



**UNIVERSIDADE FEDERAL DO CEARÁ**  
**CAMPUS QUIXADÁ**  
**CURSO DE GRADUAÇÃO EM ENGENHARIA DE COMPUTAÇÃO**

**FELIPE FEITOSA TEIXEIRA**

**APLICAÇÃO DE ALGORITMOS DE APRENDIZADO SUPERVISIONADO NA  
IDENTIFICAÇÃO DE PADRÕES ELÉTRICOS DE CORRENTE E TENSÃO**

**QUIXADÁ**

**2025**

FELIPE FEITOSA TEIXEIRA

APLICAÇÃO DE ALGORITMOS DE APRENDIZADO SUPERVISIONADO NA  
IDENTIFICAÇÃO DE PADRÕES ELÉTRICOS DE CORRENTE E TENSÃO

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Engenharia de Computação do Campus Quixadá da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de bacharel em Engenharia de Computação.

Orientador: Prof. Dr. Luis Rodolfo Rebouças Coutinho.

QUIXADÁ

2025

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)

---

- T266a Teixeira, Felipe Feitosa.  
Aplicação de algoritmos de aprendizado supervisionado na identificação de padrões elétricos de corrente e tensão / Felipe Feitosa Teixeira. – 2026.  
61 f. : il. color.
- Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Campus de Quixadá, Curso de Ciência da Computação, Quixadá, 2026.  
Orientação: Prof. Dr. Luis Rodolfo Rebouças Coutinho.
1. Aprendizado de Máquina. 2. Random Forest. 3. Wavelet. 4. Engenharia de Características. 5. Computação na Borda. I. Título.

CDD 004

---

FELIPE FEITOSA TEIXEIRA

APLICAÇÃO DE ALGORITMOS DE APRENDIZADO SUPERVISIONADO NA  
IDENTIFICAÇÃO DE PADRÕES ELÉTRICOS DE CORRENTE E TENSÃO

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Engenharia de Computação do Campus Quixadá da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de bacharel em Engenharia de Computação.

Aprovada em: 19/01/2026.

BANCA EXAMINADORA

---

Prof. Dr. Luis Rodolfo Rebouças  
Coutinho (Orientador)  
Universidade Federal do Ceará (UFC)

---

Prof. Dr. André Ribeiro Braga  
Universidade Federal do Ceará (UFC)

---

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

À minha família, por sempre acreditar em mim e caminhar ao meu lado em todos os momentos. À minha mãe, pelo cuidado, carinho e dedicação constantes, que me sustentaram mesmo nos dias mais difíceis. Ao meu pai, pelo apoio firme, e pela segurança que transmitiu, me lembrando de que nunca estive sozinho nessa jornada.

## **AGRADECIMENTOS**

Ao Prof. Dr. Luis Rodolfo Rebouças Coutinho, pela orientação atenciosa, pelas contribuições ao longo do desenvolvimento deste trabalho e pela confiança no meu potencial.

Aos professores da banca examinadora, pelos comentários, críticas construtivas e sugestões que enriqueceram este trabalho. À Universidade Federal do Ceará (UFC), pelo ensino de qualidade e pelo ambiente acadêmico que possibilitou minha formação. Aos meus familiares, pelo apoio incondicional durante toda a minha trajetória acadêmica. Aos amigos e colegas de curso, pelas trocas de experiências, pelo incentivo mútuo e pelos momentos de aprendizado compartilhado ao longo da graduação.

“Não há ventos favoráveis para quem não sabe  
onde quer chegar.” (SÊNECA, 65 d.C., p. 12.)

## RESUMO

O monitoramento não intrusivo de cargas consolidou-se como uma tecnologia chave para a eficiência energética em redes inteligentes, permitindo a desagregação do consumo sem a necessidade de múltiplos sensores. Este trabalho propõe e valida uma metodologia computacional para a identificação automática de cargas elétricas, fundamentada em técnicas de Ciência de Dados e Aprendizado de Máquina. Diferentemente de abordagens que dependem de coleta de dados em ambiente controlado, o estudo utiliza a unificação de duas bases de dados públicas heterogêneas (PLAID e COOLL), garantindo maior robustez e capacidade de generalização a ruídos. A estratégia de extração de características adota uma abordagem híbrida, combinando a análise espectral via Transformada Rápida de Fourier (FFT) e a decomposição tempo-frequência via Transformada Wavelet Discreta (DWT) com descritores estatísticos no domínio do tempo, como Fator de Crista e Curtose. Foram avaliados comparativamente três algoritmos de classificação supervisionada: *k-Nearest Neighbors* (k-NN), *Support Vector Machines* (SVM) e *Random Forest*. Os resultados experimentais, validados sobre um conjunto de teste independente, demonstraram a superioridade do modelo *Random Forest*, que alcançou um *F1-Score* de 0,95, superando as limitações de desbalanceamento de classes observadas nos demais modelos. Adicionalmente, a análise de custo computacional confirmou a viabilidade técnica teórica da implementação do algoritmo em sistemas embarcados de baixo custo (*Edge Computing*), validando a solução para futuras aplicações em dispositivos de IoT.

**Palavras-chave:** NILM; aprendizado de máquina; random forest; wavelet; engenharia de características; computação na borda.

## ABSTRACT

Non-Intrusive Load Monitoring (NILM) has established itself as a key technology for energy efficiency in smart grids, enabling consumption disaggregation without the need for multiple sensors. This work proposes and validates a computational methodology for the automatic identification of electrical loads, grounded in Data Science and Machine Learning techniques. Unlike approaches that rely on data collection in controlled environments, this study utilizes the unification of two heterogeneous public datasets (PLAID and COOLL), ensuring greater robustness and generalization capability against noise. The feature extraction strategy adopts a hybrid approach, combining spectral analysis via Fast Fourier Transform (FFT) and time-frequency decomposition via Discrete Wavelet Transform (DWT) with statistical descriptors in the time domain, such as Crest Factor and Kurtosis. Three supervised classification algorithms were comparatively evaluated: *k-Nearest Neighbors* (k-NN), *Support Vector Machines* (SVM), and *Random Forest*. The experimental results, validated on an independent test set, demonstrated the superiority of the *Random Forest* model, which achieved an F1-Score of 0.95, overcoming the class imbalance limitations observed in the other models. Additionally, the computational cost analysis confirmed the theoretical technical feasibility of implementing the algorithm in low-cost embedded systems (Edge Computing), validating the solution for future applications in IoT devices.

**Keywords:** NILM; machine learning; random forest; wavelet transform; feature engineering; edge computing.

## LISTA DE FIGURAS

Figura 1 – Conceito geral da técnica NILM: as cargas são identificadas por meio de uma única medição no ponto de entrada do sistema elétrico. . . . .	19
Figura 2 – Exemplo de forma de onda de corrente de um liquidificador (esquerda) e seu respectivo espectro de harmônicas de corrente, destacando as componentes ímpares (direita). . . . .	21
Figura 3 – Exemplo de sensores comumente utilizados para aquisição de sinais em NILM de baixo custo: (esq.) transformador de corrente TC; (dir.) sensor de tensão ZMPT101B. . . . .	23
Figura 4 – Exemplo ilustrativo de uma matriz de confusão em um problema de classificação com três classes. . . . .	31
Figura 5 – Preservação da morfologia do sinal após a reamostragem: (a) Sinal original em alta frequência; (b) Sinal padronizado em 30 kHz. . . . .	41
Figura 6 – Distribuição das classes no conjunto de dados final após a filtragem de relevância estatística. . . . .	43
Figura 7 – Comparação espectral entre carga não-linear e linear: (a) Laptop, apresentando múltiplos picos harmônicos; (b) Aquecedor, concentrando energia apenas na fundamental (60 Hz). . . . .	44
Figura 8 – Esquema da estratégia de validação cruzada com $k = 5$ iterações, onde os dados são rotacionados entre treino e teste para garantir a robustez estatística. . . . .	49
Figura 9 – Comparativo de desempenho final entre k-NN, SVM e Random Forest (com Wavelet), destacando a discrepância observada no modelo SVM. . . . .	52
Figura 10 – Matriz de confusão do modelo <i>Random Forest</i> no conjunto de teste, evidenciando as classes com maior taxa de confusão mútua. . . . .	53
Figura 11 – Importância das características no cenário base (apenas FFT e estatísticas temporais), mostrando dependência de métricas globais. . . . .	54
Figura 12 – Importância das características no cenário híbrido, destacando a predominância dos descritores Wavelet (cD2, cA5). . . . .	55

## LISTA DE TABELAS

Tabela 1 – Comparativo entre classificadores supervisionados aplicados em NILM . . .	29
Tabela 2 – Comparativo entre as arquiteturas de Computação em Nuvem e Computação na Borda . . . . .	33
Tabela 3 – Comparativo de metodologias entre os trabalhos relacionados e a proposta atual	37
Tabela 4 – Resumo das métricas de desempenho no conjunto de teste (Cenário Base: Apenas FFT e Tempo) . . . . .	51
Tabela 5 – Resumo das métricas de desempenho no conjunto de teste (Cenário Híbrido: Com Wavelet) . . . . .	52
Tabela 6 – Análise de viabilidade para sistemas embarcados: Complexidade e Desempenho	56

## LISTA DE ABREVIATURAS E SIGLAS

COOLL	<i>Controlled On/Off Loads Library</i>
DHT	Distorção Harmônica Total
DWT	<i>Discrete Wavelet Transform</i>
FFT	<i>Fast Fourier Transform</i>
FIR	<i>Finite Impulse Response</i>
IA	Inteligência Artificial
IIR	<i>Infinite Impulse Response</i>
IoT	<i>Internet of Things</i>
k-NN	<i>k-Nearest Neighbors</i>
LED	<i>Light Emitting Diode</i>
NILM	<i>Non-Intrusive Load Monitoring</i>
PCA	<i>Principal Component Analysis</i>
PDS	Processamento Digital de Sinais
PLAID	<i>Plug-Level Appliance Identification Dataset</i>
RMS	<i>Root Mean Square</i>
STFT	<i>Short-Time Fourier Transform</i>
SVM	<i>Support Vector Machines</i>

## LISTA DE SÍMBOLOS

$i(t)$	Corrente elétrica em função do tempo
$v(t)$	Tensão elétrica em função do tempo
$f_s$	Frequência de amostragem
$n$	Índice temporal discreto
$N$	Número total de amostras (ou número de classes)
$C$	Fator de Crista
$I_{RMS}$	Corrente eficaz ( <i>Root Mean Square</i> )
$I_{pico}$	Valor de pico da corrente
$K$	Curtose
$\bar{x}$	Média aritmética do sinal
$E_{cD}$	Energia dos coeficientes de detalhe (Wavelet)
$E_{cA}$	Energia dos coeficientes de aproximação (Wavelet)
$\sigma$	Desvio padrão
$TP$	Verdadeiro positivo ( <i>True Positive</i> )
$TN$	Verdadeiro negativo ( <i>True Negative</i> )
$FP$	Falso positivo ( <i>False Positive</i> )
$FN$	Falso negativo ( <i>False Negative</i> )
$k$	Número de vizinhos (k-NN) ou número de dobras ( <i>fold</i> s)

## SUMÁRIO

1	<b>INTRODUÇÃO</b>	16
1.1	<b>Motivação</b>	17
1.2	<b>Objetivo geral</b>	17
1.3	<b>Objetivos Específicos</b>	18
2	<b>FUNDAMENTAÇÃO TEÓRICA</b>	19
2.1	<b>Identificação de Cargas Elétricas e seu Papel em Sistemas Inteligentes</b>	19
2.1.1	<i>Sinais Elétricos e suas Propriedades Fundamentais</i>	20
2.1.2	<i>Digitalização e Pré-Processamento dos Sinais</i>	21
2.1.3	<i>Cargas Elétricas e Assinaturas Características</i>	22
2.2	<b>Extração e Representação de Assinaturas Elétricas</b>	23
2.2.1	<i>Análise Temporal e Morfologia do Sinal</i>	24
2.2.2	<i>Análise Espectral e Distorção Harmônica</i>	24
2.2.3	<i>Abordagens de Tempo-Frequência</i>	25
2.2.4	<i>Engenharia de Características (Feature Engineering)</i>	26
2.3	<b>Algoritmos de Classificação Supervisionada</b>	26
2.3.1	<i>k-Nearest Neighbors (k-NN)</i>	26
2.3.2	<i>Árvores de Decisão e Florestas Aleatórias</i>	27
2.3.3	<i>Máquinas de Vetores de Suporte (SVM)</i>	27
2.3.4	<i>Redes Neurais Artificiais</i>	28
2.3.5	<i>Comparativo e Escolha do Algoritmo</i>	28
2.4	<b>Métricas de Avaliação de Desempenho</b>	29
2.4.1	<i>Acurácia</i>	29
2.4.2	<i>Precisão e Revocação</i>	30
2.4.3	<i>F1-Score</i>	30
2.4.4	<i>Matriz de Confusão</i>	31
2.4.5	<i>Curvas ROC e AUC</i>	31
2.4.6	<i>Validação Cruzada e Overfitting</i>	31
2.4.7	<i>Interpretação e Limitações</i>	32
2.5	<b>Computação na Borda e TinyML</b>	32
3	<b>TRABALHOS RELACIONADOS</b>	34

3.1	<b>STFT e Kernel PCA para Identificação de Cargas</b> . . . . .	34
3.2	<b>Classificação de Cargas Elétricas Idênticas</b> . . . . .	34
3.3	<b>Monitoramento e Identificação de Cargas com ML</b> . . . . .	35
3.4	<b>Classificação de Cargas Elétricas em Tempo Real</b> . . . . .	35
3.5	<b>Machine Learning para Previsão de Carga</b> . . . . .	35
3.6	<b>Redes Neurais para Predição de Curvas de Carga</b> . . . . .	36
3.7	<b>Análise Comparativa</b> . . . . .	36
4	<b>METODOLOGIA</b> . . . . .	38
4.1	<b>Tipo de Pesquisa e Abordagem Experimental</b> . . . . .	38
4.2	<b>Seleção e Compreensão dos Dados</b> . . . . .	39
4.2.1	<i>Base de Dados PLAID</i> . . . . .	39
4.2.2	<i>Base de Dados COOLL</i> . . . . .	40
4.2.3	<i>Integração e Heterogeneidade dos Dados</i> . . . . .	40
4.3	<b>Preparação e Engenharia de Dados</b> . . . . .	40
4.3.1	<i>Integração e Reamostragem (Downsampling)</i> . . . . .	41
4.3.2	<i>Detecção de Eventos e Limpeza dos Dados</i> . . . . .	42
4.4	<b>Engenharia de Características (Feature Engineering)</b> . . . . .	43
4.4.1	<i>Análise Espectral e Decomposição Harmônica</i> . . . . .	43
4.4.2	<i>Descritores Estatísticos Temporais</i> . . . . .	44
4.4.3	<i>Análise Tempo-Frequência via Wavelet</i> . . . . .	45
4.5	<b>Modelagem e Treinamento Computacional</b> . . . . .	46
4.5.1	<i>Algoritmos de Classificação Selecionados</i> . . . . .	47
4.5.2	<i>Estratégia de Treinamento e Otimização</i> . . . . .	47
4.5.2.1	<i>Padronização dos Dados</i> . . . . .	47
4.5.2.2	<i>Protocolo de Validação Cruzada</i> . . . . .	48
4.6	<b>Avaliação de Desempenho</b> . . . . .	48
4.6.1	<i>Métricas de Classificação</i> . . . . .	49
4.6.2	<i>Matriz de Confusão</i> . . . . .	50
5	<b>RESULTADOS E DISCUSSÃO</b> . . . . .	51
5.1	<b>Desempenho Global dos Classificadores</b> . . . . .	51
5.2	<b>Análise Detalhada de Erros (Matriz de Confusão)</b> . . . . .	52
5.3	<b>Contribuição da Transformada Wavelet e Relevância das Características</b> . . . . .	54

<b>5.4</b>	<b>Viabilidade Computacional e Complexidade Algorítmica . . . . .</b>	<b>55</b>
<b>6</b>	<b>CONCLUSÃO . . . . .</b>	<b>57</b>
	<b>REFERÊNCIAS . . . . .</b>	<b>58</b>

## 1 INTRODUÇÃO

A busca pela eficiência energética tornou-se um pilar fundamental no desenvolvimento das chamadas *Smart Grids* e no contexto da *Industry 4.0*. A otimização de recursos e a redução de desperdícios exigem uma granularidade cada vez maior na medição do consumo. Tradicionalmente, isso demandaria a instalação de múltiplos sensores físicos em cada ponto de carga, uma abordagem financeiramente inviável e complexa. Como alternativa, surge o conceito de monitoramento não intrusivo de cargas, ou *Non-Intrusive Load Monitoring* (NILM), que propõe o monitoramento individualizado de equipamentos a partir de um único ponto de medição central, utilizando técnicas avançadas de análise de dados.

