Autoencoders Beat PCA for Low-Dimension DGA-based Fault Diagnosis of Power Transformers

Thales W. Cabral, Eduardo R. de Lima, José Cândido S. Santos Filho, and Luís Geraldo P. Meloni

***Abstract*— Energy utility companies are investing in advanced monitoring systems using efficient data processing methodologies to mitigate the impacts of power transformer malfunctions on supply stability. Reducing data processing volume is crucial for achieving efficiency. In this context, our contributions include (i) proposing a fault diagnosis system that maintains high performance even under severe dimensionality reduction, (ii) introducing two Autoencoder structures, (iii) conducting pioneering tests of the Adafactor optimizer in dissolved gas analysis using Autoencoders, and (iv) comparing our solution with Principal Component Analysis (PCA), one of the most well-established techniques in the literature. Results confirm that our proposed system outperforms PCA, particularly in scenarios requiring severe dimensionality reduction.**

***Keywords*— Machine learning, Autoencoders, Dimensionality reduction, DGA, Fault diagnosis.**

I. Introduction

Instability in energy supply due to malfunctioning equipment in electrical substations has encouraged energy utilities to work for a more stable and environmentally sustainable electricity supply. Faults in vital equipment, such as power transformers, can reduce the efficiency of the electrical system or render it inoperable [1]. The concern to prevent such scenarios motivates distribution companies to invest in monitoring systems and data processing, focusing on solutions for diagnosing transformer faults [2].

Dissolved Gas Analysis (DGA) relies on seven types of gases to diagnose the condition of power transformers: hydrogen (H_2), methane (CH_4), ethane (C_2H_6), ethylene (C_2H_4), acetylene (C_2H_2), carbon monoxide (CO), and carbon dioxide (CO_2) [3]. Different circumstances generate these gases. The corona effect produces H_2, oil decomposition at low temperatures produces CH_4 and C_2H_6, and oil decomposition at high temperatures produces C_2H_4. Electric arcs produce C_2H_2, while insulation paper decomposition produces CO and CO_2.

Traditionally, DGA-based fault diagnosis of power transformers used interpretation techniques [4]. But most modern, state-of-the-art DGA solutions employ Machine Learning (ML) techniques [5]. Efficient decision-making mechanisms include in their structures architectures such as Support Vector Machines (SVM) [5], k-Nearest Neighbor (k-NN) [6], and Artificial Neural Networks (ANN) [7].

Thales W. Cabral, Department of Communications, University of Campinas, Campinas-SP, e-mail: t264377@dac.unicamp.br; Eduardo R. de Lima, Eldorado Research Institute, e-mail: eduardo.lima@eldorado.org.br; José Cândido S. Santos Filho, Department of Communications, University of Campinas, Campinas-SP, e-mail: jcssf@unicamp.br; and Luís Geraldo P. Meloni, Department of Communications, University of Campinas, Campinas-SP, e-mail: meloni@unicamp.br. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

However, despite the advancements in ML, practical challenges still affect monitoring systems. Companies face high operating costs regarding the installation and maintenance of these systems, particularly in terms of data processing. Therefore, there is a practical demand for methods that reduce the amount of data. Furthermore, some gaps in fault diagnosis strategies remain unexplored, especially in the phase before the decision-making stage, where techniques to reduce data dimensionality can be applied.

