

Reconhecimento de Cargas em Sistemas Residenciais de Gestão de Energia por meio de Algoritmos Baseados em Árvores de Decisão

Thales W. Cabral, Dimas Augusto M. Lemes, Fernando B. Neto, Eduardo R. de Lima, Gustavo Fraidenraich e Luís Geraldo P. Meloni

Resumo— Para o gerenciamento eficiente da demanda energética doméstica, avanços tecnológicos recentes trouxeram equipamentos inovadores como o *Home Energy Management System* (HEMS). Como uma extensão de trabalhos anteriores, este artigo adota novas alternativas na cadeia de treinamento rápido dos modelos no reconhecimento de cargas em HEMS. As contribuições são: (i) a inclusão do *Extreme Gradient Boosting* (XGBoost) no reconhecimento de cargas, (ii) adição de técnicas robustas para extração de características e (iii) novas análises empregando tais técnicas, juntamente com os antigos modelos baseados em árvores. Por fim, os resultados indicam o XGBoost como alternativa vencedora alcançando valores superiores de desempenho em relação aos métodos concorrentes.

Palavras-Chave— Aprendizado de máquina, Reconhecimento de Cargas, XGBoost.

Abstract— For the efficient management of domestic energy demand, technology has carried innovative equipment such as the *Home Energy Management System* (HEMS). As an extension of previous works, this article adopts new alternatives for the fast training chain of models to load recognition in HEMS. The contributions are (i) the inclusion of *Extreme Gradient Boosting* (XGBoost) to load recognition, (ii) the addition of robust techniques for feature extraction, and (iii) new analyses employing these techniques along with the old tree-based models. Finally, the results indicate XGBoost as the winning alternative, achieving superior performance values compared to rivals.

Keywords— Machine learning, Load Recognition, XGBoost.

I. INTRODUÇÃO

A demanda humana por eletricidade tem impulsionado o consumo energético em diferentes setores da sociedade. Com 27% de participação no consumo global de energia, o setor residencial exerce uma influência significativa no consumo total de energia [1]. Para lidar com o aumento da demanda energética nas residências, a evolução tecnológica trouxe consigo equipamentos inovadores capazes de monitorar o consumo doméstico de energia e promover a sustentabilidade ambiental,

Thales W. Cabral, Departamento de comunicações, Universidade Estadual de Campinas, Campinas-SP, e-mail: thaleswulcab@gmail.com, Dimas Augusto M. Lemes, Departamento de comunicações, Universidade Estadual de Campinas, Campinas-SP, e-mail: diguto.mendes@gmail.com, Fernando Bauer, Copel Distribuição S.A., Curitiba-PR, e-mail: fernando.bauer@copel.com; Eduardo R. de Lima, Instituto de Pesquisas Eldorado, e-mail: eduardo.lima@eldorado.org.br, Gustavo Fraidenraich, Departamento de comunicações, Universidade Estadual de Campinas, Campinas-SP, e-mail: gf@decom.fee.unicamp.br, Luís Geraldo P. Meloni, Departamento de comunicações, Universidade Estadual de Campinas, Campinas-SP, e-mail: meloni@unicamp.br. Os autores agradecem ao Instituto Eldorado de Pesquisas e à Copel Distribuição S.A. pelo financiamento desta pesquisa no escopo do projeto P&D Aneel PD-PD-02866-0508/2019.

como é o caso do *Home Energy Management System* (HEMS) [2]. Este sistema permite o monitoramento individualizado do consumo de energia de cada eletrodoméstico, é capaz de alertar os usuários sobre possíveis oportunidades de economia, identificar distúrbios nos aparelhos, bem como gerar relatórios de consumo. Além disso, o HEM permite incorporar funcionalidades adicionais, como técnicas de desagregação de cargas [3]. No entanto, alguns pesquisadores adicionam métodos de reconhecimento de carga para maximizar a sua robustez.