O desafio técnico do NILM reside na desagregação de cargas em ambientes ruidosos, onde dispositivos distintos podem apresentar potências ativas semelhantes. A simples leitura de consumo não é suficiente para diferenciar, por exemplo, cargas resistivas lineares de cargas eletrônicas não-lineares. Estas últimas introduzem distorções harmônicas significativas na rede elétrica, exigindo uma abordagem fundamentada em Processamento Digital de Sinais (PDS). A extração de características nos domínios da frequência, do tempo e do tempo-frequência se torna um pré-requisito para a identificação correta dos estados de operação.

Nesse cenário, a aplicação de algoritmos de aprendizado de máquina supervisionado se torna uma ferramenta robusta para o reconhecimento de padrões complexos. Ao correlacionar parâmetros estatísticos com a decomposição espectral via Transformada Rápida de Fourier (*Fast Fourier Transform* (FFT)) e a análise de transientes via Transformada Wavelet Discreta (*Discrete Wavelet Transform* (DWT)), é possível criar assinaturas elétricas únicas. Essa abordagem permite que o sistema classifique dispositivos com base na morfologia da onda de corrente e em seus comportamentos transientes, superando as limitações dos medidores convencionais.

Este trabalho propõe o desenvolvimento e validação de um modelo computacional para a classificação automática de cargas, utilizando bases de dados reais e heterogêneas, como o *Plug-Level Appliance Identification Dataset* (PLAID) e a *Controlled On/Off Loads Library* (COOLL). A metodologia prioriza uma engenharia de características (*feature engineering*) rigorosa, visando não apenas a precisão do modelo, mas também a sua viabilidade de implementação em sistemas embarcados de baixo custo. Além de demonstrar que, a partir de um tratamento matemático adequado dos sinais elétricos e do uso de classificadores eficientes, é possível processar essas informações na borda (*Edge Computing*). Isso viabiliza soluções de *Internet of Things* (IoT) acessíveis e escaláveis, capazes de operar em *hardware* com recursos limitados,

contribuindo para a modernização do monitoramento energético residencial e industrial.

## 1.1 Motivação

A busca por maior eficiência no uso da energia elétrica tem impulsionado o desenvolvimento de tecnologias inteligentes de monitoramento, fundamentais para a consolidação das *Smart Grids*. Entre essas tecnologias, destaca-se o NILM, técnica que visa identificar cargas elétricas individuais a partir de medições agregadas de corrente e tensão, realizadas em um único ponto da instalação elétrica, eliminando a necessidade de sensoriamento distribuído intrusivo.

Apesar de sua aplicabilidade promissora, o NILM enfrenta desafios significativos em cenários reais. A crescente presença de cargas não-lineares, como eletrônicos e iluminação *Light Emitting Diode* (LED), introduz distorções harmônicas na rede que tornam as assinaturas elétricas de diferentes dispositivos muito semelhantes. A variabilidade nas condições de operação e os ruídos inerentes à aquisição de dados dificultam a correta distinção das cargas apenas por potência ativa. Isso exige a aplicação de técnicas robustas de extração de características que vão além da filtragem básica; torna-se fundamental interpretar a morfologia do sinal nos domínios da frequência e tempo-frequência para capturar as nuances que diferenciam, por exemplo, cargas resistivas de fontes chaveadas.

Nesse cenário, surge a motivação deste trabalho: investigar o uso conjunto de PDS e aprendizado de máquina supervisionado como base metodológica para sistemas de identificação de cargas de alto desempenho. A proposta consiste em construir um modelo que extraia, dos sinais brutos de corrente, assinaturas discriminativas utilizando ferramentas como a FFT, a DWT e a análise estatística temporal, permitindo sua classificação por algoritmos eficientes. Essa abordagem alinha-se à necessidade de processamento na borda (*Edge Computing*), buscando viabilizar a implementação futura de algoritmos inteligentes em *hardware* de baixo custo e recursos limitados, contribuindo para a democratização de soluções IoT de monitoramento energético.

## 1.2 Objetivo geral

Neste sentido, este trabalho tem por objetivo Avaliar comparativamente o desempenho de algoritmos de aprendizado supervisionado na identificação de cargas elétricas não-intrusiva (NILM), baseada na análise espectral, temporal e tempo-frequência de sinais de corrente.

O estudo foca na aplicação conjunta de técnicas de PDS, notadamente a FFT e a DWT, e algoritmos de aprendizado de máquina supervisionado, visando a classificação precisa de dispositivos lineares e não-lineares, bem como a verificação da viabilidade técnica para futura implementação em sistemas embarcados.

### 1.3 Objetivos Específicos

- Integrar as bases de dados públicas PLAID e COOLL, padronizando suas taxas de amostragem, para compor um cenário de teste heterogêneo e desafiador.
- Investigar a contribuição da Transformada Wavelet Discreta (DWT) na extração de características, avaliando sua eficácia em comparação a abordagens puramente espectrais (FFT) e temporais.
- Avaliar comparativamente o desempenho dos classificadores supervisionados *k-Nearest Neighbors* (k-NN), *Support Vector Machines* (SVM) e *Random Forest*, identificando o modelo mais robusto para lidar com o desbalanceamento de classes.
- Analisar a complexidade computacional e a importância das características extraídas, visando validar a viabilidade teórica de aplicação em sistemas de computação na borda (*Edge Computing*).

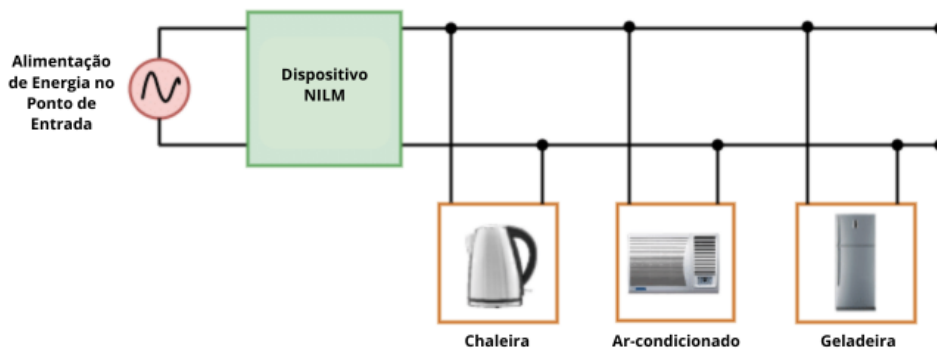
## 2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta os conceitos fundamentais que sustentam o desenvolvimento deste trabalho, com foco na identificação de cargas elétricas a partir da análise de sinais de corrente e tensão. Inicialmente, aborda-se o contexto da eficiência energética e a motivação para o monitoramento do consumo individual de dispositivos, diferenciando as estratégias de medição local e geral. Em seguida, são discutidas as assinaturas elétricas como representação característica dos dispositivos e os fundamentos de PDS aplicados à extração dessas informações. Por fim, são abordados os principais algoritmos de classificação supervisionada e métricas de avaliação utilizadas neste tipo de tarefa.

### 2.1 Identificação de Cargas Elétricas e seu Papel em Sistemas Inteligentes

Com o avanço das tecnologias digitais e a crescente demanda por soluções sustentáveis, tornou-se essencial desenvolver sistemas de monitoramento energético capazes de identificar, com precisão, os dispositivos em uso dentro de uma instalação elétrica, conforme representado na Figura 1. A identificação precisa de cargas não apenas viabiliza estratégias de eficiência energética, como também permite detectar padrões de uso, realizar manutenções preditivas e integrar mecanismos de controle automatizado, aspectos fundamentais para a consolidação de ambientes conectados em redes elétricas inteligentes (Shibata *et al.*, 2021).

Figura 1 – Conceito geral da técnica NILM: as cargas são identificadas por meio de uma única medição no ponto de entrada do sistema elétrico.



Fonte: Gopinath *et al.* (2020, p. 5).

Para implementar tal monitoramento, a literatura classifica as abordagens em duas categorias principais, baseadas na arquitetura de medição adotada:

1. **Medição Intrusiva (ou Medição Local):** Consiste na instalação de sensores dedicados

em cada ponto de consumo, como tomadas inteligentes ou medidores individuais para cada eletrodoméstico. Embora ofereça dados diretos e de alta precisão, essa abordagem apresenta custo elevado de *hardware* e manutenção, além de ser complexa para implementar em instalações já existentes.

2. **Medição Não Intrusiva (ou Medição Geral):** Baseia-se em um único ponto de medição centralizado (geralmente no quadro de distribuição ou medidor de entrada). A identificação do consumo individual é realizada via *software*, por meio da desagregação matemática do sinal total medido.

Este trabalho insere-se no contexto da segunda abordagem, conhecida como NILM (*Non-Intrusive Load Monitoring*). Essa técnica busca identificar os dispositivos presentes no sistema a partir da análise inteligente dos sinais agregados de corrente e tensão, empregando técnicas de PDS e algoritmos de aprendizado de máquina para inferir quais equipamentos estão em operação sem a necessidade de invadir a instalação interna da residência (Hart, 1992).

O NILM é considerado uma abordagem promissora devido à sua viabilidade econômica e flexibilidade de aplicação. Trabalhos como os de Pereira (2017) e Paixao *et al.* (2016) demonstram a eficácia de metodologias que combinam PDS com técnicas de aprendizado de máquina na identificação de padrões específicos de funcionamento de cargas. Nesse sentido, Bouhouras *et al.* (2012) mostraram ser possível diferenciar até mesmo lâmpadas fluorescentes compactas por meio da análise criteriosa dos sinais elétricos coletados.

O primeiro passo para a aplicação dessas técnicas é compreender que cada dispositivo imprime uma *assinatura elétrica* única no sinal agregado ao ser ativado. Essas assinaturas podem ser analisadas tanto no domínio do tempo quanto nos domínios da frequência e tempo-frequência, e constituem a base para os sistemas de classificação automáticos que integram soluções modernas de NILM. Os tópicos a seguir exploram os fundamentos físicos, matemáticos e computacionais necessários para a extração, digitalização e modelagem dessas assinaturas.

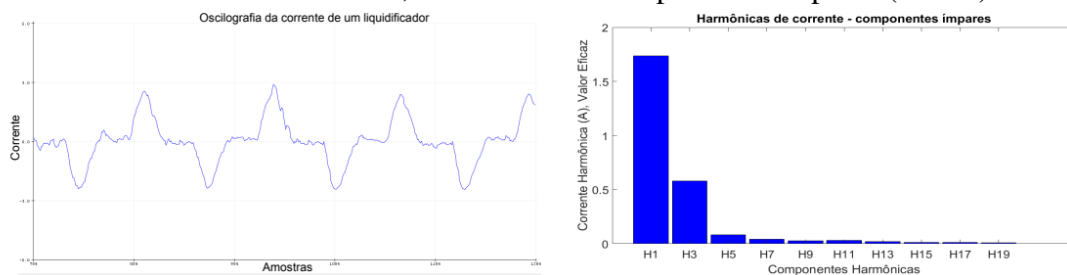
### 2.1.1 *Sinais Elétricos e suas Propriedades Fundamentais*

Os sinais elétricos de corrente  $i(t)$  e tensão  $v(t)$  são funções do tempo que expressam, respectivamente, o fluxo de cargas elétricas e a diferença de potencial em um circuito. Em sistemas de corrente alternada (CA), amplamente utilizados em ambientes residenciais e industriais, esses sinais se apresentam de forma periódica, geralmente senoidal, com frequência de 50 Hz ou 60 Hz, dependendo da localidade. Essa característica cíclica possibilita a análise detalhada tanto

no domínio do tempo quanto no domínio da frequência (Oppenheim *et al.*, 1999).

A relação entre corrente e tensão depende do tipo de carga conectada ao circuito. Cargas resistivas, como aquecedores e lâmpadas incandescentes, apresentam corrente e tensão em fase. Já cargas indutivas, como motores e transformadores, introduzem defasagem positiva entre as grandezas. Cargas capacitivas, por sua vez, tendem a apresentar defasagem negativa e, em certos casos, geram distorções que afetam a forma de onda do sinal medido. Na Figura 2 é ilustrado um exemplo prático desse comportamento.

Figura 2 – Exemplo de forma de onda de corrente de um liquidificador (esquerda) e seu respectivo espectro de harmônicas de corrente, destacando as componentes ímpares (direita).



Fonte: (Shibata *et al.*, 2021)

À esquerda, observa-se a forma de onda de corrente de um liquidificador durante sua operação, apresentando distorções típicas de cargas não-lineares. À direita, o espectro de harmônicas de corrente evidencia a presença significativa de componentes ímpares, que representam assinaturas características utilizadas em sistemas de identificação de cargas.

### 2.1.2 Digitalização e Pré-Processamento dos Sinais

Para que os sinais analógicos de corrente e tensão sejam processados digitalmente, é necessário submetê-los a um processo de amostragem e quantização. Segundo o Teorema de Nyquist-Shannon, a frequência de amostragem deve ser pelo menos o dobro da maior frequência presente no sinal, de modo a evitar *aliasing* e preservar as informações relevantes (Oppenheim *et al.*, 1999). Trabalhos como os de Gritli *et al.* (2017) e Bouhouras *et al.* (2019) relatam que taxas de amostragem superiores a 3,7 kHz são recomendadas para garantir a captura precisa de componentes harmônicas até a 31ª ordem da frequência fundamental de 60 Hz, faixa onde residem as principais características das cargas não-lineares.

Após a digitalização, o pré-processamento dos sinais desempenha papel fundamental na padronização dos dados e na redução da dimensionalidade. Entre as etapas comuns destacam-se:

- **Normalização:** ajuste da amplitude dos sinais para um intervalo padronizado, eliminando variações decorrentes da magnitude da tensão da rede;
- **Segmentação e Detecção de Eventos:** divisão dos dados em janelas temporais baseadas nos instantes de acionamento (*event detection*), isolando o regime permanente da carga;
- **Reamostragem (*Downsampling*):** adequação da taxa de amostragem para uniformizar bases de dados heterogêneas e reduzir o custo computacional, sem perda significativa de informação espectral relevante.