This work is motivated by the practical need to reduce the data dimensionality that feeds the decision-maker stage of a system for fault diagnosis in power transformers. This system is under development as part of a research project that entails cooperation between the University of Campinas and the Eldorado Research Institute. Our key objective is to guarantee a good balance between performance and high dimensionality reduction. To this end, this study proposes a modern approach based on Autoencoders (AEs) [8] and ML models for fault diagnosis via DGA. We use AEs due to their potential to enhance class separability by learning intrinsic data patterns. For this task, we propose two variants of AE architectures: one with intermediate layers and the other without. On the other hand, for the decision-maker stage, we choose models whose low computational complexity contributes to the system structure. For this purpose, we employ the ML models SVM, k-NN, and Extreme Learning Machine (ELM) [9]. This study has the following main contributions: (i) proposing a fault diagnosis system that maintains high performance even under severe dimensionality reduction; (ii) introducing two AE architectures aimed at low-complexity DGA data processing; (iii) conducting pioneering tests of the Adafactor optimizer [10] in fault diagnosis using AEs; and (iv) establishing a fair comparison between the proposed AE-based method for fault diagnosis and a classical, well-established dimensionality reduction technique, namely Principal Component Analysis (PCA). Through this comparison, we seek to answer critical questions: Can the proposed AE-based approach preserve system performance under severe dimensionality reduction? Can it compete with a widely recognized technique like PCA? As confirmed in Section IV, the AE-based approach indeed surpasses PCA in most performance metrics when severe dimensionality reduction is pursued.

The remainder of this paper is organized as follows. Section II presents the proposed approach for fault diagnosis in power transformers. Section III presents the performance metrics. Section IV discusses the results. Finally, Section V outlines the main conclusions.

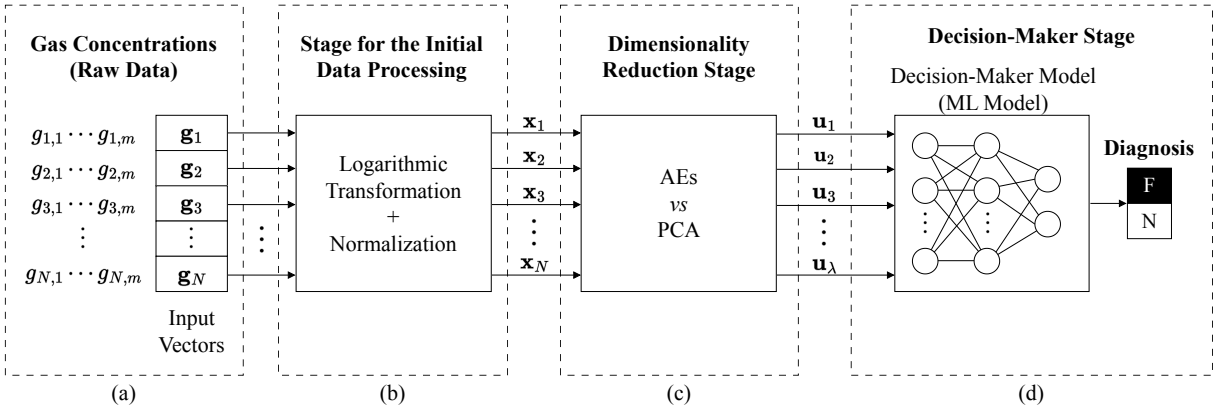

Fig. 1. Overview of the proposed system for fault diagnosis in power transformers. The system begins with (a) the concentrations of seven gases as input. Next, (b) the data treatment stage processes the input data. Then, (c) dimensionality reduction is performed using either AEs or PCA. Finally, (d) the decision-maker stage uses ML models to provide the fault diagnosis.

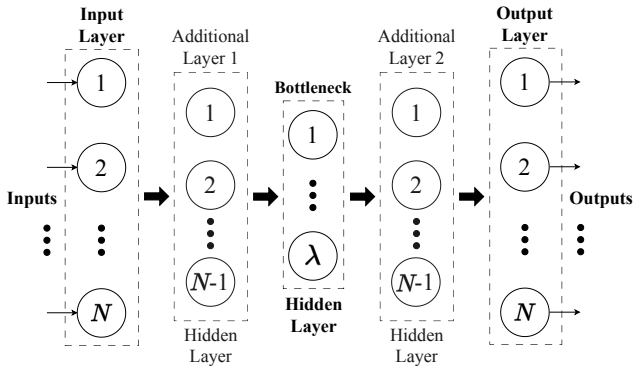

Fig. 2. Proposed structure for Autoencoder 2 (AE2). Autoencoder 1 (AE1) has a single hidden layer (bottleneck).