O reconhecimento de carga é o processo de identificação do tipo de dispositivo em operação [4]. Segundo [5], na prática, se o usuário mudar o aparelho conectado à tomada inteligente, um sistema de reconhecimento de carga permite que o HEMS identifique automaticamente o tipo de aparelho ligado à tomada. Na literatura, existem diversas formas de executar o reconhecimento de carga. Em [6], os autores empregaram a *Convolutional Neural Network* (CNN) no reconhecimento de cargas e encontraram uma acurácia de 83,33%. Em [7], uma acurácia de 90,04% foi obtida através do emprego da *Hilbert Transform Long Short-Term Memory* (HT-LSTM). Em [8], os pesquisadores alcançaram 94,40% de acurácia ao utilizar o *Nearest Neighbors* (k -NN) e a *Artificial Neural Network* (ANN). Em [9], os autores empregaram o k -NN e a *Support Vector Machine* (SVM) para efetuar o reconhecimento de cargas e obtiveram um resultado de 95,40% para a acurácia. No entanto, mais análises podem ser realizadas buscando novas alternativas para o reconhecimento de cargas.

Este artigo é uma extensão do trabalho [5], onde modelos baseados em árvores, como *Decision Tree* (DT) e *Random Forest* (RF), são parte de uma cadeia para o treinamento rápido dos modelos, capaz de realizar o reconhecimento de cargas a partir de imagens provenientes de sinais de potência ativa. Além disso, a referência [5] emprega a *Principal Component Analysis* (PCA) como ferramenta de extração de características e, conseqüentemente, redução de dimensionalidade. O presente estudo possui as seguintes contribuições: (i) emprego do *Extreme Gradient Boosting* (XGBoost) para o reconhecimento de cargas, alcançando valores de 95,99% para acurácia, 95,81% de F_1 -Score médio ponderado e 0,9296 para o índice Kappa; (ii) utilização das técnicas PCA, *Independent Component Analysis* (ICA) e *Locally Linear Embedding* (LLE) para a extração de características no reconhecimento de cargas realizado pelo XGBoost e (iii) aplicação das técnicas ICA e LLE no reconhecimento de cargas realizado pelos modelos DT e RF. Os resultados apresentam uma acurácia de 95,44%

