, ao compararmos os resultados com 4 dispositivos aos resultados com 5 dispositivos, observamos 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 Fig. 14. Em resumo, não houve oscilações significativas nos resultados de acurácia e F1-score dos classificadores. Nessa perspectiva, concluímos 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.

Entretanto, ao analisarmos os resultados deste experimento e os resultados do experimento com dispositivos desconhecidos relatados na Seção 4.1.3, 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, optamos por considerar esses casos como

pertencentes ao mesmo dispositivo, não vendo a necessidade de uma classificação tão específica.

4.2 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 simples de monitoramento consiste em 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, realizamos um experimento prático utilizando dados reais provenientes de dispositivos elétricos, os quais estão armazenados em nosso conjunto de dados. No decorrer deste experimento, concentramos nossa 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 forno micro-ondas e geladeira pre-

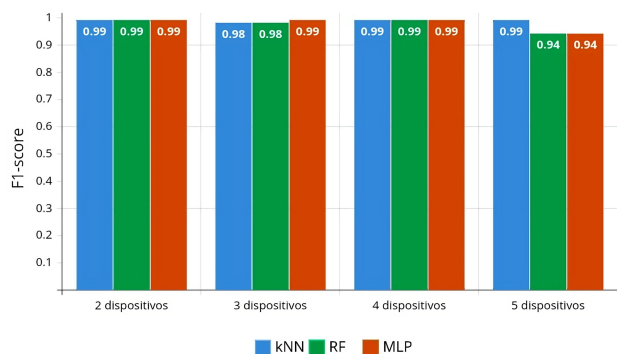


Figura 14: Resultados de F1-score dos classificadores com a inclusão de novos dispositivos no conjunto de dados.

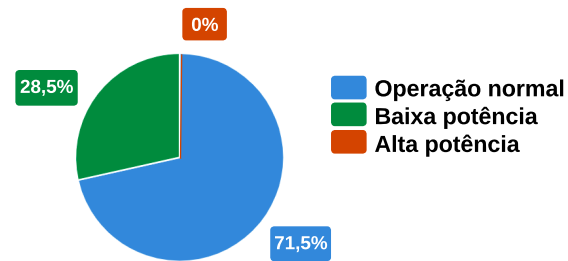


Figura 15: Percentual de entradas classificadas como operação normal, baixa potência e alta potência.

sentes no conjunto de dados. Essa média serviu como referência para avaliar o desempenho individual de cada dispositivo. Em seguida, criamos um conjunto de dados de teste com entradas dos dispositivos forno 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 Fig. 15 apresenta um gráfico dos resultados deste experimento. Após a análise das entradas dos dispositivos no conjunto de teste, observamos 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 não tenhamos encontrado nenhuma amostra com alta potência, a presença de um número significativo de entradas com baixa potência é um indicativo que deve acionar 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 de-

sempenho, 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 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.

5 Considerações Finais

Neste artigo, estudamos a classificação de dispositivos elétricos com base em sua assinatura de carga no domínio da frequência, utilizando algoritmos de aprendizado de máquina. A partir de medições realizadas em bancada, construímos um conjunto de dados abrangente que incorpora as amplitudes das 20 primeiras componentes de frequência do sinal de corrente elétrica, contemplando diversos dispositivos domésticos reais bastante populares. Destarte, sobre os dados reais, conduzimos simulações com diversos modelos de aprendizagem de máquinas atuando como classificadores, adotando a abordagem de validação cruzada, e avaliamos os resultados com base em métricas de acurácia e F1-score.

Alcançamos resultados promissores, com uma taxa de acerto de aproximadamente 99% ao empregar o modelo *k*NN. Além disso, conseguimos 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. O requisito de um número reduzido de dados para treinamento e classificação é crucial em ambientes IoT, sobretudo pela escassez ou raciocínio de recursos energéticos. Portanto, a abordagem no domínio da frequência desempenha um papel significativo nesse contexto. Este trabalho valida resultados de classificação com uma abordagem que requer menos dados do que o estado da arte.

Em comparação com outros trabalhos do estado da arte, este estudo se destaca pela utilização do domínio da frequência para a representação dos sinais de corrente. Essa abordagem resulta em uma redução significativa da quantidade de dados processados, o que a torna mais adequada para aplicações em ambientes com restrições computacionais, como sistemas embarcados e IoT. Enquanto muitos estudos utilizam sinais no domínio do tempo, imagens multidimensionais ou redes neurais complexas, que exigem maior poder de processamento, nossa abordagem utiliza classificadores mais leves, alcançando desempenho equivalente ou superior.