II. Proposed Approach for Fault Diagnosis in Power Transformers

According to Fig. 1, our proposal encompasses a complete system for fault diagnosis in power transformers. At first, as per Fig. 1(a), the gas concentrations feed the input of this system. In this phase, the system uses m samples of the N gas concentrations as input vectors: $g_{1,1} \cdots g_{1,m} \triangleq \mathbf{g}_1$, $g_{2,1} \cdots g_{2,m} \triangleq \mathbf{g}_2$, $g_{N,1} \cdots g_{N,m} \triangleq \mathbf{g}_N$. As depicted in Fig. 1(b), the system initiates the data treatment phase. As mentioned in [6], the magnitude order is one of the features highly informative regarding gas concentrations. Re-scaling the data using a logarithmic transformation is one approach to providing the AI model with information on variations in magnitude order. Then, given a sample g, this procedure generates the new sample z through $z \triangleq \log_{10}(g)$. In addition to this procedure, we apply standardization to avoid convergence problems with AI models, ensuring numerically stable operations. In this case, we use the following normalization procedure: $x \triangleq (z - \mathbb{E}[z])/\sqrt{\mathbb{V}[z]}$, where $\mathbb{E}[\cdot]$ is the expectation operator and $\mathbb{V}[\cdot]$ is the variance operator. This way, we obtain $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$, $\cdots$, $\mathbf{x}_N$. Then, the system sends the processed data to the stage dedicated to dimensionality reduction, according to Fig. 1(c). Here, we propose the implementation of AEs due to their capability to enhance separability between classes insofar as they learn patterns intrinsic to the data. At this stage, we introduce two variants of AE structures. In the structure outlined in Fig. 2 (Autoencoder 2, AE2), we propose an AE with a bottleneck (the hidden layer for dimensionality control) and two additional layers, one positioned between the input layer and the bottleneck and another between the bottleneck and the output layer. Both representations consider a number N of neurons for the input layer. Autoencoder 1 (AE1, not shown) is a variant of AE2 without these two additional layers. Furthermore, in both AEs, the dimensionality reduction, reflected in the size of the bottleneck (denoted by λ), depends upon the extent to which we aim to compress the data dimensionality. For AE1, λ can range from N to one, and for AE2, from $N-1$ to one. A smaller value indicates a more pronounced reduction in dimensionality.

Fig. 3 illustrates the dimensionality reduction procedure for both AEs, where we plug the autoencoder bottleneck directly (without decoder) into the decision-maker model. In this manner, the AE feeds the decision-maker stage with the transformed data (with dimensionality λ). It is worth mentioning that, in this study, we tested two different optimizers for the AEs, the well-known Adam and one of its current rivals — the Adafactor. The dimensionality reduction procedure via AEs is detailed in Algorithm 1. For benchmarking purposes, we also tested the classic PCA as an alternative technique to AEs. In this case, we varied the data dimensionality (λ) from the maximum (seven) to the minimum (one) to evaluate this technique under severe conditions of dimensionality reduction.

Next, according to Fig. 1(d), the transformed data feeds the decision-making stage. Here, we employ ELM, k-NN, and SVM models, i.e., models whose low computational complexity contributes to the system structure. The process of optimizing the ML models is described in Algorithm 2. In the end, the system diagnoses the status of the power transformer as faulty (F) or normal (N). To improve the performance of the models, we can find some of their optimal hyperparameters. In the ELM model, we can seek the appropriate number of neurons in the hidden layer for the ELM model, $h^{\text{ELM}}_{(\text{optimal})}$. Similarly, we can explore the optimum number of neighbors for the k-NN, $h^{k\text{-NN}}_{(\text{optimal})}$. In the SVM, we can search for

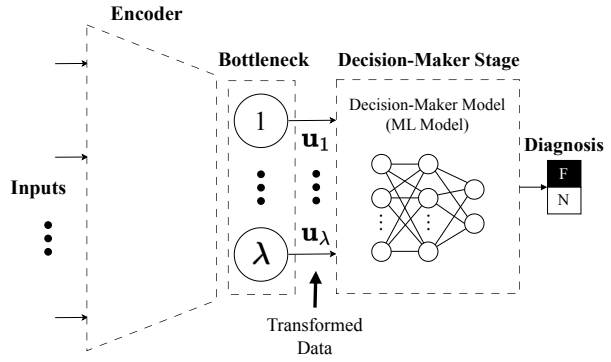

Fig. 3. Dimensionality reduction procedure using Autoencoders.