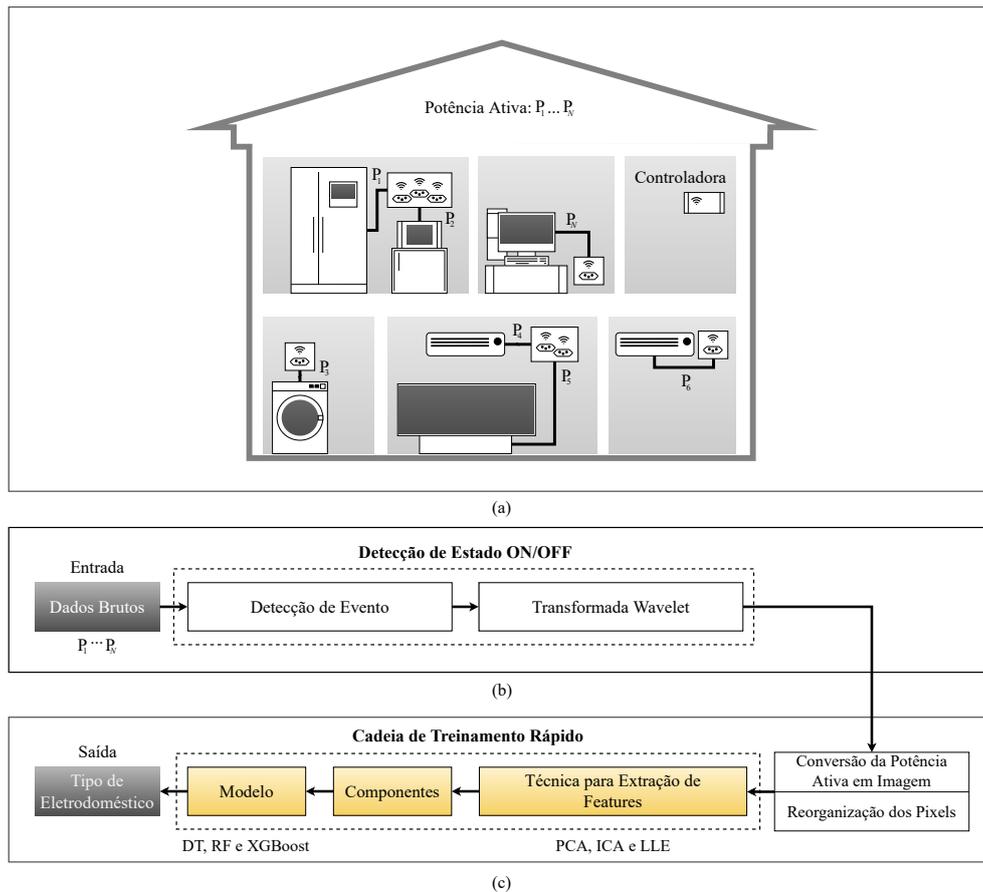


Fig. 1. Visão geral do sistema: (a) medições de potência ativa na aplicação HEMS, (b) detecção da atividade dos aparelhos via transformada Wavelet e (c) reconhecimento de cargas através da cadeia de treinamento rápido (adaptado de [5]).

ao empregar a ICA juntamente com o modelo RF e, além disso, obtivemos acurácias de 95,77% e 95,99% ao utilizar a ICA e a PCA, respectivamente, com o XGBoost. É válido salientar que os valores obtidos são maiores que os dos trabalhos citados anteriormente, inclusive são superiores aos dos modelos baseados em árvore, DT e RF, do trabalho de referência [5]. Por fim, este projeto é realizado através de um consórcio de pesquisa e desenvolvimento envolvendo a Universidade Estadual de Campinas, o Instituto de Pesquisas Eldorado e a Copel Distribuição S.A., que gerou outros trabalhos relacionados ao tema HEMS como [3], [10] e, inclusive, o trabalho de referência [5].

O restante deste trabalho está organizado da seguinte forma: A seção II apresenta as ferramentas necessárias que contemplam o sistema para o reconhecimento de cargas. A seção III apresenta a base de dados e as métricas de desempenho. A seção IV discute os resultados obtidos. Finalmente, a seção V, elenca as principais conclusões.

II. SISTEMA PARA O RECONHECIMENTO DE CARGAS

A Figura 1 ilustra uma visão abrangente do sistema, onde a estratégia de reconhecimento de cargas é responsável por determinar quais os tipos de aparelhos estão em operação numa residência. A Figura 1(a) apresenta um sistema HEMS, que é constituído por uma controladora e tomadas inteligentes.

As tomadas inteligentes são responsáveis por coletar os dados dos eletrodomésticos. Posteriormente, os dados coletados são transmitidos utilizando protocolos de comunicação como Wi-Fi, Bluetooth ou Wireless Smart Ubiquitous Network (Wi-SUN). Por sua vez, a controladora recebe os dados transmitidos. A controladora é formada pelo hardware NXP i.Mx6, conforme [10]. Os dados podem ser analisados localmente ou enviados à nuvem. Ao final, o sistema pode enviar relatórios e alertas para o consumidor. A Figura 1(b) ilustra procedimento para se detectar a operação dos eletrodomésticos através dos estados ON/OFF. Após a detecção, os segmentos com as atividades dos aparelhos são convertidos em imagens [5] e, posteriormente, os pixels são reorganizados, conforme [5]. Por fim, a Figura 1(c) apresenta a cadeia para treinamento rápido dos modelos, onde técnicas para a extração de características, como a PCA, ICA e LLE são empregadas. O presente artigo traz como novidade a utilização das técnicas ICA e LLE na redução de dimensionalidade. Além disso, as classes dos eletrodomésticos são determinadas através de modelos como DT, RF e XGBoost. Neste estágio, este estudo também inova ao empregar o XGBoost para o reconhecimento de cargas.