Essas etapas garantem que os algoritmos de extração de características operem sobre dados consistentes. Vale ressaltar que, embora a aplicação de filtros digitais *Finite Impulse Response* (FIR) e *Infinite Impulse Response* (IIR) seja tradicionalmente utilizada para a remoção de ruídos (Rodríguez-Silva; Makonin, 2019), no contexto de identificação de cargas modernas deve-se ter cautela. Filtros passa-baixa muito agressivos podem atenuar inadvertidamente componentes de alta frequência e transientes rápidos que compõem a assinatura singular de dispositivos eletrônicos. Portanto, o pré-processamento deve priorizar a preservação da morfologia do sinal e de seu conteúdo harmônico, em detrimento de uma suavização excessiva.

### 2.1.3 Cargas Elétricas e Assinaturas Características

Cada tipo de carga elétrica imprime um padrão específico nos sinais elétricos capturados, o que é conhecido como *assinatura elétrica*. Esses padrões são determinados por fatores como a defasagem entre corrente e tensão, o valor médio de consumo, a distorção harmônica gerada e os comportamentos transientes de partida e desligamento.

Cargas resistivas apresentam assinaturas regulares e de baixa complexidade espectral. Já cargas indutivas, como motores de geladeiras ou ventiladores, tendem a produzir picos de corrente no momento da partida, com componentes de frequência mais altas. Cargas capacitivas, por sua vez, podem introduzir perturbações específicas no sinal de tensão.

Alguns estudos mostraram que até mesmo dispositivos eletronicamente semelhantes, como múltiplas lâmpadas fluorescentes, podem apresentar assinaturas suficientemente distintas para fins de classificação. Essas assinaturas são compostas por atributos como valor RMS, potência ativa e reativa, fator de potência e, quando necessário, componentes extraídas por meio de transformadas como a FFT, *Short-Time Fourier Transform* (STFT) ou DWT. Essas representações vetoriais constituem a base dos algoritmos de aprendizado supervisionado e não supervisionado, que serão discutidos em seções posteriores.

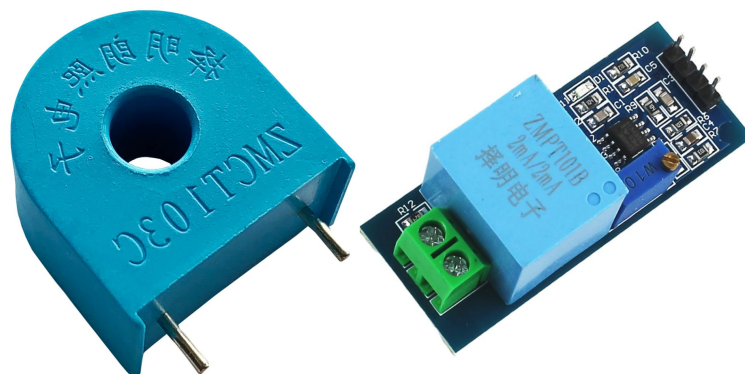
## 2.2 Extração e Representação de Assinaturas Elétricas

A correta identificação de cargas por meio do monitoramento não intrusivo depende diretamente da capacidade do sistema em traduzir sinais brutos em informações úteis. Essas informações, conhecidas como *assinaturas elétricas*, correspondem a padrões matemáticos únicos que cada aparelho imprime na rede durante sua operação. O sucesso de um algoritmo de classificação não está apenas na sua complexidade, mas principalmente na qualidade dessas características extraídas.

Visando a viabilidade de implementação em larga escala de soluções de NILM, a literatura técnica cita o uso de sensores de baixo custo. Tipicamente, utiliza-se o transformador de corrente (TC) não invasivo para a leitura de corrente, operando por indução magnética, e módulos de isolamento (como o ZMPT101B) para a leitura de tensão. Embora apresentem limitações de linearidade comparados a equipamentos de laboratório, Ahammed *et al.* (2021) demonstram que esses sensores fornecem precisão suficiente para aplicações de desagregação de cargas quando combinados a um pré-processamento adequado.

Neste trabalho, esses sensores são citados como referência tecnológica para a arquitetura de *hardware* visada em aplicações futuras de *Edge Computing*. Ressalta-se, entretanto, que os experimentos aqui apresentados foram conduzidos com bases de dados públicas (PLAID e COOLL) que servem como *ground truth* para validar os algoritmos de reconhecimento de padrões propostos.

Figura 3 – Exemplo de sensores comumente utilizados para aquisição de sinais em NILM de baixo custo: (esq.) transformador de corrente TC; (dir.) sensor de tensão ZMPT101B.



Fonte: Imagem adaptada de anúncios comerciais dos sensores TC e ZMPT101B.

A seguir, discutem-se as estratégias de extração de características adotadas, divididas entre a análise da forma de onda no tempo e a decomposição espectral na frequência.

### 2.2.1 Análise Temporal e Morfologia do Sinal

A análise no domínio do tempo é a abordagem mais direta e computacionalmente leve. Tradicionalmente, utilizam-se métricas de potência como o valor eficaz (*Root Mean Square* (RMS)) e a potência ativa para distinguir cargas. No entanto, em um cenário residencial moderno, muitos dispositivos possuem potências nominais semelhantes, o que torna o uso exclusivo do RMS insuficiente para uma desagregação precisa (Hart, 1992).

Para superar essa limitação, a literatura aponta para o uso de estatísticas de ordem superior e descritores de forma. Estes parâmetros buscam quantificar a morfologia da onda, permitindo diferenciar cargas lineares de não-lineares. Dentre eles, destacam-se:

- **Fator de Crista ( $C$ ):** Definido pela razão entre o valor de pico e o valor RMS da onda ( $C = I_{pico}/I_{rms}$ ). Para uma carga resistiva senoidal pura,  $C \approx 1,41$ . Já para cargas eletrônicas retificadas, esse valor tende a ser significativamente maior. Schirmer e Mporas (2019) avaliam essa métrica (denominada por eles como *Peak2RMS*) e destacam sua relevância na caracterização de sinais elétricos.
- **Curtose:** É uma medida estatística que descreve a "cauda" da distribuição de probabilidade do sinal, indicando o grau de impulsividade ou a presença de picos abruptos (*outliers*). Em aplicações de NILM, a curtose é utilizada para enriquecer o vetor de características, auxiliando na distinção de aparelhos com padrões de consumo não-gaussianos (Schirmer; Mporas, 2019).

A vantagem desses atributos é que eles são invariantes à escala; ou seja, permitem diferenciar um dispositivo pequeno de um grande apenas pelo formato da onda, sendo vitais para a robustez de algoritmos de aprendizado de máquina.

### 2.2.2 Análise Espectral e Distorção Harmônica

Enquanto a análise temporal descreve a envoltória do sinal, a análise no domínio da frequência revela a composição interna da energia consumida. A ferramenta matemática fundamental para isso é a Transformada Rápida de Fourier (FFT), que decompõe o sinal periódico em uma soma de senóides com frequências múltiplas da fundamental (60 Hz).

Cargas não-lineares, como fontes chaveadas e inversores de frequência, operam comutando a corrente milhares de vezes por segundo. Isso gera distorções harmônicas, tipicamente nas ordens ímpares (3<sup>a</sup>, 5<sup>a</sup>, 7<sup>a</sup> harmônicas). Conforme demonstrado por Kang *et al.*

(2020), a magnitude dessas componentes harmônicas funciona como uma "impressão digital" do circuito interno do aparelho. Por exemplo, uma lâmpada fluorescente compacta e uma lâmpada *LED* podem ter a mesma potência, mas apresentam espectros de frequência distintos devido às diferentes topologias de seus drivers eletrônicos.

Para quantificar o nível de deformação da onda de corrente em relação a uma senóide pura, utiliza-se a métrica da Distorção Harmônica Total (DHT), ou THD do inglês (*Total Harmonic Distortion*). A DHT é definida matematicamente pela razão entre o valor eficaz (RMS) de todas as componentes harmônicas somadas e o valor eficaz da componente fundamental, conforme expresso na Equação 2.1:

$$\text{DHT} = \frac{\sqrt{\sum_{n=2}^{\infty} I_n^2}}{I_1} \quad (2.1)$$

Em que  $I_n$  representa o valor eficaz da  $n$ -ésima componente harmônica e  $I_1$  corresponde ao valor eficaz da componente fundamental (60 Hz). Valores elevados de DHT indicam que o dispositivo drena energia de forma pulsada ou abrupta, característica intrínseca de retificadores e conversores estáticos presentes na eletrônica de potência (Pomilio, 1998).

### 2.2.3 Abordagens de Tempo-Frequência

Existem ainda técnicas híbridas que analisam como as frequências variam ao longo do tempo, como a Transformada de Fourier de Tempo Curto (STFT) e a Transformada Wavelet Discreta (DWT). A STFT utiliza uma janela fixa para segmentar o sinal, o que impõe uma limitação na resolução tempo-frequência conhecida como Princípio da Incerteza de Heisenberg.

Para superar essa limitação, a DWT emprega uma análise multirresolução, decompondo o sinal em coeficientes de aproximação (baixas frequências) e detalhes (altas frequências) através de sucessivos níveis de filtragem. Essa característica torna a DWT particularmente eficaz para isolar transientes rápidos de carga e ruídos de chaveamento que ocorrem em curtos intervalos de tempo, conforme demonstrado por Guo *et al.* (2020).

Embora essas técnicas possam gerar representações de alta dimensionalidade (como espectrogramas), a extração de estatísticas (como energia e desvio padrão) dos níveis de decomposição permite condensar a informação relevante em um vetor compacto. Em sistemas embarcados focados em processamento na borda (*Edge Computing*), essa estratégia viabiliza o uso de informações de tempo-frequência sem comprometer a eficiência computacional, pri-

vilegiando vetores de características unidimensionais em detrimento de matrizes complexas (Ahammed *et al.*, 2021).

#### **2.2.4 Engenharia de Características (Feature Engineering)**

A etapa final do processamento consiste na construção do vetor de características que alimentará o classificador. A tendência atual na literatura não é escolher entre tempo ou frequência, mas sim realizar a fusão dessas informações.

Essa abordagem, conhecida como Engenharia de Características, concatena métricas estatísticas (para capturar a forma da onda e a potência) com os coeficientes da FFT (para capturar a não-linearidade) e, neste trabalho, com os descritores da DWT (para capturar transientes). Trabalhos como o de Paixao *et al.* (2016) mostram que essa redundância controlada de informações aumenta a robustez do classificador, permitindo que algoritmos como *Random Forest* ou SVM encontrem hiperplanos de separação ótimos mesmo em ambientes ruidosos.

### **2.3 Algoritmos de Classificação Supervisionada**

No contexto do monitoramento não intrusivo de cargas (NILM), a etapa de classificação supervisionada tem por objetivo associar um vetor de características extraídas dos sinais elétricos a uma determinada carga em operação. Para isso, diversos algoritmos de aprendizado supervisionado têm sido utilizados na literatura, com diferentes níveis de complexidade, desempenho e requisitos computacionais.

Esses modelos são treinados a partir de um conjunto de dados rotulado, onde cada vetor de características está associado a uma classe correspondente a uma carga conhecida. Após o treinamento, espera-se que o classificador generalize o conhecimento aprendido, sendo capaz de identificar corretamente novas instâncias. A seguir, são apresentados os principais modelos supervisionados aplicados na tarefa de identificação de cargas elétricas.

#### **2.3.1 *k*-Nearest Neighbors (*k*-NN)**

O algoritmo *k*-Nearest Neighbors é um classificador baseado em instâncias que atribui a classe de uma amostra desconhecida com base na classe predominante entre os *k* vizinhos mais próximos, segundo uma métrica de distância. A distância Euclidiana é a mais comum, mas também são utilizadas Manhattan, Minkowski, entre outras.

A simplicidade de implementação e a ausência de fase explícita de treinamento são as principais vantagens do k-NN. No entanto, ele é sensível a ruídos, atributos irrelevantes e à escala dos dados, além de apresentar maior custo computacional no momento da inferência.

Pereira (2017) utilizaram o k-NN para classificar cargas residenciais com base em características extraídas por STFT e *kernel PCA*, alcançando acurácia superior a 80%. Já Devlin e Hayes (2019) mostraram que, mesmo sem redução de dimensionalidade, o k-NN se manteve competitivo frente a classificadores mais complexos, desde que o conjunto de atributos fosse bem definido.

### 2.3.2 *Árvores de Decisão e Florestas Aleatórias*

As Árvores de Decisão (*Decision Trees*) são classificadores baseados em regras hierárquicas, que dividem o espaço de atributos em regiões associadas a diferentes classes. A principal vantagem das árvores está na interpretabilidade e na facilidade de identificar os atributos mais relevantes para a tomada de decisão.

Já as Florestas Aleatórias (*Random Forests*) representam uma evolução robusta desse conceito. O algoritmo opera sob o princípio de *Ensemble Learning*, utilizando a técnica de *Bagging (Bootstrap Aggregating)*. Isso significa que o modelo é formado por um conjunto de árvores de decisão independentes, onde cada uma é treinada com uma subamostra aleatória dos dados originais.

A classificação final é obtida pela votação majoritária de todas as árvores. Essa arquitetura confere ao *Random Forest* uma vantagem significativa em cenários de dados desbalanceados (comuns em NILM, onde algumas cargas são acionadas raramente): o "comitê" de árvores consegue mitigar o viés em direção à classe majoritária, capturando nuances das classes minoritárias que um classificador linear simples poderia ignorar. Segundo Kang *et al.* (2020), essa abordagem em conjunto com harmônicos de baixa ordem permite classificar eletrodomésticos com alto desempenho e tolerância a ruídos.

### 2.3.3 *Máquinas de Vetores de Suporte (SVM)*

As Máquinas de Vetores de Suporte (SVM) são classificadores que buscam o hiperplano ótimo que separa duas ou mais classes com a maior margem possível. Em casos em que os dados não são linearmente separáveis, utilizam-se funções *kernel*, como a *Radial Basis Function (RBF)*, para projetar os dados em um espaço de maior dimensionalidade onde a separação se

torne viável.

Paixao *et al.* (2016) aplicaram SVMs à identificação de cargas “idênticas”, observando desempenho competitivo em relação a k-NN e RNAs, especialmente com atributos adequadamente selecionados. Apesar da boa acurácia em margens bem definidas, a SVM pode apresentar maior custo computacional e sensibilidade à escolha de hiperparâmetros, especialmente em datasets com forte sobreposição de classes.

#### **2.3.4 Redes Neurais Artificiais**

As Redes Neurais Artificiais (RNAs) consistem em camadas de unidades de processamento (neurônios) interconectadas, com pesos ajustados iterativamente durante o treinamento. São capazes de modelar relações não lineares complexas, sendo altamente eficazes quando há grandes volumes de dados.

Em Ahammed *et al.* (2021), os autores aplicaram RNAs simples para classificação de cargas em tempo real, utilizando sensores de corrente (TC) e tensão (ZMPT), com acurácia acima de 90%. Arquiteturas mais sofisticadas, como redes convolucionais (*CNN*) e redes recorrentes (*RNN*), também vêm sendo exploradas com sucesso (Guo *et al.*, 2020). Entretanto, RNAs exigem maior capacidade computacional (*hardware*), além de apresentarem menor interpretabilidade do que modelos baseados em árvores.