Algorithm 1 Approach for dimensionality reduction via AEs

Input: Desired AE architecture (AE1 or AE2), bottleneck size (λ), optimizer (Adam or Adafactor), number of training epochs (ψ), amount of gases involved in monitoring via DGA (N), treated data ($\mathbf{x}_1$, $\mathbf{x}_2$, $\cdots$, $\mathbf{x}_N$).

Output: $\{\mathbf{u}_{1,i}, \mathbf{u}_{2,i}, \mathbf{u}_{3,i}, \cdots, \mathbf{u}_{\lambda,i}\}_{i=1}^{q}$ and $\{\mathbf{u}_{1,i}, \mathbf{u}_{2,i}, \mathbf{u}_{3,i}, \cdots, \mathbf{u}_{\lambda,i}\}_{i=1}^{p}$.

1: first method:
Split the samples of treated data ($\mathbf{x}_1$, $\mathbf{x}_2$, $\cdots$, $\mathbf{x}_N$) into a set of q training samples: $\{\mathbf{x}_{1,i}, \mathbf{x}_{2,i}, \mathbf{x}_{3,i}, \cdots, \mathbf{x}_{N,i}\}_{i=1}^{q}$, and a set of p test samples: $\{\mathbf{x}_{1,i}, \mathbf{x}_{2,i}, \mathbf{x}_{3,i}, \cdots, \mathbf{x}_{N,i}\}_{i=1}^{p}$.
2: second method:
Select the AE structure (AE1 or AE1).
3: third method:
Load the parameters N, λ, ψ, and optimizer (Adam or Adafactor) for the selected AE structure.
Apply dimensionality reduction:
Control the size of the bottleneck via the λ value.
4: fourth method:
Train the selected AE employing the loaded parameters and applying the training set $\{\mathbf{x}_{1,i}, \mathbf{x}_{2,i}, \mathbf{x}_{3,i}, \cdots, \mathbf{x}_{N,i}\}_{i=1}^{q}$ to the input and output of the model.
5: fifth method:
Collect the set of transformed data from the bottleneck, i.e., $\{\mathbf{u}_{1,i}, \mathbf{u}_{2,i}, \mathbf{u}_{3,i}, \cdots, \mathbf{u}_{\lambda,i}\}_{i=1}^{q}$.
6: sixth method:
Testing the AE using the test set $\{\mathbf{x}_{1,i}, \mathbf{x}_{2,i}, \mathbf{x}_{3,i}, \cdots, \mathbf{x}_{N,i}\}_{i=1}^{p}$.
7: seventh method:
Collect the set of transformed data from the bottleneck: $\{\mathbf{u}_{1,i}, \mathbf{u}_{2,i}, \mathbf{u}_{3,i}, \cdots, \mathbf{u}_{\lambda,i}\}_{i=1}^{p}$.
return $\{\mathbf{u}_{1,i}, \mathbf{u}_{2,i}, \mathbf{u}_{3,i}, \cdots, \mathbf{u}_{\lambda,i}\}_{i=1}^{q}$ and $\{\mathbf{u}_{1,i}, \mathbf{u}_{2,i}, \mathbf{u}_{3,i}, \cdots, \mathbf{u}_{\lambda,i}\}_{i=1}^{p}$.

the optimum value for the regularization hyperparameter C, $h^{\text{SVM}}_{(\text{optimal})}$. For this purpose, we apply the Grid Search with the K-fold Cross-Validation technique [11].