A. Detecção de Estado ON/OFF e Geração de Imagens

A partir dos sinais de potência ativa como dados brutos, é necessário encontrar os instantes de atividade dos aparelhos

para, posteriormente, determinar as cargas conectadas à rede. Neste momento, é importante salientar que o presente trabalho utiliza somente a potência ativa como única informação acerca dos eletrodomésticos. Para realizar tal tarefa, pode-se empregar técnicas que identificam transições de estado dos eletrodomésticos. Devido a simplicidade de utilização e minimização do tempo de execução, o presente trabalho utiliza a *Discrete Wavelet Transform* (DWT), sobretudo os coeficientes de detalhe. Conforme apresentado no trabalho anterior do nosso grupo, a utilização de coeficientes de detalhe de nível 1 é suficiente para identificar a atividade dos eletrodomésticos. Segundo [11], os coeficientes de detalhe $W_{m,n}$ da DWT são dados por

$$W_{m,n} = \int_{-\infty}^{+\infty} g(t)\psi_{m,n}(t)dt, \quad (1)$$

onde $g(t)$ é um sinal contínuo criado a partir das medições discretas do sinal em estudo, onde as funções de base da transformada $\psi_{m,n}(t)$ oferecem detalhes de escala e posição, descritos por m e n respectivamente. O período de integração de infinito negativo até infinito positivo está relacionado com a história do sinal, intervalo que depende de configuração prática.

Quando surgem os instantes de atividade do eletrodoméstico, como congelador na Figura 2(a), os coeficientes de detalhe de nível 1, conforme a Figura 2(b), apresentam valores superiores a zero. Os picos de maior magnitude, que podem ser observados na Figura 2(b), indicam o início e o fim dos ciclos de atividade de cada aparelho. Este processo produz segmentos de potência ativa dos aparelhos em operação.

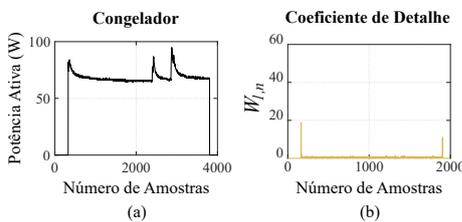


Fig. 2. Detecção de atividade através da potência ativa empregando a transformada de Wavelet.

Após a detecção de eventos, o segmento com a curva de potência ativa é convertido em pixels pretos e brancos, onde os pixels pretos representam o sinal elétrico e os pixels brancos o *background* da imagem. Dessa forma, cada imagem contém um segmento com um ciclo de funcionamento do aparelho. Posteriormente, é preciso reorganizar os pixels da imagem. Neste caso, todas as linhas da imagem são convertidas em um vetor coluna \mathbf{x} de tamanho k . Assim, cada vetor \mathbf{x} representa uma imagem. Logo, para d imagens geradas anteriormente, tal processo produz um conjunto de \mathbf{x} vetores. Ao final, a partir dos vetores \mathbf{x} , cria-se uma matriz $\mathbf{X}_{k \times d}$.

B. Técnicas para Extração de Características

Principal Component Analysis (PCA): A PCA é uma técnica de redução de dimensionalidade capaz de preservar a informação mais relevante contida nos dados [3]. Conforme [5], assume-se como entrada para a técnica a matriz de dados

$\mathbf{X}_{k \times d}$, onde se deseja reduzir a quantidade de pixels. No processo de redução de dimensionalidade, define-se um valor limiar δ de acordo com a *Accumulative Contributory Ratio* (ACR) [3], $\delta = (\sum_{i=1}^q a_i) / (\sum_{i=1}^k a_i)$ para $q < k$, onde $k = 32 \times 32 = 1024$ pixels e a_i são os autovalores. Dessa forma, via ACR, encontra-se os q componentes principais que sintetizam a informação mais relevante dos dados. Utiliza-se aqui o limiar definido em [5], no qual $\delta = 0,99$, onde se encontra $q = 35$, conforme mostra a Figura 3. Mais informações acerca do algoritmo da PCA podem ser encontradas em [12].

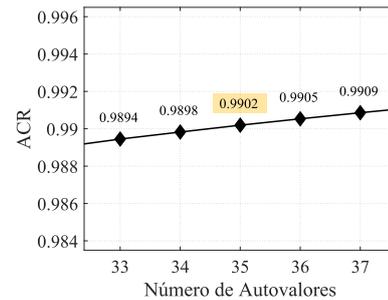


Fig. 3. Contribuição via ACR.