Ressalta-se que, devido à incompatibilidade com a premissa de baixo custo computacional para sistemas embarcados simples, as redes neurais são apresentadas neste trabalho apenas como referência do estado da arte, não sendo implementadas experimentalmente.

#### **2.3.5 Comparativo e Escolha do Algoritmo**

A seleção do algoritmo ideal depende de uma série de fatores, como o tamanho e a qualidade do conjunto de dados, a presença de ruídos, a complexidade da tarefa, o custo computacional disponível e a necessidade de interpretabilidade dos resultados.

Em aplicações onde o volume de dados é limitado e o sistema precisa ser interpretável, modelos como k-NN e *Random Forest* se mostram vantajosos. Por outro lado, RNAs e SVMs tendem a oferecer maior acurácia quando os dados são abundantes ou os padrões são altamente não lineares. A Tabela 1 apresenta um resumo das principais características teóricas dos algoritmos discutidos. Com base nessas características, optou-se por limitar os experimentos deste trabalho ao k-NN, SVM e *Random Forest*.

Tabela 1 – Comparativo entre classificadores supervisionados aplicados em NILM

Algoritmo	Vantagens	Desvantagens	Aplicações típicas em NILM
k-NN	Simples, não requer treinamento, bom para dados bem separados	Sensível a ruído, alta complexidade em tempo de inferência	Classificação básica com conjuntos moderados
Árvore de Decisão	Fácil interpretação, rápido treinamento	Propenso a sobreajuste em conjuntos ruidosos	Avaliação de atributos e decisões rápidas
<i>Random Forest</i>	Robusto a ruído, lida bem com desbalanceamento	Maior uso de memória para armazenar as árvores	Reconhecimento de padrões complexos e dados redundantes
SVM	Alta acurácia em margens claras	Alto custo computacional, sensível a hiperparâmetros	Separação de classes binárias ou com pouca sobreposição
Rede Neural Artificial	Modela relações complexas, bom desempenho com muitos dados	Menor interpretabilidade e alta demanda computacional	Ambientes com grandes volumes de dados e múltiplas classes

Fonte: Elaborada pelo autor.

## 2.4 Métricas de Avaliação de Desempenho

A avaliação do desempenho dos algoritmos de classificação é uma etapa essencial em qualquer sistema baseado em aprendizado supervisionado, especialmente no contexto de identificação de cargas elétricas por meio de monitoramento não intrusivo. As métricas estatísticas são utilizadas para quantificar o quão bem o modelo foi capaz de associar corretamente um vetor de características a sua respectiva classe, contribuindo para uma análise crítica do comportamento do classificador.

As métricas mais comuns derivam da *matriz de confusão*, que resume as previsões realizadas em comparação com os rótulos reais dos dados. Entretanto, a escolha da métrica ideal deve considerar o tipo de problema (binário ou multiclasse), o balanceamento das classes e os objetivos práticos da aplicação. Nos próximos tópicos, são apresentadas as principais métricas utilizadas em NILM.

### 2.4.1 Acurácia

A acurácia mede a proporção de amostras corretamente classificadas em relação ao total de amostras avaliadas:

$$\text{Acurácia} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.2)$$

Em que:

- *TP (True Positive)*: amostras positivas corretamente classificadas;
- *TN (True Negative)*: amostras negativas corretamente classificadas;
- *FP (False Positive)*: amostras negativas incorretamente classificadas como positivas;
- *FN (False Negative)*: amostras positivas incorretamente classificadas como negativas.

Neste contexto, *TP*, *TN*, *FP* e *FN* representam contagens absolutas derivadas da matriz de confusão. Essa métrica é apropriada quando as classes estão balanceadas. Entretanto, pode fornecer uma falsa impressão de desempenho em contextos desbalanceados, onde a maioria das amostras pertence a uma única classe dominante (Chicco; Jurman, 2020).

### 2.4.2 *Precisão e Revocação*

A precisão (*Precision*) e a revocação (*Recall*) oferecem uma avaliação mais refinada do desempenho para cada classe:

$$\text{Precisão} = \frac{TP}{TP + FP}, \quad \text{Revocação} = \frac{TP}{TP + FN} \quad (2.3)$$

A precisão indica a proporção de amostras classificadas como positivas que realmente pertencem àquela classe, enquanto a revocação mede a capacidade do modelo de identificar corretamente todas as amostras positivas reais.

Essas métricas são úteis para identificar se o classificador apresenta tendência a muitos falsos positivos (alta revocação, baixa precisão) ou falsos negativos (alta precisão, baixa revocação). Em NILM, tais desequilíbrios podem significar que uma carga foi detectada quando não estava em operação, ou não foi detectada quando estava ligada (Klemenjak *et al.*, 2020).

### 2.4.3 *F1-Score*

O F1-score é a média harmônica entre precisão e revocação, fornecendo um balanço entre ambas:

$$\text{F1-score} = 2 \cdot \frac{\text{Precisão} \cdot \text{Revocação}}{\text{Precisão} + \text{Revocação}} \quad (2.4)$$

Essa métrica é especialmente útil em contextos onde existe um trade-off entre precisão e revocação, como ocorre em cenários de classificação de cargas semelhantes. Trabalhos como os de Guo *et al.* (2020) e Ahammed *et al.* (2021) a adotam como principal critério de avaliação.

#### 2.4.4 Matriz de Confusão

A matriz de confusão apresenta uma visão detalhada dos acertos e erros por classe. Cada linha representa a classe real, enquanto cada coluna representa a classe prevista.

Figura 4 – Exemplo ilustrativo de uma matriz de confusão em um problema de classificação com três classes.

		CLASSE I					CLASSE N-I				
		Predicted = 1	Predicted = 2	Predicted = ...	Predicted = N-1	Predicted = N	Predicted = 1	Predicted = 2	Predicted = ...	Predicted = N-1	Predicted = N
Real = 1	Real = 1	TP	FN	FN	FN	FN	TN	TN	TN	FP	TN
	Real = 2	FP	TN	TN	TN	TN	TN	TN	TN	FP	TN
	Real = ...	FP	TN	...	TN	TN	TN	TN	...	FP	TN
	Real = N-1	FP	TN	TN	TN	TN	FN	FN	FN	TP	FN
	Real = N	FP	TN	TN	TN	TN	TN	TN	TN	FP	TN

Fonte: Adaptado de (Amuedo, 2021).

Em Paixao *et al.* (2016), observou-se que dispositivos “idênticos”, como lâmpadas fluorescentes, apresentaram altos índices de confusão mútua, o que evidencia a importância da qualidade das características extraídas.

#### 2.4.5 Curvas ROC e AUC

As curvas ROC (*Receiver Operating Characteristic*) ilustram a relação entre a taxa de verdadeiros positivos (revocação) e a taxa de falsos positivos para diferentes limiares de decisão. A métrica AUC (*Area Under the Curve*) resume o desempenho em um único valor entre 0 e 1.

Apesar de serem mais comuns em problemas binários, essas métricas podem ser aplicadas a classificações multiclasse por meio de estratégias como *one-vs-rest*. Em NILM, utilizaram AUC como forma de validar a robustez de seus modelos frente a diferentes níveis de ruído.

#### 2.4.6 Validação Cruzada e Overfitting

A validação cruzada é uma técnica amplamente utilizada para estimar o desempenho generalizável de um modelo. A abordagem mais comum é o *k-fold cross-validation*, em que o conjunto de dados é dividido em *k* subconjuntos e o modelo é treinado e testado *k* vezes com

combinações diferentes.

Esse processo ajuda a evitar o *overfitting*, ou seja, a tendência do modelo de memorizar os dados de treinamento, prejudicando seu desempenho em dados novos. A escolha de  $k = 5$  ou  $k = 10$  é comum na literatura por equilibrar variância e viés na estimativa de desempenho.

#### 2.4.7 *Interpretação e Limitações*

Embora as métricas descritas sejam essenciais para avaliar classificadores, elas não devem ser utilizadas de forma isolada. Em aplicações reais de NILM, outras dimensões de avaliação, como latência, interpretabilidade, robustez ao ruído e escalabilidade, são igualmente relevantes. Portanto, a análise quantitativa deve sempre ser acompanhada de uma interpretação qualitativa dos resultados obtidos, considerando o contexto da aplicação final.

### 2.5 **Computação na Borda e *TinyML***

Tradicionalmente, soluções de IoT operam sob um paradigma centralizado, no qual os dados coletados por sensores são transmitidos para servidores remotos (*Cloud Computing*) para processamento e armazenamento. Embora eficaz para análises históricas, essa arquitetura apresenta limitações críticas para sistemas de NILM operando em tempo real. A transmissão contínua de sinais de corrente em alta frequência exige grande largura de banda, introduz latência na rede e levanta preocupações sobre privacidade (Shi *et al.*, 2016).

Como alternativa, a Computação na Borda (*Edge Computing*) propõe deslocar o processamento computacional da nuvem para a "borda" da rede, ou seja, para o próprio dispositivo onde o dado é gerado. A Tabela 2 apresenta um comparativo detalhado entre as características dessas duas arquiteturas, destacando as vantagens do processamento local para aplicações sensíveis ao tempo.

Para viabilizar a inteligência artificial nesses dispositivos com recursos limitados (como o microcontrolador ESP32, que possui pouca memória RAM e armazenamento), surgiu o campo do *Tiny Machine Learning (TinyML)*. Esta área foca na otimização de modelos para que consumam miliwatts de energia e realizem inferências locais (Warden; Situnayake, 2019).

A aplicação de *TinyML* em NILM exige uma etapa de extração de características eficiente (como a FFT e a análise estatística discutidas na seção 2.2) e a escolha de classificadores leves, como *Random Forests*, que podem ser convertidos para código C otimizado, garantindo

Tabela 2 – Comparativo entre as arquiteturas de Computação em Nuvem e Computação na Borda

<b>Computação em Nuvem (<i>Cloud</i>)</b>	<b>Computação na Borda (<i>Edge</i>)</b>
Baseia-se em servidores centralizados armazenados em <i>data centers</i> remotos de larga escala.	É uma infraestrutura altamente distribuída, localizada fisicamente próxima aos dispositivos e usuários.
Processa os dados em servidores centrais, distantes da fonte geradora.	Processa os dados no local ( <i>on site</i> ) rapidamente, analisando em tempo real sem foco em armazenamento massivo.
Adequada para aplicações que não são sensíveis ao tempo (toleram latência).	Ideal para baixa latência, crítica para sistemas onde cada milissegundo conta.
Oferece capacidades de processamento e armazenamento virtualmente ilimitadas.	Por processar no dispositivo, possui poder de computação e armazenamento limitados ao <i>hardware</i> local.
Adequada para análises profundas de longo prazo (histórico).	Melhor para análises rápidas e tomada de decisão imediata.
Requer conectividade constante com a Internet.	Pode operar sem conectividade com a Internet (funcionamento <i>offline</i> ).
Custos elevados associados à conectividade, movimentação de dados e largura de banda.	Menores requisitos de largura de banda e latência, resultando em maior desempenho e menores custos operacionais.

Fonte: Elaborada pelo autor, adaptado de XenonStack (2024).

independência da nuvem conforme as vantagens listadas na Tabela 2. Embora o presente trabalho não implemente classificação em tempo real, os resultados discutidos servem como base conceitual para futuras extensões nessa direção.

### 3 TRABALHOS RELACIONADOS

O avanço das tecnologias de sensoriamento, Internet das Coisas (IoT) e, notavelmente, Inteligência Artificial (IA), tem revolucionado o setor elétrico. Dentro das *Smart Grids* (Redes Elétricas Inteligentes), a capacidade de monitorar e analisar padrões de consumo de energia é fundamental para otimizar a gestão da demanda, promover a eficiência energética e identificar comportamentos atípicos. Neste contexto, a aplicação de algoritmos de aprendizado de máquina supervisionado para a identificação de padrões em sinais elétricos de corrente e tensão emerge como uma área de pesquisa promissora e de grande impacto.

Esta seção apresenta e discute trabalhos relevantes da literatura que fundamentam e contextualizam a pesquisa proposta, destacando metodologias, resultados e lacunas que o presente estudo busca explorar. A análise prioriza estudos que empregam técnicas de processamento de sinais (especialmente aquelas baseadas em Transformada de Fourier), extração de características para formar assinaturas elétricas, e algoritmos de classificação por aprendizado de máquina.

#### 3.1 STFT e Kernel PCA para Identificação de Cargas

A identificação de padrões e a definição de assinaturas elétricas a partir de sinais de corrente e tensão representam o cerne de diversas aplicações no monitoramento de energia. Neste sentido, o trabalho de Pereira (2017) apresenta uma metodologia robusta para a identificação de cargas elétricas através da extração de características. O estudo emprega a Transformada de Fourier de Tempo Curto (STFT) em conjunto com o *Kernel PCA* para definir as assinaturas das cargas a partir de medições de corrente. A validação, realizada com o classificador *kNN*, demonstrou uma taxa de acerto de 83%, evidenciando a eficácia da abordagem baseada em Fourier e a capacidade de separação das informações. Este trabalho é de particular relevância para a presente pesquisa, pois valida a aplicação de técnicas de Fourier para a criação de assinaturas elétricas, uma etapa fundamental para a classificação de padrões.

#### 3.2 Classificação de Cargas Elétricas Idênticas

Complementando essa abordagem, Paixao *et al.* (2016) explora a classificação inteligente para a identificação de cargas elétricas, mesmo aquelas consideradas "idênticas". O estudo utiliza diretamente sinais elétricos de tensão e corrente adquiridos em diferentes frequências de amostragem, a partir dos quais são extraídas 14 características distintas para

compor a assinatura elétrica das cargas. A aplicação de métodos de classificação inteligente alcançou acurácias de até 84,79% na identificação de uma carga específica e até 72,5% na identificação da configuração de um arranjo complexo. Este trabalho ressalta a importância de uma engenharia de características detalhada a partir dos sinais brutos para diferenciar padrões sutis, reforçando a metodologia proposta neste TCC.

### **3.3 Monitoramento e Identificação de Cargas com ML**

Ainda no escopo da identificação de cargas, Shibata *et al.* (2021) apresenta um sistema de monitoramento e desagregação do consumo de cargas (NILM), capaz de identificar individualmente eletrodomésticos. O sistema utiliza sensores de corrente para coletar dados e identificar a assinatura de cada dispositivo por meio de um modelo de aprendizado de máquina para classificação. Com uma acurácia de 93,01% na identificação de seis tipos de eletrodomésticos, este estudo reforça a viabilidade e a eficácia da aplicação de *Machine Learning* na identificação de padrões de consumo baseados em dados de corrente, servindo como um forte precedente para a pesquisa atual.

### **3.4 Classificação de Cargas Elétricas em Tempo Real**

A aplicação de aprendizado de máquina em dados elétricos estende-se à classificação de padrões em tempo real e em condições variadas. O estudo de Ahammed *et al.* (2021) propõe um sistema de classificação de cargas elétricas não intrusivo em tempo real (*RT-NILC*), demonstrando robustez mesmo sob variabilidade da tensão de alimentação. O trabalho utiliza dados de tensão RMS, corrente RMS, potência ativa e fator de potência, processados por modelos de classificação *multilabel* que alcançaram acurácia superior a 94%. Este estudo destaca a capacidade dos algoritmos de classificação em lidar com dados elétricos em cenários práticos e variáveis, e a pertinência de utilizar múltiplos parâmetros derivados dos sinais de corrente e tensão para a identificação de padrões.