Uma limitação da nossa abordagem, no entanto, reside na sua priorização da simplicidade e viabilidade computacional,

o que pode resultar em um desempenho inferior para tarefas mais complexas que exigem contextualização temporal ou espacial profunda. Contudo, a proposta equilibra de forma eficaz desempenho, eficiência e aplicabilidade prática, consolidando-se como uma solução promissora em relação aos trabalhos existentes.

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 podem ser dados rumo à investigação da desagregação de cargas elétricas, permitindo a identificação de sinais de corrente provenientes de múltiplos dispositivos. Adicionalmente, almejamos aprimorar a identificação de dispositivos desconhecidos, visando a determinar se eles fazem parte do conjunto de dados ou não. Por fim, também visualizamos potencial na implementação de políticas que restrinjam equipamentos específicos em diferentes áreas da residência ou escritório, tais como limitações de quotas de consumo de energia, horários de utilização preferenciais e permitidos, e até funcionalidades de controle de permissões por usuário.

Referências

- Baharuddin, R. R., Niswar, M., Ilham, A. A. e Kashihara, S. (2022). Crab molting identification using machine learning classifiers, *2021 International Seminar on Machine Learning, Optimization, and Data Science (ISMOL)*, pp. 295–300. <https://doi.org/10.1109/ISMOL53584.2022.9743136>.
- Barbi, I. (2021). A theorem on power superposition in resistive networks, *IEEE Transactions on Circuits and Systems II: Express Briefs* 68(7): 2362–2363. <https://doi.org/10.1109/TCSII.2021.3059053>.
- Barron, L. (2021). The road to a smarter future: The smart city, connected cars and autonomous mobility, *2021 26th International Conference on Automation and Computing (ICAC)*, pp. 1–6. <https://doi.org/10.23919/ICAC50006.2021.9594125>.
- Chin, C. S. e Zhang, J. (2021). Wavelet scattering transform for multiclass support vector machines in audio devicess classification system, *2021 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*, pp. 735–740. <https://doi.org/10.1109/AIM46487.2021.9517547>.
- Chou, M., Thourn, K. e Chea, R. (2022). Multi-scale electrical appliance load signature for non-intrusive load monitoring classification, *2022 14th International Conference on Software, Knowledge, Information Management and Applications (SKIMA)*, pp. 79–83. <https://doi.org/10.1109/SKIMA57145.2022.10029485>.
- Gaikwad, N. B., Khare, S. K., Mendhe, D., Mir, H., Kosta, S. e Acharya, U. R. (2025). Fpga soc implementation of adaptive deep neural network based multimodal edge intelligence for internet of medical things, *IEEE Access* pp. 1–1. <https://doi.org/10.1109/ACCESS.2025.3592729>.