Independent Component Analysis (ICA): A ICA é comumente empregada para aplicações onde se deseja estimar o perfil de fontes de sinais cujas fontes são estatisticamente independentes [13]. No aprendizado de máquina, é possível empregar a ICA para efetuar a extração de características, além de realizar a redução de dimensionalidade dos dados [14]. Para manter um comparativo justo, empregou-se a ICA na matriz de dados de entrada $\mathbf{X}_{k \times d}$, reduzindo também a dimensão dos dados para $q = 35$. Mais informações sobre o algoritmo da ICA estão em [13].

Locally Linear Embedding (LLE): LLE é uma técnica de redução não linear de dimensionalidade, onde características dos dados de entrada residentes em um espaço de alta dimensionalidade são projetadas em espaço de menor dimensionalidade [15]. Neste estudo, LLE mapeia a matriz de entrada $\mathbf{X}_{k \times d}$ para um espaço de dimensão reduzida, com q componentes. Para manter um comparativo justo, mantivemos $q = 35$ para o algoritmo LLE. Os principais passos do algoritmo LLE estão em [15].

C. Modelos de Aprendizado

Decision Tree (DT): DT é um algoritmo de aprendizado supervisionado que define nós de decisão relacionados através de uma hierarquia. Assim, o algoritmo determina rotas ao longo da árvore para obter resultados de classificação. No presente trabalho, efetua-se uma busca de hiperparâmetros via *Grid Search* (GS) com *K-fold Cross-Validation* (K-CV), conforme [5] e o dicionário de hiperparâmetros presente nesta mesma referência, para determinar o valor mais adequado da profundidade máxima da árvore. Ao empregar $K = 10$, juntamente com a PCA, o trabalho [5] obteve um valor de 50 para a profundidade máxima da árvore. Em concordância com o trabalho [5], mantivemos $K = 10$, no entanto, obtivemos valores de profundidade máxima de 70 e 90 para as técnicas ICA e LLE, respectivamente.

Random Forest (RF): O RF é uma arquitetura de *ensemble* que emprega várias árvores em seu processo decisório. Ao final, a maioria dos votos é o resultado do algoritmo. Em sua estrutura, o número de árvores envolvidos na floresta é um dos hiperparâmetros mais importantes do RF [5]. Através do dicionário de hiperparâmetros oriundo de [5], encontrou-se uma quantidade de 60 árvores envolvidas na floresta ao utilizar $K = 10$ com a PCA. Conforme [5], mantivemos $K = 10$, porém, obtivemos valores de 100 árvores aplicando ICA e 10 árvores utilizando LLE.

Extreme Gradient Boosting (XGBoost): Diferentemente do RF, que produz as árvores de decisão e obtém as previsões de forma paralela, o XGBoost é uma arquitetura sequencial [16]. Dessa forma, cada árvore subsequente no XGBoost depende do resultado da árvore anterior [17]. Assim como o RF, neste trabalho, o número de árvores envolvidas na floresta para o XGBoost também deve ser definido. A referência [5] não emprega o XGBoost, porém, mantivemos no GS com K -CV o valor de $K = 10$ e o mesmo dicionário para o número de árvores. Neste caso, encontramos 70 árvores com o emprego da PCA, 40 árvores com a utilização da ICA e 40 com o uso da LLE.

III. BASES DE DADOS E MÉTRICAS DE DESEMPENHO

A. Bases de Dados

Para avaliar a eficiência do estudo proposto, consideramos leituras de potência ativa de eletrodomésticos oriundos da base de dados REDD [18], essas medições foram coletadas em uma frequência de 1/3 Hz. Tal base de dados está disponível na literatura e possui dados de medições reais de eletrodomésticos. Na residência 1, utilizada neste trabalho, o REDD contempla os seguintes elementos: forno, congelador, máquina de lavar louça, forno de cozinha, iluminação, máquina de lavar e secar roupa, micro-ondas, tomada do banheiro, bomba de calor, forno e medição desconhecida.