### **3.5 Machine Learning para Previsão de Carga**

Além da identificação direta de padrões, a aplicação de aprendizado de máquina em redes elétricas abrange desafios como a previsão de carga e a detecção de anomalias, que fornecem *insights* metodológicos e conceituais valiosos. A tese de doutorado de Johannesen

(2022) explora a aplicação de *Machine Learning* para previsão de carga, investigando técnicas de regressão como *kNN* e *Random Forest*. Embora o foco seja a previsão, a tese aborda a importância de um pré-processamento robusto, incluindo a Análise de Componentes Principais (*Principal Component Analysis (PCA)*) para redução de dimensionalidade, uma técnica que pode ser aplicada para otimizar as "assinaturas elétricas" no presente trabalho. Adicionalmente, a investigação da detecção de anomalias nessa tese é relevante, pois a identificação de padrões incomuns complementa a identificação de padrões normais, fornecendo uma perspectiva mais ampla sobre a análise de sinais elétricos.

### **3.6 Redes Neurais para Predição de Curvas de Carga**

Por fim, o artigo de Urzagasti (2022) também demonstra a aplicação de modelos de aprendizado de máquina na predição de curvas de carga. Embora sua tarefa seja a previsão e não a classificação de padrões de sinais brutos, o estudo exemplifica o uso de Redes Neurais Artificiais (ANN) e *Long Short-Term Memory (LSTM)* em dados elétricos, com acurácia superior a 0,95. Este trabalho contribui para o panorama geral da aplicação de *Machine Learning* no setor elétrico, ressaltando a capacidade desses modelos em extrair informações valiosas de dados energéticos, validando o campo de atuação do presente TCC.

A análise desses trabalhos prévios revela a consolidação do aprendizado de máquina como uma ferramenta essencial para a compreensão e gestão de sistemas elétricos. A presente pesquisa se insere nesse cenário, buscando aprofundar a identificação de padrões elétricos de corrente e tensão através da engenharia de características baseada em Transformada de Fourier Rápida e da aplicação de algoritmos de aprendizado supervisionado, com o objetivo de contribuir para o aprimoramento do monitoramento e da eficiência energética.

### **3.7 Análise Comparativa**

Para facilitar a visualização das diferenças entre as abordagens existentes na literatura e a proposta deste trabalho, elaborou-se a Tabela 3. Nela, são comparados o objetivo principal de cada pesquisa, os tipos de características extraídas dos sinais e o algoritmo de aprendizado de máquina utilizado como base.

Ao observar a Tabela 3, nota-se que a maioria dos trabalhos tende a focar isoladamente em características do domínio do tempo (como RMS e Potência) ou do domínio da

Tabela 3 – Comparativo de metodologias entre os trabalhos relacionados e a proposta atual

Trabalhos / Autores	Objetivo		Extração de Carac.		Família de Alg.	
	Identif.	Previsão	Tempo	Freq.	ML	RN
Pereira (2017)	✓			✓	✓	
Paixao <i>et al.</i> (2016)	✓		✓		✓	
Shibata <i>et al.</i> (2021)	✓		✓		✓	
Ahammed <i>et al.</i> (2021)	✓		✓			✓
Johannesen (2022)		✓	✓		✓	
Urzagasti (2022)		✓	✓			✓
<b>Este Trabalho</b>	✓		✓	✓	✓	

Fonte: Elaborada pelo autor.

Nota: Legenda: Identif. (Identificação de Cargas); Tempo (Estatísticas/RMS); Freq. (Espectral/FFT); ML (Clássico: SVM, k-NN); RN (Redes Neurais).

frequência. O diferencial da proposta deste TCC reside na abordagem híbrida. O modelo desenvolvido busca integrar a robustez das estatísticas temporais com a capacidade de discriminação da FFT. Além disso, a escolha dos algoritmos propostos mostrou-se essencial para a correta identificação de cargas não-lineares, preenchendo uma lacuna observada em estudos voltados apenas para previsão ou que utilizam métodos lineares mais simples.

## 4 METODOLOGIA

Este capítulo descreve as etapas metodológicas adotadas para o desenvolvimento do sistema de identificação de cargas elétricas a partir de sinais de corrente e tensão. A abordagem contempla desde a aquisição e preparação dos dados até a modelagem e avaliação do desempenho dos classificadores aplicados.

### 4.1 Tipo de Pesquisa e Abordagem Experimental

Este trabalho caracteriza-se como uma pesquisa de natureza aplicada, com o objetivo de avaliar comparativamente estratégias computacionais para a identificação automática de cargas elétricas. A abordagem adotada possui caráter experimental e quantitativo, uma vez que parte do processamento de bases de dados reais, segue para uma etapa de engenharia de características e culmina na avaliação estatística do desempenho de algoritmos de aprendizado de máquina supervisionado.

A pesquisa classifica-se como desenvolvimento tecnológico na área de Ciência de Dados aplicada a *Smart Grids*. Diferentemente de abordagens focadas puramente em *hardware*, este estudo concentra-se na investigação de técnicas de PDS e aprendizado de máquina. Os experimentos foram conduzidos utilizando bases de dados de referência na literatura (PLAID e COOLL), que contêm sinais de corrente e tensão de alta frequência, permitindo verificar a robustez dos modelos frente a diferentes cenários de operação.

Do ponto de vista metodológico, o estudo segue os princípios do método científico empírico, fundamentando-se na observação, hipótese e análise sistemática de dados. Todas as etapas, desde a manipulação dos sinais até a classificação e validação, foram implementadas em ambiente computacional de código aberto, utilizando a linguagem Python e bibliotecas especializadas como NumPy, SciPy e scikit-learn.

A adoção dessa abordagem visa garantir a reprodutibilidade dos experimentos, a transparência na manipulação dos dados e a aplicabilidade prática dos resultados, alinhando o trabalho às diretrizes atuais de pesquisa em monitoramento não-intrusivo de energia (Klemenjak *et al.*, 2020).

## 4.2 Seleção e Compreensão dos Dados

Para o desenvolvimento e validação dos modelos de identificação de cargas, optou-se pela utilização de bases de dados públicas consolidadas na literatura de NILM. Essa escolha estratégica, em detrimento da coleta primária de dados em ambiente restrito, permite: (i) garantir a reprodutibilidade dos resultados; (ii) comparar o desempenho do modelo proposto com outros trabalhos do estado da arte; e (iii) expor os algoritmos a uma maior variabilidade de cenários de ruído e tipos de cargas.

Foram selecionados dois *datasets* com características complementares: o PLAID (*Plug Load Appliance Identification Dataset*) e o COOLL (*Controlled On/Off Loads Library*). A combinação dessas fontes de dados visa mitigar o viés de um ambiente específico e testar a capacidade de generalização do sistema frente a diferentes taxas de amostragem e condições de instalação.

### 4.2.1 Base de Dados PLAID

O *dataset* PLAID (*Plug Load Appliance Identification Dataset*) é uma referência na pesquisa de identificação de cargas, consistindo em medições de alta resolução de corrente e tensão. Os dados utilizados neste trabalho provêm da versão PLAID1, coletada durante o verão de 2013 em mais de 56 residências localizadas em Pittsburgh, Pensilvânia (EUA).

A base abrange 11 tipos distintos de aparelhos eletrodomésticos: ar-condicionado, lâmpada fluorescente compacta, geladeira, secador de cabelo, *laptop*, micro-ondas, máquina de lavar, lâmpada incandescente, aspirador de pó, ventilador e aquecedor. Cada tipo é representado por dezenas de instâncias de diferentes marcas e modelos, garantindo uma alta variabilidade intra-classe (Gao *et al.*, 2014).

As medições foram realizadas a uma frequência de amostragem de 30 kHz. Para cada aparelho, foram coletados de três a seis ciclos de operação para cada transição de estado. Esses dados brutos foram submetidos a um pós-processamento pelos autores originais para extrair janelas de poucos segundos, contendo tanto o comportamento em regime permanente quanto o transiente de partida (quando disponível). Essa característica torna o PLAID ideal para validar algoritmos que dependem da análise de transientes rápidos e harmônicas em cenários residenciais reais, sujeitos a ruídos e distorções da rede elétrica.

#### 4.2.2 Base de Dados COOLL

Em contrapartida, o *dataset* COOLL oferece um ambiente de alta precisão e controle. Esta base contém medições de 42 tipos de eletrodomésticos, totalizando 840 amostras, adquiridas em laboratório com uma taxa de amostragem elevada de 100 kHz.

Diferentemente do PLAID, os sensores utilizados no COOLL possuem altíssima sensibilidade e o ambiente elétrico é livre de interferências externas significativas. A inclusão desta base no estudo permite analisar o comportamento ideal das assinaturas elétricas (harmônicas e forma de onda) sem a contaminação de ruídos de fundo. Isso é crucial para entender se eventuais erros de classificação decorrem de limitações do algoritmo ou da qualidade do sinal adquirido (Picon *et al.*, 2016).

#### 4.2.3 Integração e Heterogeneidade dos Dados

A utilização conjunta dessas duas bases impõe um desafio de engenharia de dados, dado que elas possuem frequências de amostragem e formatos distintos (30 kHz no PLAID vs 100 kHz no COOLL). Para viabilizar o treinamento unificado dos modelos de aprendizado de máquina, foi necessário estabelecer um procedimento de padronização.

Optou-se por adotar a frequência de 30 kHz como padrão para todo o projeto, visto que é a frequência nativa do *dataset* mais desafiador (PLAID) e suficiente para capturar as componentes espectrais de interesse. Consequentemente, os sinais provenientes da base COOLL foram submetidos a um processo de adequação da taxa de amostragem. Essa estratégia permitiu a concatenação das bases em um conjunto de dados híbrido e heterogêneo, ideal para treinar um modelo generalista capaz de operar em diferentes condições de aquisição.

### 4.3 Preparação e Engenharia de Dados

A eficácia de modelos de aprendizado de máquina depende intrinsecamente da qualidade e da consistência dos dados de entrada. Visto que os sinais provenientes das bases PLAID e COOLL foram adquiridos em contextos distintos e possuem características brutas heterogêneas, torna-se indispensável uma etapa rigorosa de pré-processamento.

Esta seção descreve o *pipeline* de tratamento de dados desenvolvido, que engloba desde a uniformização das taxas de amostragem até a detecção de eventos e limpeza de ruídos. O objetivo dessas transformações é converter os sinais brutos em um formato estruturado e

padronizado, preservando as informações espectrais relevantes para a classificação, ao mesmo tempo em que se reduz a dimensionalidade para viabilizar o processamento em sistemas com recursos limitados.

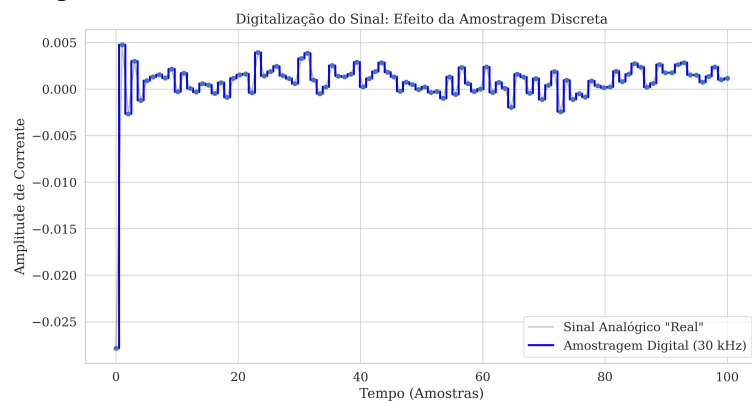
#### 4.3.1 Integração e Reamostragem (Downsampling)

O primeiro desafio técnico abordado foi a discrepância entre as frequências de amostragem das bases selecionadas: 30 kHz para o PLAID e 100 kHz para o COOLL. Para permitir o treinamento conjunto dos algoritmos, estabeleceu-se a frequência de 30 kHz como a taxa de amostragem padrão do sistema ( $f_s$ ).

Optou-se por realizar o processo de subamostragem (*downsampling*) nos sinais da base COOLL, reduzindo sua resolução temporal para equiparar-se à do PLAID. Este procedimento foi realizado utilizando técnicas de decimação após a aplicação de filtros *anti-aliasing*, garantindo a integridade espectral na faixa de interesse e prevenindo a distorção do sinal. Do ponto de vista teórico, essa redução é plenamente justificada pelo Teorema da Amostragem de Nyquist-Shannon. Uma taxa de  $f_s = 30$  kHz permite a reconstrução perfeita de frequências de até 15 kHz. Considerando que as harmônicas mais relevantes para a identificação de cargas residenciais situam-se tipicamente abaixo de 2 kHz (aprox. 33ª harmônica), a redução para 30 kHz não acarreta perda de informação discriminativa.

A Figura 5 ilustra o efeito visual e espectral dessa padronização, demonstrando que a morfologia da onda (fundamental para o cálculo do Fator de Crista) e os picos de harmônicas são preservados após o processamento.

Figura 5 – Preservação da morfologia do sinal após a reamostragem: (a) Sinal original em alta frequência; (b) Sinal padronizado em 30 kHz.



Fonte: Elaborada pelo autor.

### 4.3.2 Detecção de Eventos e Limpeza dos Dados

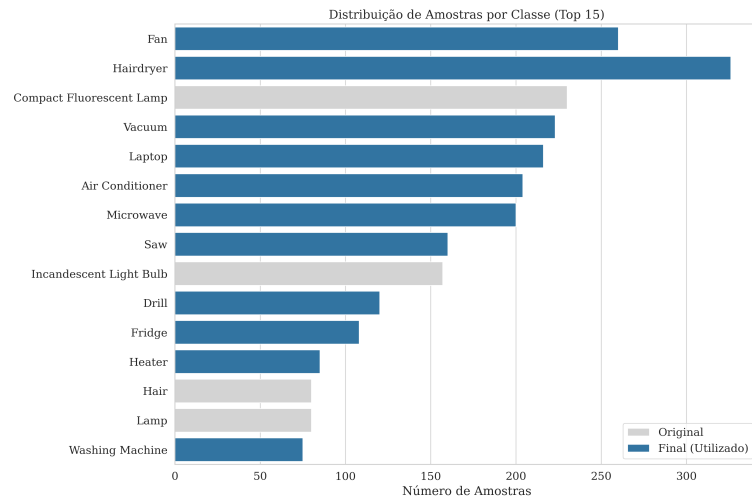
Apesar de os bancos de dados selecionados já apresentarem os sinais segmentados em janelas de operação, observou-se que diversas amostras continham longos períodos de inatividade ou ruídos de baixa intensidade que não correspondiam ao funcionamento efetivo da carga. A presença desses trechos de "silêncio" prejudica a extração de características, pois dilui a relevância estatística dos parâmetros calculados, como o valor RMS e a curtose.