- Illakya, T., Keerthana, B., Murugan, K., Venkatesh, P., Manikandan, M. e Maran, K. (2024). The role of the internet of things in the telecom sector, *2024 International Conference on Communication, Computing and Internet of Things (IC3IoT)*, pp. 1–5. <https://doi.org/10.1109/IC3IoT60841.2024.10550390>.
- Inibhunu, C. e Carolyn McGregor, A. (2020). A privacy preserving framework for smart cities utilising iot, smart buildings and big data, *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)*, pp. 1096–1103. <https://doi.org/10.1109/HPCC-SmartCity-DSS50907.2020.00197>.
- Jadhav, T., Tanishq, Jagadish, S., Gaikwad, M., Wagle, S., Jayaswal, R. e Patil, S. (2023). Predicting urban land cover using classification: A machine learning approach, *2023 IEEE 11th Region 10 Humanitarian Technology Conference (R10-HTC)*, pp. 450–454. <https://doi.org/10.1109/R10-HTC57504.2023.10461930>.
- Kaloev, M. e Krastev, G. (2023). Tailored learning rates for reinforcement learning: A visual exploration and guideline formulation, *2023 7th International Symposium on Innovative Approaches in Smart Technologies (ISAS)*, pp. 1–7. <https://doi.org/10.1109/ISAS60782.2023.10391644>.
- Lu, Y., Ye, T. e Zheng, J. (2022). Decision tree algorithm in machine learning, *2022 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA)*, pp. 1014–1017. <https://doi.org/10.1109/AEECA55500.2022.9918857>.
- Massidda, L., Marrocu, M. e Manca, S. (2020). Non-intrusive load disaggregation by convolutional neural network and multilabel classification, *Applied Sciences* 10(4). <https://doi.org/10.3390/app10041454>.
- Meena, R. e Dubey, S. (2021). Smart houses with the application of energy management system & smart grid, *2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, pp. 947–952. <https://doi.org/10.1109/ICCCIS51004.2021.9397063>.
- Nuran, A. S., Murti, M. A. e Suratman, F. Y. (2023). Non-intrusive load monitoring method for appliance identification using random forest algorithm, *2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC)*, pp. 0754–0758. <https://doi.org/10.1109/CCWC57344.2023.10099248>.
- Palomsba, V., Brunaccini, G., Aloisio, D., Dino, G. E., Luna, M., La Tona, G., Carmela Di Piazza, M., Frazzica, A. e Sergi, F. (2022). A smart thermal management system for high efficiency data center cooling in microgrid environment, *2022 AEIT International Annual Conference (AEIT)*, pp. 1–6. <https://doi.org/10.23919/AEIT56783.2022.9951798>.
- Rebonatto, M. T., da Paixão, O. J. P. e Spalding, L. E. S. (2023). Detecção e alerta de equipamentos não permitidos em quartos hospitalares por meio da supervisão da corrente elétrica, *Revista Brasileira de Computação Aplicada* 15(2): 73–87. <https://doi.org/10.5335/rbca.v15i2.13920>.
- Richardson, C., Shah, S. e Yuan, X. (2022). Semi-supervised machine learning for analyzing covid-19 related twitter data for asian hate speech, *2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 1643–1648. <https://doi.org/10.1109/ICMLA55696.2022.00252>.
- Rocha, F., Dantas, L. C., Santos, L. F., Ferreira, S., Soares, B., Fernandes, A., Cavalcante, E. e Batista, T. (2020). Energy efficiency in smart buildings: An iot-based air conditioning control system, in A. Casaca, S. Katkooi, S. Ray e L. Strous (eds), *Internet of Things. A Confluence of Many Disciplines*, Springer International Publishing, Cham, pp. 21–35. https://doi.org/10.1007/978-3-030-43605-6_2.
- Shahid, K., Gulzar, M. M., Raza, M. A., Habib, S. e Ahmad, H. (2024). Machine learning based design of energy management system using non-intrusive load monitoring strategy, *2024 International Conference on Control, Automation and Diagnosis (ICCAD)*, pp. 1–6. <https://doi.org/10.1109/ICCAD60883.2024.10553879>.
- Sheu, M.-H., Yang, E.-C., Wang, S.-T., Juang, W.-H. e Lai, S.-C. (2024). 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* pp. 1–1. <https://doi.org/10.1109/TIM.2024.3457933>.
- Sljepcevic, D., Zeppelzauer, M., Unglaube, F., Kranzl, A., Breiteneder, C. e Horsak, B. (2023). Explainable machine learning in human gait analysis: A study on children with cerebral palsy, *IEEE Access* pp. 1–1. <https://doi.org/10.1109/ACCESS.2023.3289986>.
- Subramanian, S., Tseng, B., Barbieri, R. e Brown, E. N. (2021). Unsupervised machine learning methods for artifact removal in electrodermal activity, *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 399–402. <https://doi.org/10.1109/EMBC46164.2021.9630535>.
- Swarnkar, R., Harikrishnan, R., Thakur, P. e Singh, G. (2023). Electric vehicle lithium-ion battery ageing analysis under dynamic condition: A machine learning approach, *SAIEE Africa Research Journal* 114(1): 4–13. <https://doi.org/10.23919/SAIEE.2023.9962788>.
- Wu, X., Gao, Y. e Jiao, D. (2019). Multi-label classification based on random forest algorithm for non-intrusive load monitoring system, *Processes* 7(6). <https://doi.org/10.3390/pr7060337>.
- Yang, J., Kwon, Y. e Kim, D. (2021). Regional smart city development focus: The south korean national strategic smart city program, *IEEE Access* 9: 7193–7210. <https://doi.org/10.1109/ACCESS.2020.3047139>.

Zeng, Z., Liu, J. e Yuan, Y. (2024). A generalized nyquist-shannon sampling theorem using the koopman operator, *IEEE Transactions on Signal Processing* 72: 3595–3610.
<https://doi.org/10.1109/TSP.2024.3436610>.