B. Métricas de Desempenho

Este trabalho emprega acurácia, F_1 -Score médio ponderado (F_1) e índice Kappa (\mathcal{K}) como métricas para avaliação de desempenho. Assumindo TP, TN, FP, e FN, como verdadeiro positivo, verdadeiro negativo, falso positivo e falso negativo, respectivamente, pode-se apresentar a formulação canônica das métricas supracitadas [19], como a acurácia em (2), F_1 em (3) e \mathcal{K} em (4). A acurácia representa a proporção de instâncias corretamente classificadas em relação ao total de instâncias, logo, diz respeito ao desempenho global do modelo. Logo, a acurácia pode ser definida por

$$\text{Acurácia} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (2)$$

É de comum conhecimento que o F_1 -Score é a média harmônica entre precisão e *recall*, entretanto, nem sempre cada aparelho produz o mesmo número de eventos. Para incorporar este efeito empregamos o F_1 -Score médio ponderado [20], o qual nomeamos de F_1 , conforme

$$F_1 = \frac{1}{J} \sum_i i \times \left[\frac{2 \times \text{TP}}{2 \times \text{TP} + 1 \times (\text{FN} + \text{FP})} \right] \quad (3)$$

Por fim, conforme [6], o índice Kappa, o qual nomeamos de \mathcal{K} , mede a concordância entre a predição realizada pelo classificador e o valor esperado. Este índice possui uma escala de $[-1, 1]$, onde -1 indica nenhuma concordância e 1 indica concordância total. Assim, \mathcal{K} pode ser representado por

$$\mathcal{K} = \frac{2 \times (\text{TP} \times \text{TN} - \text{FN} \times \text{FP})}{(\text{TP} + \text{FP}) \times (\text{FP} + \text{TN}) + (\text{TP} + \text{FN}) \times (\text{FN} + \text{TN})} \quad (4)$$

IV. RESULTADOS

Nesta secção, utiliza-se a base de dados REDD nos ensaios. É fundamental ressaltar que após o estágio de detecção de estados ON/OFF, o sistema converte os segmentos detectados em imagens, com resolução de 32×32 pixels. Este processo produz 4609 imagens, as quais foram particionadas em 80% para treinamento e 20% para teste. Os dados de teste foram reservados e apenas dados de treinamento participam da busca de hiperparâmetros via GS com K -CV. Assim, apenas dados de teste participam do computo das métricas de desempenho presentes nas tabelas I, II e III.

A Tabela I apresenta uma comparação entre os modelos DT, RF e XGBoost ao utilizar a PCA para a extração de características. Neste cenário, o XGBoost apresentou os maiores valores para a acurácia, F_1 e \mathcal{K} , com 95,99%, 95,81% e 0,9296, respectivamente. Considerando as métricas acurácia e F_1 , o XGBoost alcançou um ganho de desempenho acima 1,6% em relação aos demais classificadores. Por fim, para a métrica \mathcal{K} , o XGBoost apresentou um ganho acima de 2,4% em relação às arquiteturas DT e RF.

TABELA I
COMPARATIVO ENTRE OS MODELOS COM O EMPREGO DA PCA.

Modelo	Acurácia	F_1	\mathcal{K}
DT	94,14% (ref. [5])	94,11% (ref. [5])	0,9015 (ref. [5])
RF	94,36% (ref. [5])	94,12% (ref. [5])	0,9050 (ref. [5])
XGBoost	95,99%	95,81%	0,9296

A Tabela II aborda a comparação entre os modelos DT, RF e XGBoost ao empregar a ICA no processo de extração de características. Neste caso, o XGBoost novamente apresentou os maiores valores para a acurácia (95,77%), F_1 (95,57%) e \mathcal{K} (0,9257). Ao utilizar a ICA, o XGBoost obteve ganhos de desempenho aproximados acima de 0,3%, 1,5% e 0,5% para acurácia, F_1 e \mathcal{K} , respectivamente, quando comparado com os modelos DT e RF.

TABELA II
COMPARATIVO ENTRE OS MODELOS COM O EMPREGO DA ICA.

Modelo	Acurácia	F_1	\mathcal{K}
DT	94,03%	94,06%	0,8955
RF	95,44%	94,97%	0,9199
XGBoost	95,77%	95,57%	0,9257

A Tabela III traz a comparação entre os modelos DT, RF e XGBoost ao aplicar a técnica LLE no para a extração de características. Mais uma vez, o XGBoost apresentou os maiores valores para a acurácia (94,03%), F_1 (93,67%) e \mathcal{K} (0,8954). Ao utilizar a técnica LLE, o XGBoost obteve ganhos

de desempenho acima de 1,8%, 1,7% e 3,2%, para acurácia, F_1 e \mathcal{K} , respectivamente, em relação aos demais modelos.