Para mitigar esse problema, implementou-se um algoritmo de detecção de eventos baseado em limiar de corrente (*current thresholding*). Definiu-se empiricamente um limiar de ativação de 0,5 A. Apenas os segmentos de sinal cuja corrente instantânea superasse esse valor foram considerados válidos para a etapa de treinamento. Esse procedimento assegura que o classificador seja alimentado exclusivamente com dados representativos do estado "ligado" dos equipamentos, eliminando o ruído de fundo do modo *standby*.

Adicionalmente, realizou-se uma limpeza estatística baseada na representatividade das classes. Ao analisar a distribuição original dos dados unificados, constatou-se um forte desbalanceamento: enquanto certas cargas (como lâmpadas e ventiladores) possuíam centenas de exemplos, outras apresentavam uma quantidade insuficiente de amostras para garantir a generalização do modelo.

Para assegurar a robustez estatística durante a validação cruzada, foram descartadas as classes que possuíam menos de 20 instâncias válidas. A Figura 6 apresenta o impacto desse filtro na distribuição das classes, evidenciando o cenário final de dados que foi utilizado para o treinamento dos modelos. Nota-se que, mesmo após a limpeza, o conjunto de dados permanece desbalanceado, o que justifica a adoção posterior de métricas de avaliação ponderadas, como o *F1-Score Macro*.

Figura 6 – Distribuição das classes no conjunto de dados final após a filtragem de relevância estatística.



Fonte: Elaborada pelo autor.

#### 4.4 Engenharia de Características (*Feature Engineering*)

Após a etapa de limpeza e padronização, os dados ainda se encontram em sua forma bruta: vetores de séries temporais com milhares de pontos por janela. A utilização direta desses sinais brutos em algoritmos de aprendizado de máquina clássicos (como *Random Forest* ou SVM) é inviável devido à alta dimensionalidade e à redundância de informações.

Portanto, faz-se necessária a aplicação de técnicas de Engenharia de Características para transformar o sinal bruto em um conjunto compacto de atributos discriminativos, compondo a chamada "assinatura elétrica" da carga. A estratégia adotada neste trabalho é híbrida, integrando a análise espectral (frequência) e a análise estatística (tempo) com a decomposição via Wavelets (tempo-frequência), conforme detalhado a seguir.

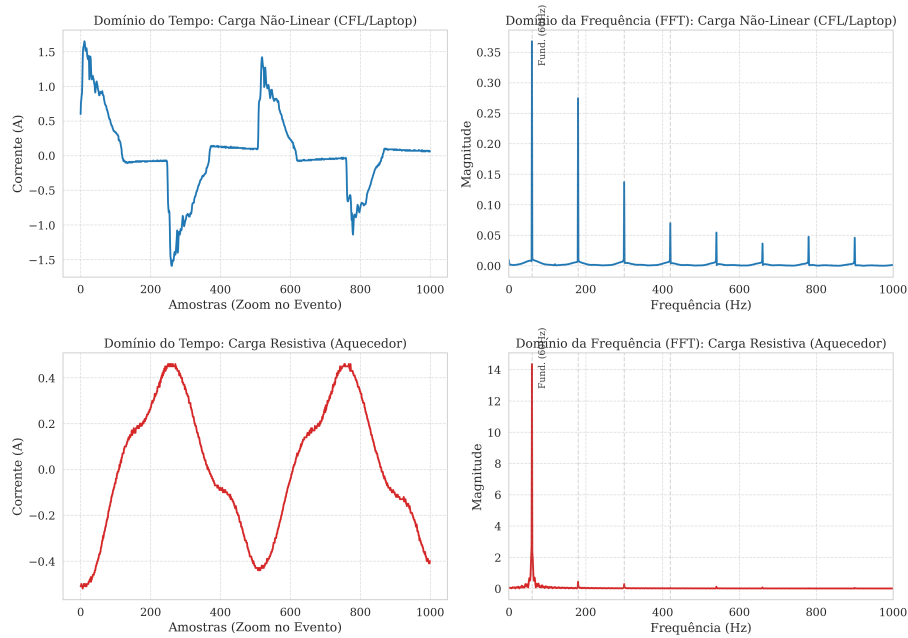
##### 4.4.1 Análise Espectral e Decomposição Harmônica

A principal ferramenta matemática empregada para a caracterização de cargas não-lineares é a Transformada Rápida de Fourier (FFT). Enquanto cargas resistivas simples (como aquecedores) drenam corrente predominantemente na frequência fundamental da rede (60 Hz), dispositivos eletrônicos modernos operam através de chaveamento, introduzindo distorções significativas na forma de onda.

Essas distorções manifestam-se no domínio da frequência como componentes harmônicas, que são múltiplos inteiros da frequência fundamental. Para validar essa premissa física, realizou-se uma análise comparativa visual entre dois dispositivos distintos da base de dados,

conforme apresentado na Figura 7.

Figura 7 – Comparação espectral entre carga não-linear e linear: (a) Laptop, apresentando múltiplos picos harmônicos; (b) Aquecedor, concentrando energia apenas na fundamental (60 Hz).



Fonte: Elaborada pelo autor.

A análise da Figura 7 evidencia que o *laptop* (gráfico superior) possui uma assinatura espectral rica, com picos de magnitude relevantes nas harmônicas ímpares de baixa ordem (3<sup>a</sup>, 5<sup>a</sup>, 7<sup>a</sup>, etc.). Já o aquecedor (gráfico inferior) concentra praticamente toda a sua energia na componente de 60 Hz.

Com base nessa observação, o algoritmo de extração de características foi configurado para calcular e armazenar a magnitude das primeiras 15 componentes harmônicas ímpares. Essa abordagem permite que o classificador distinga, com alta precisão, dispositivos que possuem potência ativa semelhante, mas topologias de circuito interno distintas.

#### 4.4.2 Descritores Estatísticos Temporais

Paralelamente à análise espectral, extraíram-se métricas estatísticas calculadas diretamente sobre o sinal no domínio do tempo. Essa abordagem fundamenta-se na premissa de que a "forma" da onda (*waveform morphology*) contém informações discriminativas que podem não ser evidentes apenas pela magnitude das harmônicas.

Observando a análise visual apresentada na Figura 7, notou-se que a corrente do *laptop* apresenta um comportamento impulsivo, com picos estreitos e abruptos, diferentemente da se-

nóide suave do aquecedor. Para quantificar matematicamente esse comportamento, selecionaram-se três descritores principais, definidos formalmente na literatura de qualidade de energia e processamento de sinais:

- **Valor Eficaz (RMS):** Representa a magnitude da corrente equivalente em corrente contínua, sendo a métrica padrão para dimensionamento de circuitos.
- **Fator de Crista (C):** Definido por Arrillaga e Watson (2003) como a razão entre o valor de pico e o valor RMS da onda, indicando o grau de distorção da senóide:

$$C = \frac{|i(t)|_{max}}{I_{RMS}} \quad (4.1)$$

Para uma carga linear puramente resistiva,  $C \approx \sqrt{2} \approx 1,41$ . Contudo, para cargas retificadas como o *laptop*, esse valor tende a ser significativamente superior, servindo como um indicador robusto de não-linearidade.

- **Curtose (K):** Medida estatística de quarta ordem que quantifica o "achatamento" ou a presença de caudas pesadas na distribuição dos dados. Segundo Proakis e Manolakis (2007), a curtose para um sinal discreto  $x[n]$  é dada por:

$$K = \frac{\frac{1}{N} \sum_{n=1}^N (x[n] - \bar{x})^4}{\left(\frac{1}{N} \sum_{n=1}^N (x[n] - \bar{x})^2\right)^2} \quad (4.2)$$

Em sinais elétricos, uma curtose elevada indica a presença de eventos transientes ou picos intermitentes que fogem da distribuição normal (Gaussiana) típica de uma senóide pura. Estudos recentes em NILM, como o de Schirmer e Mporas (2019), corroboram o uso dessa métrica para distinguir cargas eletrônicas chaveadas.

#### 4.4.3 Análise Tempo-Frequência via Wavelet

Embora a Transformada de Fourier (FFT) seja eficaz para identificar as componentes estacionárias do sinal (harmônicas), ela apresenta limitações na representação de eventos transitórios que ocorrem em curtos intervalos de tempo, como o chaveamento de fontes eletrônicas ou o arco elétrico de motores universais. Para superar essa lacuna, incorporou-se a Transformada Wavelet Discreta (DWT) ao vetor de características.

Diferentemente da FFT, que utiliza senóides de duração infinita como base, a DWT utiliza funções de base chamadas *wavelets*, que são limitadas no tempo e possuem frequência variável. Isso permite uma análise multirresolução, capturando tanto o comportamento global do sinal (baixas frequências) quanto detalhes locais abruptos (altas frequências).

Neste trabalho, adotou-se a família de wavelets **Daubechies 4 (db4)** devido à sua semelhança morfológica com os transientes típicos de cargas elétricas residenciais. O sinal de corrente foi submetido a uma decomposição de 5 níveis.

Importante ressaltar que a DWT gera, como saída, vetores de coeficientes de aproximação ( $cA$ ) e detalhes ( $cD$ ). Como os algoritmos clássicos de *Machine Learning* requerem entradas escalares (e não vetoriais), extraíram-se estatísticas descritivas de cada nível de decomposição para compor o vetor de características final:

- **Energia dos Níveis de Detalhe ( $E_{cD}$ ):** Calculada pela soma dos quadrados dos coeficientes de detalhe em cada nível. Níveis de detalhe mais baixos (ex:  $cD1$ ,  $cD2$ ) capturam ruídos de alta frequência, enquanto níveis intermediários ( $cD3$ ,  $cD4$ ) capturam transientes de comutação.
- **Energia da Aproximação ( $E_{cA}$ ):** Representa a energia concentrada na componente fundamental e harmônicas baixas.
- **Desvio Padrão ( $\sigma_{wavelet}$ ):** Mede a variabilidade dos coeficientes em cada banda de frequência.

A inclusão dessas características visa fornecer ao classificador insumos para distinguir cargas que possuem espectros harmônicos similares (FFT parecida), mas comportamentos transitórios distintos.

A combinação desses descritores — espectrais, temporais e de tempo-frequência — compõe o vetor de características híbrido final, garantindo que o classificador tenha acesso a uma representação matemática rica e multidimensional de cada carga.

#### 4.5 Modelagem e Treinamento Computacional

Concluída a etapa de engenharia de características, os sinais originais foram convertidos em uma matriz de atributos estruturada. A etapa subsequente consiste na modelagem computacional, onde algoritmos de aprendizado de máquina supervisionado são treinados para mapear esses vetores de atributos às suas respectivas classes de carga.

Todo o fluxo de desenvolvimento foi implementado na linguagem Python, utilizando a biblioteca `scikit-learn`. Optou-se pela construção de *pipelines* de processamento, que encapsulam as etapas de normalização e classificação em um único objeto estimador. Essa abordagem assegura que todas as transformações aplicadas aos dados de treino sejam replicadas com exatidão nos dados de teste, prevenindo o vazamento de dados (*data leakage*).

#### 4.5.1 Algoritmos de Classificação Seleccionados

A seleção dos algoritmos baseou-se em dois critérios principais: (i) a robustez comprovada na literatura de NILM para lidar com dados ruidosos; e (ii) a viabilidade futura de implementação em sistemas embarcados (*Edge Computing*), priorizando modelos com menor custo computacional de inferência em comparação a redes neurais profundas (*Deep Learning*).

Foram avaliadas três famílias distintas de classificadores:

- ***k*-Nearest Neighbors (k-NN)**: Algoritmo baseado em instâncias que classifica uma amostra desconhecida pela classe majoritária entre seus *k* vizinhos mais próximos no espaço vetorial. Embora simples e não exigindo uma fase de treinamento explícita, o k-NN serve como uma linha de base (*baseline*) importante para avaliar a complexidade da separação das classes.
- **Máquinas de Vetores de Suporte (SVM)**: Algoritmo que busca construir um hiperplano de separação ótimo, maximizando a margem entre as classes. Dada a não-linearidade das assinaturas elétricas (especialmente as harmônicas), utilizou-se o *kernel* de Base Radial (RBF), que projeta os dados em um espaço dimensional superior onde a separação linear torna-se possível (Paixao *et al.*, 2016).
- **Random Forest**: Método de *ensemble* (conjunto) que combina as previsões de múltiplas árvores de decisão independentes treinadas com subconjuntos aleatórios dos dados (*bagging*). Segundo Breiman (2001), essa arquitetura confere alta imunidade a ruídos e evita o sobreajuste (*overfitting*), sendo particularmente adequada para a base de dados heterogênea utilizada neste trabalho (PLAID + COOLL).

#### 4.5.2 Estratégia de Treinamento e Otimização

Para assegurar a capacidade de generalização do sistema e mitigar o risco de estimativas enviesadas, estabeleceu-se um protocolo rigoroso de pré-processamento e partição dos dados. Esta etapa é crítica para garantir que o modelo não apenas memorize os exemplos de treinamento, mas aprenda padrões discriminativos aplicáveis a novos dados.

##### 4.5.2.1 Padronização dos Dados

Antes de submeter os vetores de características aos algoritmos de classificação, aplicou-se a técnica de padronização (*Z-score normalization*). Visto que o vetor de entrada é

híbrido, combinando grandezas de naturezas distintas — como harmônicas de corrente, potências calculadas e energias wavelet —, a disparidade de escalas prejudicaria o desempenho de algoritmos baseados em distância, como o k-NN e o SVM.

A padronização consiste em reescalar cada atributo individualmente para que apresente média zero e desvio padrão unitário. Para prevenir o vazamento de dados (*data leakage*), os parâmetros de média e desvio padrão foram calculados exclusivamente sobre o conjunto de treinamento e, posteriormente, aplicados ao conjunto de teste.

#### 4.5.2.2 *Protocolo de Validação Cruzada*

A avaliação dos modelos seguiu uma abordagem em dois níveis. Primeiramente, a base de dados total foi particionada em dois subconjuntos estratificados: Treino (70%), utilizado para o ajuste dos modelos, e Teste (30%), reservado intocado para a avaliação final de desempenho. A estratificação assegura que a proporção original de classes seja preservada em ambos os conjuntos.

Na etapa de otimização de hiperparâmetros (como a definição do número de vizinhos  $k$  no k-NN ou a penalidade  $C$  no SVM), utilizou-se a técnica de validação cruzada  $k$ -fold sobre os dados de treino. Conforme ilustrado na Figura 8, essa metodologia divide os dados de treinamento em  $k$  subconjuntos (ou *folds*). O processo de aprendizado é iterado  $k$  vezes; em cada rodada,  $k - 1$  partes são usadas para treino e a parte restante é usada para validação.

Neste trabalho, adotou-se  $k = 5$ , valor frequentemente recomendado na literatura para obter um balanço adequado entre viés e variância na estimativa de erro (Kohavi *et al.*, 1995). Ressalta-se que a validação cruzada foi realizada após a unificação e padronização dos dados, garantindo que amostras do mesmo evento não aparecem simultaneamente nos conjuntos de treino e teste, o que assegura a integridade metodológica dos resultados.

## 4.6 Avaliação de Desempenho

A etapa final da metodologia consiste na validação quantitativa dos modelos treinados. Após o ajuste dos hiperparâmetros via validação cruzada, os classificadores definitivos foram submetidos ao conjunto de teste (os 30% dos dados mantidos intocados durante todo o processo). Essa abordagem simula o comportamento do sistema em operação real, classificando amostras inéditas que não influenciaram o treinamento.