TABELA III

COMPARATIVO ENTRE OS MODELOS COM O EMPREGO DA LLE.

Modelo	Acurácia	F_1	\mathcal{K}
DT	91,65%	91,39%	0,8534
RF	92,19%	91,94%	0,8631
XGBoost	94,03%	93,67%	0,8954

As tabelas IV, V e VI, apresentam o tempo gasto para encontrar os hiperparâmetros, o tempo necessário para treinar os modelos e o tempo de inferência, ambas em segundos. A partir da Tabela IV, verifica-se que o menor tempo gasto na busca de hiperparâmetros encontra-se na junção entre DT e LLE, ou seja, no par DT-LLE (2,843 segundos). Porém, o maior tempo para esta tarefa encontra-se no par XGBoost-ICA (296,881 segundos), que representa aproximadamente 104 vezes o tempo gasto pelo par DT-LLE. Por outro lado, conforme a Tabela V, o menor tempo necessário de treinamento é obtido pelo par RF-LLE (0,044 segundos) e o maior tempo é alcançado pelo par XGBoost-PCA (10,176 segundos), ou seja, o par XGBoost-PCA é aproximadamente 231 vezes o tempo gasto pelo par RF-LLE. Por fim, conforme a Tabela VI, o menor tempo de inferência é alcançado pelo par DT-LLE (0,0004 segundos) e o maior tempo pelo par XGBoost-PCA (3,4190 segundos), ou seja, este resultado é 8547 vezes o tempo gasto pelo par DT-LLE. Assim, observa-se que o desempenho superior do XGBoost é obtido em detrimento de tempo de busca, treinamento e inferência mais prolongados em comparação às outras arquiteturas.

TABELA IV

TEMPO EM SEGUNDOS PARA A BUSCA DE HIPERPARÂMETROS.

Modelo	PCA	ICA	LLE
DT	6,587 (ref. [5])	3,575	2,843
RF	25,827 (ref. [5])	33,415	19,47
XGBoost	296,487	296,881	264,889

TABELA V

TEMPO EM SEGUNDOS PARA O TREINAMENTO.

Modelo	PCA	ICA	LLE
DT	0,162 (ref. [5])	0,052	0,047
RF	0,334 (ref. [5])	0,734	0,044
XGBoost	10,176	6,806	2,314

TABELA VI

TEMPO EM SEGUNDOS PARA A INFERÊNCIA.

Modelo	PCA	ICA	LLE
DT	0,0008 (ref. [5])	0,0007	0,0004
RF	0,0214 (ref. [5])	0,0206	0,0027
XGBoost	3,4190	2,5409	2,1324

V. CONCLUSÃO

Como uma extensão do trabalho [5], este artigo adota novas alternativas na cadeia para treinamento rápido dos modelos no reconhecimento de cargas, como a inclusão do XGBoost e das técnicas de extração de características PCA, ICA e LLE, além dos testes dos modelos DT e RF com ambas as técnicas ICA e

LLE. Destacam-se os resultados de acurácia do RF com ICA (95,44%), do XGBoost com a ICA (95,77%) e do XGBoost com a PCA (95,99%). Além disso, o par XGBoost-PCA também valores superiores para F_1 (95,81%) e \mathcal{K} (0,9296). No entanto, o desempenho superior do XGBoost requer um maior tempo de busca de hiperparâmetros, de treinamento e de inferência. Por fim, conclui-se que o XGBoost, bem como as técnicas ICA e LLE, são alternativas bastante promissoras para aplicações no reconhecimento de carga de alto desempenho em sistemas residenciais de gestão de energia.

REFERÊNCIAS

- [1] T.-Y. Kim and S.-B. Cho, "Predicting residential energy consumption using cnn-lstm neural networks," *Energy*, vol. 182, pp. 72–81, 2019.
- [2] B. Mahapatra and A. Nayyar, "Home energy management system (hems): Concept, architecture, infrastructure, challenges and energy management schemes," *Energy Systems*, pp. 1–27, 2019.
- [3] D. A. M. Lemes, T. W. Cabral, G. Fraidenraich, L. G. P. Meloni, E. R. De Lima, and F. B. Neto, "Load disaggregation based on time window for hems application," *IEEE Access*, vol. 9, pp. 70746–70757, 2021.
- [4] A. Faustine and L. Pereira, "Multi-label learning for appliance recognition in nilm using fryze-current decomposition and convolutional neural network," *Energies*, vol. 13, no. 16, p. 4154, 2020.
- [5] T. W. Cabral, D. A. M. Lemes, G. Fraidenraich, F. B. Neto, E. R. de Lima, and L. G. P. Meloni, "High-reliability load recognition in home energy management systems," *IEEE Access*, vol. 11, pp. 31244–31261, 2023.
- [6] L. Matindife, Y. Sun, and Z. Wang, "Image-based mains signal disaggregation and load recognition," *Complex & Intelligent Systems*, vol. 7, no. 2, pp. 901–927, 2021.
- [7] S. Heo, H. Kim, et al., "Toward load identification based on the Hilbert transform and sequence to sequence long short-term memory," *IEEE Transactions on Smart Grid*, vol. 12, no. 4, pp. 3252–3264, 2021.
- [8] S. Mian Qaisar and F. Alsharif, "Signal piloted processing of the smart meter data for effective appliances recognition," *Journal of Electrical Engineering & Technology*, vol. 15, no. 5, pp. 2279–2285, 2020.
- [9] S. M. Qaisar and F. Alsharif, "Event-driven system for proficient load recognition by interpreting the smart meter data," *Procedia Computer Science*, vol. 168, pp. 210–216, 2020.
- [10] L. C. Ferreira, A. D. R. Borchardt, G. D. S. Cardoso, D. A. M. Lemes, G. R. D. R. de Sousa, F. B. Neto, E. R. de Lima, G. Fraidenraich, P. Cardieri, and L. G. P. Meloni, "Edge computing and microservices middleware for home energy management systems," *IEEE Access*, vol. 10, pp. 109663–109676, 2022.
- [11] P. Addison, "Wavelet transforms and the ecg: A review," *Physiological measurement*, vol. 26, pp. R155–99, 11 2005.
- [12] J. Shlens, "A tutorial on principal component analysis," *arXiv preprint arXiv:1404.1100*, 2014.
- [13] E. Oja and A. Hyvarinen, "Independent component analysis: algorithms and applications," *Neural networks*, vol. 13, no. 4-5, pp. 411–430, 2000.
- [14] M. Fang, Y. Yu, W. Zhang, H. Wu, M. Deng, and J. Fang, "High performance computing of fast independent component analysis for hyperspectral image dimensionality reduction on mic-based clusters," in *2015 44th International Conference on Parallel Processing Workshops*, pp. 138–145, 2015.
- [15] B.-Y. Sun, X.-M. Zhang, J. Li, and X.-M. Mao, "Feature fusion using locally linear embedding for classification," *IEEE transactions on neural networks*, vol. 21, no. 1, pp. 163–168, 2009.
- [16] D. Zhang, L. Qian, B. Mao, C. Huang, B. Huang, and Y. Si, "A data-driven design for fault detection of wind turbines using random forests and xgboost," *Ieee Access*, vol. 6, pp. 21020–21031, 2018.
- [17] R. Abedi, R. Costache, H. Shafizadeh-Moghadam, and Q. B. Pham, "Flash-flood susceptibility mapping based on xgboost, random forest and boosted regression trees," *Geocarto International*, vol. 37, no. 19, pp. 5479–5496, 2022.
- [18] J. Kolter and M. Johnson, "Redd: A public data set for energy disaggregation research," *Artif. Intell.*, vol. 25, 01 2011.
- [19] A. Géron, *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. "O'Reilly Media, Inc.", 2022.
- [20] Y. Guo, Q. Song, M. Jiang, Y. Guo, P. Xu, Y. Zhang, C.-C. Fu, Q. Fang, M. Zeng, and X. Yao, "Histological subtypes classification of lung cancers on ct images using 3d deep learning and radiomics," *Academic radiology*, vol. 28, no. 9, pp. e258–e266, 2021.