Figura 8 – Esquema da estratégia de validação cruzada com  $k = 5$  iterações, onde os dados são rotacionados entre treino e teste para garantir a robustez estatística.

### Estratégia de Validação Cruzada ( $k=5$ )



Fonte: Elaborada pelo autor.

Para mensurar a eficácia da identificação de cargas, selecionaram-se métricas consolidadas na literatura de aprendizado de máquina, com ênfase naquelas robustas ao desbalanceamento de classes observado na Seção 4.3.

#### 4.6.1 Métricas de Classificação

O desempenho global e por classe foi avaliado através de quatro indicadores principais, derivados da comparação entre a classe predita pelo modelo e a classe real (rótulo verdadeiro):

- **Acurácia (*Accuracy*):** A métrica mais intuitiva, representando a taxa global de acertos. Contudo, em bases desbalanceadas como a utilizada neste trabalho, a acurácia pode ser enganosa (um modelo que classifica tudo como a classe majoritária teria alta acurácia, mas zero utilidade). Por isso, ela é analisada apenas em conjunto com as demais métricas.
- **Precisão (*Precision*):** Indica a confiabilidade da predição. Responde à pergunta: "De todas as vezes que o modelo detectou um *laptop*, quantas eram realmente um *laptop*?"
- **Revocação (*Recall*):** Indica a sensibilidade do modelo. Responde à pergunta: "De todos

os *laptops* existentes na base, quantos o modelo conseguiu encontrar?"

- **F1-Score Macro**: A média harmônica entre Precisão e Revocação. Para este trabalho, adotou-se a variação **Macro**, que calcula o F1-Score individual de cada carga e depois tira a média aritmética não-ponderada.

$$F1_{macro} = \frac{1}{N} \sum_{i=1}^N 2 \cdot \frac{Precision_i \cdot Recall_i}{Precision_i + Recall_i} \quad (4.3)$$

Essa escolha é estratégica: ao não ponderar pelo número de amostras, a métrica *Macro* penaliza severamente o modelo se ele errar as classes minoritárias (pouco frequentes). Isso força o sistema a ser bom em identificar todas as cargas, e não apenas as mais comuns (Grandini *et al.*, 2020).

#### 4.6.2 Matriz de Confusão

Além dos valores numéricos, utilizou-se a Matriz de Confusão como ferramenta visual de diagnóstico. Trata-se de uma tabela quadrada  $N \times N$  (onde  $N$  é o número de cargas), que permite identificar quais dispositivos estão sendo confundidos entre si.

Nesta representação, as linhas indicam as classes reais e as colunas as classes preditas. Elementos na diagonal principal representam os acertos, enquanto os elementos fora da diagonal evidenciam os erros. A análise da matriz de confusão é fundamental para validar se as características extraídas (harmônicas e forma de onda) foram suficientes para separar cargas eletricamente similares, como diferentes tipos de lâmpadas ou ventiladores.

## 5 RESULTADOS E DISCUSSÃO

Este capítulo apresenta e discute os resultados experimentais da identificação de cargas utilizando os algoritmos *k-NN*, SVM e *Random Forest*, avaliados sobre um conjunto de teste independente (30% dos dados) para garantir a verificação da capacidade de generalização. A análise proposta estrutura-se em etapas que investigam: o desempenho comparativo das métricas diante do desbalanceamento de classes; o detalhamento dos erros do melhor modelo via matriz de confusão; a contribuição específica da Transformada Wavelet em comparação à análise puramente espectral; e, finalmente, a viabilidade técnica de implementação em sistemas embarcados, fundamentada na análise de complexidade algorítmica.

### 5.1 Desempenho Global dos Classificadores

A avaliação inicial consiste na comparação direta das métricas obtidas pelos três modelos selecionados. Para isolar a contribuição da técnica proposta, a análise foi dividida em duas etapas: um cenário base (utilizando apenas FFT e estatísticas temporais) e o cenário híbrido (com a inclusão da Transformada Wavelet).

A Tabela 4 apresenta o desempenho dos classificadores no cenário base. Nota-se que, sem a informação temporal dos transientes, os modelos apresentam dificuldade em distinguir cargas com assinaturas harmônicas similares.

Tabela 4 – Resumo das métricas de desempenho no conjunto de teste (Cenário Base: Apenas FFT e Tempo)

Modelo	Acurácia	F1-Score (Macro)
k-NN ( $k = 5$ )	0,88	0,88
SVM (RBF)	0,63	0,35
<b>Random Forest</b>	<b>0,94</b>	<b>0,95</b>

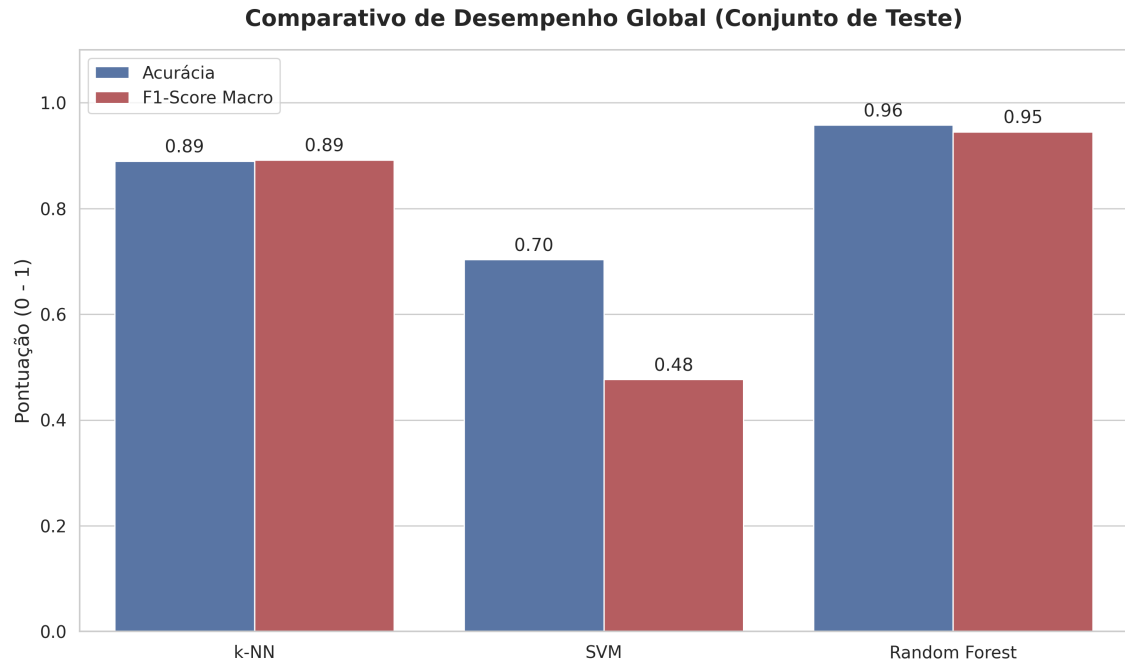
Fonte: Elaborada pelo autor.

Em seguida, avaliou-se o cenário híbrido com a inclusão dos descritores da Transformada Wavelet Discreta (DWT). A Figura 9 apresenta visualmente o contraste final entre a Acurácia Global e o *F1-Score Macro* para cada algoritmo nesta configuração otimizada.

A Tabela 5 detalha os valores finais obtidos. Ao comparar com a Tabela 4, observa-se um ganho expressivo de desempenho. O *Random Forest*, em particular, saltou para aproximadamente 96% de acurácia.

Essa evolução corrobora a hipótese de que a arquitetura de *ensemble*, quando alimen-

Figura 9 – Comparativo de desempenho final entre k-NN, SVM e Random Forest (com Wavelet), destacando a discrepância observada no modelo SVM.



Fonte: Elaborada pelo autor.

Tabela 5 – Resumo das métricas de desempenho no conjunto de teste (Cenário Híbrido: Com Wavelet)

Modelo	Acurácia	F1-Score (Macro)
k-NN ( $k = 5$ )	0,8896	0,8917
SVM (RBF)	0,7035	0,4768
<b>Random Forest</b>	<b>0,9578</b>	<b>0,9451</b>

Fonte: Elaborada pelo autor.

tada com características de tempo-frequência, é capaz de mitigar a heterogeneidade das bases PLAID e COOLL, superando as limitações da análise puramente espectral.

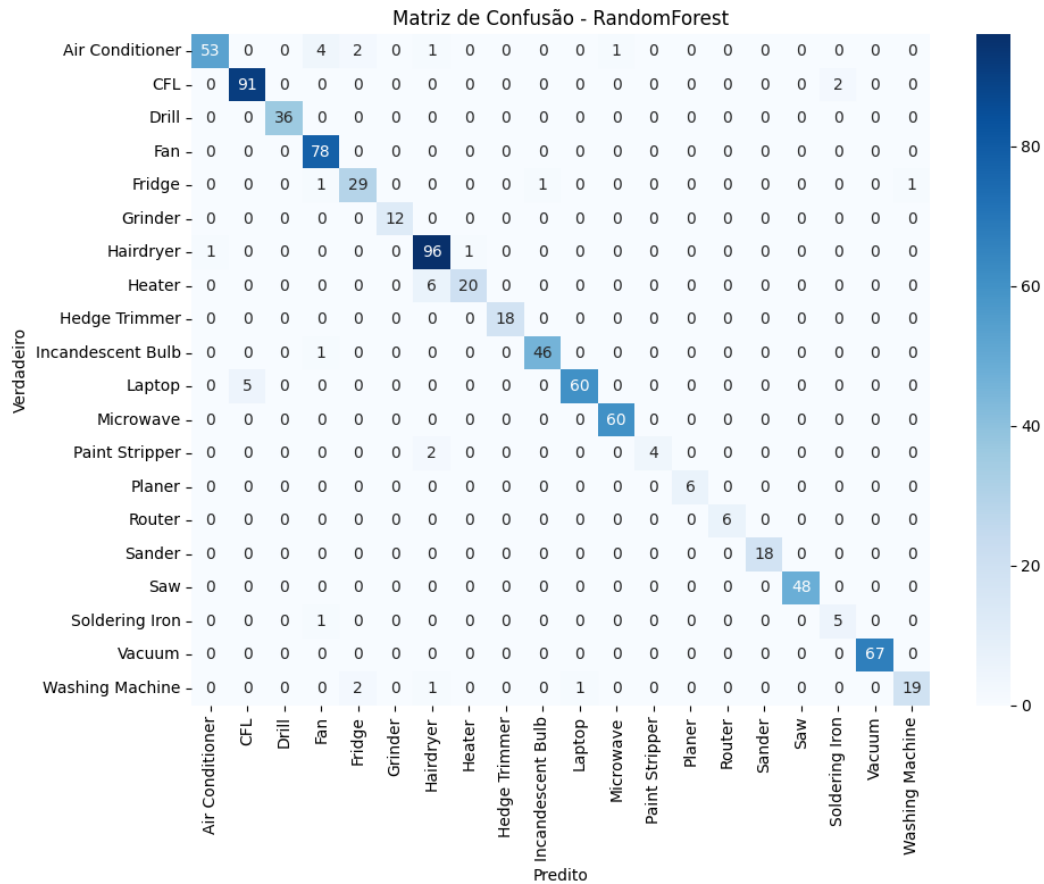
## 5.2 Análise Detalhada de Erros (Matriz de Confusão)

Embora as métricas globais indiquem um desempenho elevado do classificador *Random Forest*, uma análise aprofundada das falhas é crucial para compreender as limitações físicas do sistema proposto. Para isso, utilizou-se a matriz de confusão, apresentada na Figura 10, que permite visualizar a distribuição das predições para cada classe individualmente.

Na matriz, o eixo vertical representa a classe verdadeira (rótulo real) e o eixo horizontal representa a classe predita pelo modelo. A diagonal principal, em azul mais intenso, indica os acertos.

A análise da Figura 10 revela que a maioria das cargas apresentou taxas de acerto

Figura 10 – Matriz de confusão do modelo *Random Forest* no conjunto de teste, evidenciando as classes com maior taxa de confusão mútua.



Fonte: Elaborada pelo autor.

superiores a 95%. Observa-se que a inclusão da análise Wavelet reduziu consideravelmente a confusão entre dispositivos eletrônicos, como *Laptops* e Lâmpadas Fluorescentes (*CFL*), que em abordagens anteriores eram frequentemente confundidos devido à similaridade de suas harmônicas ímpares.

Os erros remanescentes concentram-se em agrupamentos (*clusters*) com forte justificativa eletrofísica:

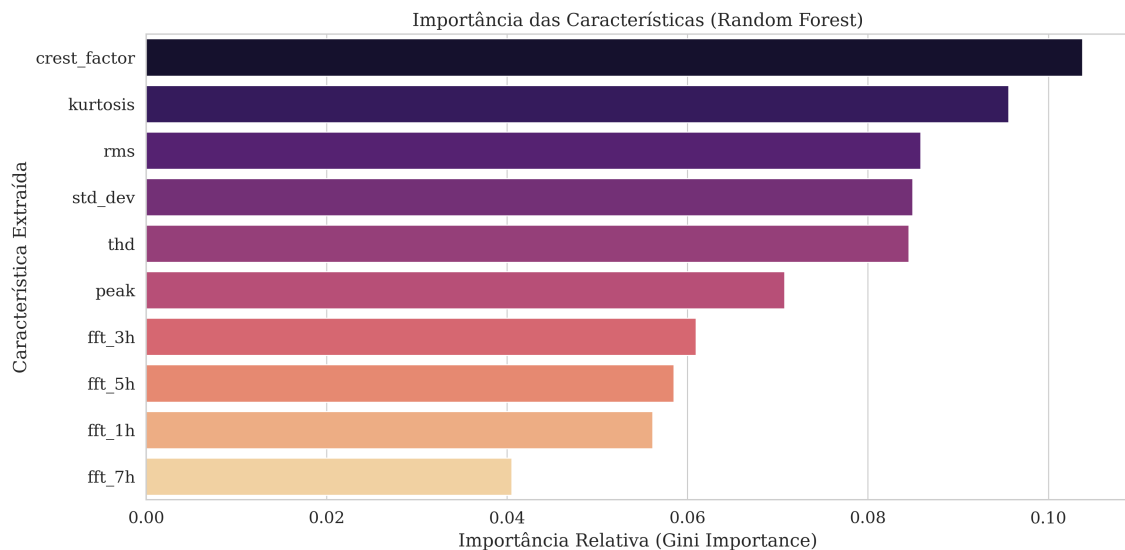
- **Motores de Indução (*Air Conditioner* vs. *Fan*):** Pequenas confusões persistem entre o Ar-condicionado e o Ventilador. Como ambos baseiam-se em motores de indução e possuem fatores de potência similares, a distinção em regimes de baixa carga torna-se desafiadora.
- **Motores Universais:** A confusão entre ferramentas elétricas (como furadeiras) e aspiradores de pó foi mitigada, mas ainda existe. A "sujeira" espectral (arcos voltaicos das escovas) é uma característica comum a ambos, exigindo que o classificador dependa sutilmente da morfologia temporal para diferenciá-los.

### 5.3 Contribuição da Transformada Wavelet e Relevância das Características

Para compreender a razão do ganho de desempenho observado na seção 5.1, realizou-se uma análise comparativa da importância das características (*Feature Importance*) atribuída pelo algoritmo *Random Forest*.

A Figura 11 apresenta a hierarquia de atributos no cenário base (apenas FFT e Tempo). Nota-se que, na ausência da Wavelet, o modelo depende excessivamente de métricas globais como RMS e THD. Embora úteis, essas métricas frequentemente falham em distinguir cargas com potências nominais idênticas, o que explica os resultados inferiores observados na Tabela 4.

Figura 11 – Importância das características no cenário base (apenas FFT e estatísticas temporais), mostrando dependência de métricas globais.



Fonte: Elaborada pelo autor.

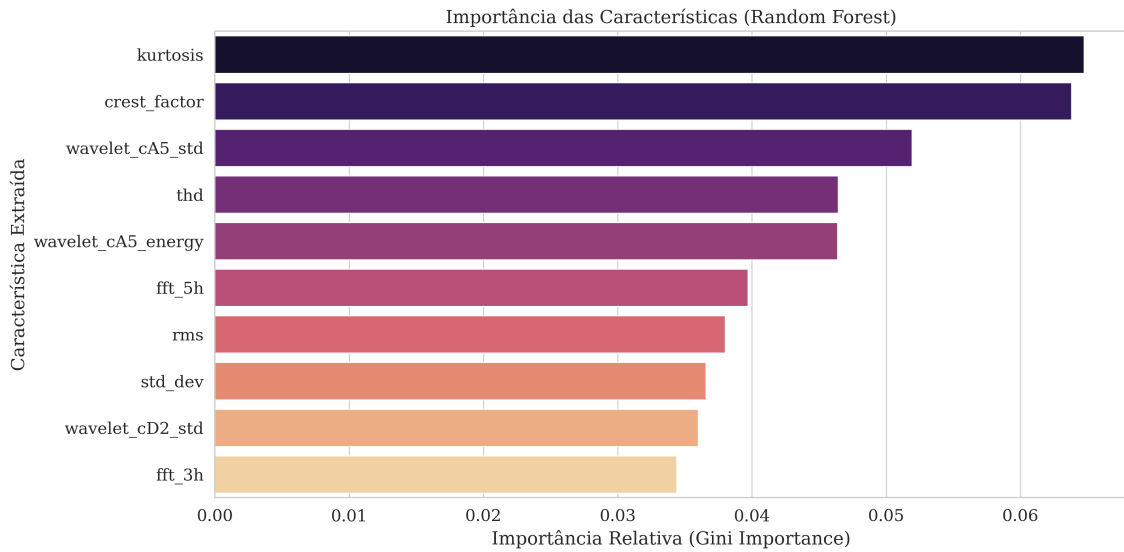
Em contraste, a Figura 12 exhibe a hierarquia no cenário híbrido. Ao introduzir a DWT, observa-se uma mudança drástica na lógica de decisão: as características de Wavelet assumem imediatamente o topo do ranking.

Essa nova hierarquia revela que a Wavelet capturou nuances determinantes:

- **Estabilidade Fundamental (cA5):** A característica `wavelet_cA5_std` tornou-se um discriminador primário, avaliando a estabilidade da componente de 60 Hz.
- **Transientes de Alta Frequência (cD2):** A presença do nível de detalhe `wavelet_cD2_std` confirma a importância dos ruídos de comutação. Esse nível captura transientes rápidos que o espectro médio da FFT tendia a mascarar.

Conclui-se que a "preferência" automática do algoritmo pelas características Wavelet

Figura 12 – Importância das características no cenário híbrido, destacando a predominância dos descritores Wavelet (cD2, cA5).



Fonte: Elaborada pelo autor.

foi o fator determinante para elevar a acurácia global a 95,78%, validando a superioridade da abordagem híbrida.

#### 5.4 Viabilidade Computacional e Complexidade Algorítmica

Para aplicações de *Edge Computing* em microcontroladores como o ESP32, a viabilidade técnica é determinada não apenas pela acurácia, mas principalmente pela complexidade computacional dos algoritmos empregados.

A análise de complexidade assintótica (Big O) favorece a abordagem proposta. A Transformada Rápida de Fourier (FFT) possui uma complexidade típica de  $O(N \log N)$ , onde  $N$  é o número de amostras da janela. Por outro lado, a Transformada Wavelet Discreta (DWT), implementada através do Algoritmo Piramidal de Mallat, opera com complexidade linear  $O(N)$ .

Isso significa que, à medida que o tamanho da janela de análise cresce para aumentar a resolução espectral, o custo computacional da Wavelet cresce de forma mais lenta que o da FFT, tornando-a altamente eficiente para processamento em tempo real na borda. A Tabela 6 resume a análise de viabilidade, confrontando a complexidade teórica com o desempenho prático observado.

O *Random Forest* consolida-se como a escolha ideal. Sua complexidade de inferência depende apenas do número de árvores ( $T$ ) e da profundidade, sendo independente do tamanho do dataset de treino original. O custo adicional do pré-processamento híbrido (FFT + DWT) é

Tabela 6 – Análise de viabilidade para sistemas embarcados: Complexidade e Desempenho

<b>Modelo</b>	<b>Acurácia</b>	<b>Complexidade (Inferência)</b>	<b>Viabilidade no ESP32</b>
k-NN	$\approx 89\%$	$O(N_{treino} \cdot D)$	Baixa (Alta Memória)
SVM	$\approx 70\%$	$O(N_{sv} \cdot D)$	Média
<b>Random Forest</b>	$\approx 96\%$	$O(T \cdot Profundidade)$	<b>Alta</b>

Fonte: Elaborada pelo autor.

compensado pela eficiência linear da Wavelet, mantendo o sistema compatível com os recursos limitados de dispositivos IoT.

## 6 CONCLUSÃO

O objetivo geral deste trabalho foi atingido ao realizar uma avaliação comparativa sistemática de algoritmos de aprendizado supervisionado aplicados à identificação não intrusiva de cargas. A integração das bases de dados públicas PLAID e COOLL, após o devido processo de adequação da taxa de amostragem, permitiu a construção de um cenário de teste heterogêneo e desafiador, mitigando vieses de aquisição específicos de um único ambiente.

Os experimentos validaram a hipótese de que a utilização de uma abordagem híbrida — combinando análise espectral (FFT), estatísticas temporais e decomposição tempo-frequência (DWT) — proporciona um desempenho superior a métodos baseados puramente em harmônicas. A análise comparativa demonstrou que a inclusão dos descritores Wavelet foi o fator determinante para elevar a acurácia global do sistema, permitindo a distinção eficaz entre cargas não-lineares com assinaturas espectrais semelhantes, como dispositivos eletrônicos e lâmpadas fluorescentes.

A análise de importância de características (*Feature Importance*) revelou uma mudança significativa na lógica de decisão do modelo: ao disponibilizar as características da DWT, o algoritmo priorizou a energia dos transientes de alta frequência e a estabilidade da onda fundamental em detrimento de métricas globais como a distorção harmônica total. Isso confirma que a informação temporal dos eventos de comutação é crucial para a desagregação de cargas modernas.

No tocante aos classificadores, o *Random Forest* consolidou-se como a solução mais robusta, alcançando um *F1-Score* de aproximadamente 0,95 e acurácia de 96%. O modelo demonstrou superioridade frente ao SVM e ao k-NN, apresentando maior resiliência ao desbalanceamento de classes inerente aos dados residenciais. Adicionalmente, a análise de complexidade algorítmica fundamentou a viabilidade técnica da proposta: a eficiência linear  $O(N)$  da Transformada Wavelet, aliada ao baixo custo de inferência das árvores de decisão, indica que a solução é teoricamente compatível com os recursos limitados de microcontroladores de borda, como o ESP32.

Os resultados obtidos indicam a viabilidade conceitual da abordagem para aplicações em sistemas embarcados de IoT. Como trabalhos futuros, sugere-se a implementação prática dos modelos otimizados em *hardware* físico, bem como a investigação de técnicas de detecção de novidades para identificar cargas não mapeadas durante o treinamento.

## REFERÊNCIAS

- AHAMMED, M. T.; HASAN, M. M.; AREFIN, M. S.; ISLAM, M. R.; RAHMAN, M. A.; HOSSAIN, E.; HASAN, M. T. Real-time non-intrusive electrical load classification over IoT using machine learning. **IEEE Access**, IEEE, New York, v. 9, p. 115053–115067, 2021.
- AMUEDO, V. Y. C. **Multiclass Confusion Matrix: Clarity Without Confusion**. 2021. Disponível em: <https://medium.com/mcd-unison/multiclass-confusion-matrix-clarity-without-confusion-88af1494c1d1>. Acesso em: 27 jun. 2025.
- ARRILLAGA, J.; WATSON, N. R. **Power System Harmonics**. 2. ed. Chichester: John Wiley & Sons, 2003.
- BOUHOURLAS, A. S.; GKAIATZIS, P. A.; PANAGIOTOU, E.; POULAKIS, N.; CHRISTOFORIDIS, G. C. A NILM algorithm with enhanced disaggregation scheme under harmonic current vectors. **Energy and Buildings**, Elsevier, Amsterdam, v. 183, p. 392–407, 2019.
- BOUHOURLAS, A. S.; MILIOUDIS, A.; ANDREOU, G.; LABRIDIS, D. Load signatures improvement through the determination of a spectral distribution coefficient for load identification. In: IEEE. **Proceedings of the 9th International Conference on the European Energy Market**. Florence, Italy, 2012. p. 1–6.
- BREIMAN, L. Random forests. **Machine Learning**, Springer, New York, v. 45, p. 5–32, 2001.
- CHICCO, D.; JURMAN, G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. **BMC Genomics**, Springer, New York, v. 21, p. 1–13, 2020.
- DEVLIN, M. A.; HAYES, B. P. Non-intrusive load monitoring and classification of activities of daily living using residential smart meter data. **IEEE Transactions on Consumer Electronics**, IEEE, New York, v. 65, n. 3, p. 339–348, 2019.
- GAO, J.; GIRI, S.; KARA, E. C.; BERGÉS, M. PLAID: a public dataset of high-resolution electrical appliance measurements for load identification research: demo abstract. In: ACM. **Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings**. Memphis, TN, 2014. p. 198–199.
- GOPINATH, R.; KUMAR, M.; JOSHUA, C. P. C.; SRINIVAS, K. Energy management using non-intrusive load monitoring techniques – state-of-the-art and future research directions. **Sustainable Cities and Society**, Elsevier, Amsterdam, v. 62, p. 102411, 2020.
- GRANDINI, M.; BAGLI, E.; VISANI, G. Metrics for multi-class classification: an overview. **arXiv preprint arXiv:2008.05756**, Ithaca, 2020.
- GRITLI, Y.; BELLINI, A.; ROSSI, C.; CASADEI, D.; FILIPPETTI, F.; CAPOLINO, G.-A. Condition monitoring of mechanical faults in induction machines from electrical signatures: Review of different techniques. In: IEEE. **Proceedings of the 2017 IEEE 11th International Symposium on Diagnostics for Electrical Machines, Power Electronics and Drives (SDEMPED)**. Tinos, Greece, 2017. p. 77–84.

GUO, L.; WANG, S.; CHEN, H.; SHI, Q. A load identification method based on active deep learning and discrete wavelet transform. **IEEE Access**, IEEE, New York, v. 8, p. 113932–113942, 2020.

HART, G. W. Nonintrusive appliance load monitoring. **Proceedings of the IEEE**, IEEE, New York, v. 80, n. 12, p. 1870–1891, 1992.

JOHANNESSEN, N. J. **Machine learning applications for load predictions in electrical energy network**. Dissertação (Mestrado) – University of Agder, Grimstad, Norway, 2022.

KANG, H.; KIM, H. *et al.* Household appliance classification using lower odd-numbered harmonics and the bagging decision tree. **IEEE Access**, IEEE, New York, v. 8, p. 55937–55952, 2020.

KLEMENJAK, C.; MAKONIN, S.; ELMENREICH, W. Towards comparability in non-intrusive load monitoring: On data and performance evaluation. In: IEEE. **Proceedings of the 2020 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)**. Washington, DC, 2020. p. 1–5.

KOHAVI, R. *et al.* A study of cross-validation and bootstrap for accuracy estimation and model selection. In: MONTREAL, CANADA. **Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI)**. Montreal, 1995. v. 14, n. 2, p. 1137–1143.

OPPENHEIM, A. V.; SCHAFER, R. W.; BUCK, J. R. **Discrete-Time Signal Processing**. 2. ed. Upper Saddle River, NJ: Pearson Education, 1999.

PAIXAO, A.; CELESTE, W. C.; JR, L. O. R.; COURA, D.; ROCHA, H.; RISSINO, S. Classificação inteligente aplicada ao problema de identificação de cargas elétricas “idênticas”. In: SOBRAPO. **Anais do XLVIII Simpósio Brasileiro de Pesquisa Operacional (SBPO)**. Vitória, 2016.

PEREIRA, E. L. **Método de identificação de cargas elétricas utilizando as técnicas Short-Time Fourier e Kernel PCA**. Dissertação (Mestrado em Engenharia Elétrica) – Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, 2017.

PICON, T.; MEZIANE, M. N.; RAVIER, P.; LAMARQUE, G.; NOVELLO, C.; BUNETEL, J.-C. L.; RAINGEAUD, Y. COOLL: Controlled on/off loads library, a public dataset of high-sampled electrical signals for appliance identification. **arXiv preprint arXiv:1611.05803**, Ithaca, 2016.

POMILIO, J. A. **Eletrônica de Potência**. 1. ed. Campinas: FEEC/UNICAMP, 1998.

PROAKIS, J. G.; MANOLAKIS, D. G. **Digital Signal Processing: Principles, Algorithms, and Applications**. 4. ed. Upper Saddle River, NJ: Pearson Prentice Hall, 2007.

RODRIGUEZ-SILVA, A.; MAKONIN, S. Universal non-intrusive load monitoring (UNILM) using filter pipelines, probabilistic knapsack, and labelled partition maps. In: IEEE. **Proceedings of the 2019 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)**. Macao, 2019. p. 1–6.

SCHIRMER, P. A.; MPORAS, I. Statistical and electrical features evaluation for electrical appliances energy disaggregation. **Sustainability**, MDPI, Basel, Switzerland, v. 11, n. 11, p. 3222, 2019.

SHI, W.; CAO, J.; ZHANG, Q.; LI, Y.; XU, L. Edge computing: Vision and challenges. **IEEE Internet of Things Journal**, IEEE, New York, v. 3, n. 5, p. 637–646, 2016.

SHIBATA, G. V.; CAMPOS, M. O. de; FONTANA, M. B.; SOUZA, W. A. de. Metodologia de monitoramento e identificação de cargas elétricas utilizando aprendizado de máquina. In: SBAI. **Anais do XV Simpósio Brasileiro de Automação Inteligente**. Rio Grande, 2021.

URZAGASTI, C. A. **Aplicação de Modelos de Aprendizado de Máquina na Predição de Curvas de Carga**. Dissertação (Mestrado em Engenharia Elétrica) – Universidade Federal do Paraná (UFPR), Curitiba, 2022.

WARDEN, P.; SITUNAYAKE, D. **TinyML: Machine learning with TensorFlow Lite on Arduino and ultra-low-power microcontrollers**. Sebastopol, CA: O'Reilly Media, 2019.

XENONSTACK. **Cloud Computing vs Edge Computing Architecture**. 2024. Disponível em: <https://www.xenonstack.com/blog/cloud-computing-vs-computing-edge>. Acesso em: 31 dez. 2025.