III. Numerical Results and Discussion

A. Evaluation Metrics

This work uses accuracy [9], F_1-Score [11], and the Jaccard index to evaluate the proposed system. The joint use of these metrics ensures a holistic comprehension of the results, providing a more robust and fair analysis. Accuracy offers a comprehensive notion of the success rate regarding the approach; we use this metric as accuracy $= (\text{TP} + \text{TN})/(\text{TP} + \text{FP} + \text{TN} + \text{FN})$, where TP is true-positive, TN is true-negative, FP is false-positive, and FN is false-negative. F_1-Score provides a more balanced view of the model's performance, especially if there is an imbalance between the classes. We employ F_1-Score as per $F_1\text{-Score} = (2 \times \text{TP})/[2 \times \text{TP} + 1 \times (\text{FN} + \text{FP})]$. Jaccard index is defined as $\text{Jaccard} = (\text{TP})/(\text{TP} + \text{FP} + \text{FN})$. The Jaccard index measures the similarity between model predictions and expected values [12], useful for overlapping classes, and it ranges from 0 (no match) to 1 (full match).

Algorithm 2 Optimization of Decision-Maker Models

Input: Selected ML model (θ), set of j candidates hyperparameters for the ML model: $\{h_i^\theta\}_{i=1}^{j}$, $\{\mathbf{u}_{1,i}, \mathbf{u}_{2,i}, \mathbf{u}_{3,i}, \cdots, \mathbf{u}_{\lambda,i}\}_{i=1}^{q}$, $\{\mathbf{u}_{1,i}, \mathbf{u}_{2,i}, \mathbf{u}_{3,i}, \cdots, \mathbf{u}_{\lambda,i}\}_{i=1}^{p}$, number of folds (K).

Output: Fault Diagnosis

1: first method:
Load θ and load $\{h_i^\theta\}_{i=1}^{j}$.
2: second method:
Split the q samples of $\{\mathbf{u}_{1,i}, \mathbf{u}_{2,i}, \mathbf{u}_{3,i}, \cdots, \mathbf{u}_{\lambda,i}\}_{i=1}^{q}$ in K folds and apply the Grid Search with K-fold Cross-Validation to obtain the optimal hyperparameter $h^\theta_{(\text{optimal})}$.
3: third method:
Train the optimized model with $h^\theta_{(\text{optimal})}$ and test them with $\{\mathbf{u}_{1,i}, \mathbf{u}_{2,i}, \mathbf{u}_{3,i}, \cdots, \mathbf{u}_{\lambda,i}\}_{i=1}^{p}$.
return Fault Diagnosis

B. Results and Analysis

Most DGA studies use few samples due to the limited availability of data in the literature, as in [1], [13], [14]. The present work employs dataset fusion to minimize this situation. We implement a fusion of the databases from [3], [4], [13]–[17]. As a result, we obtain a partial discharges category (PD) with 209 samples, an energy discharges category (D) with 206 samples, a thermal faults category (T) with 90 samples, and a normal status category (N) with 46 samples. Subsequently, we relocate these samples into normal (N) or fault (F) for binary classification.

It is worth mentioning that, from the fusion of samples, the proposed approach sets 70% of the data for training and 30% for testing, where, in Algorithm 2, only the training set takes part in the Grid Search with the K-fold Cross-Validation. For each dimension (seven to one), there is a search for hyperparameters considering the dimensionality reduction technique (PCA or AEs), the ML model used (θ), and its set of associated hyperparameters ($\{h_i^\theta\}_{i=1}^{j}$). For each hyperparameter search, we used the following sets associated with the models: $\{h_i^{\text{ELM}}\}_{i=1}^{4} = \{1, 10, 100, 1000\}$, $\{h_i^{\text{SVM}}\}_{i=1}^{4} = \{1, 10, 100, 1000\}$, and $\{h_i^{k\text{-NN}}\}_{i=1}^{10} = \{1, 2, 3, \ldots, 10\}$. These values were chosen to ensure the convergence of the search without significantly compromising computational time. For

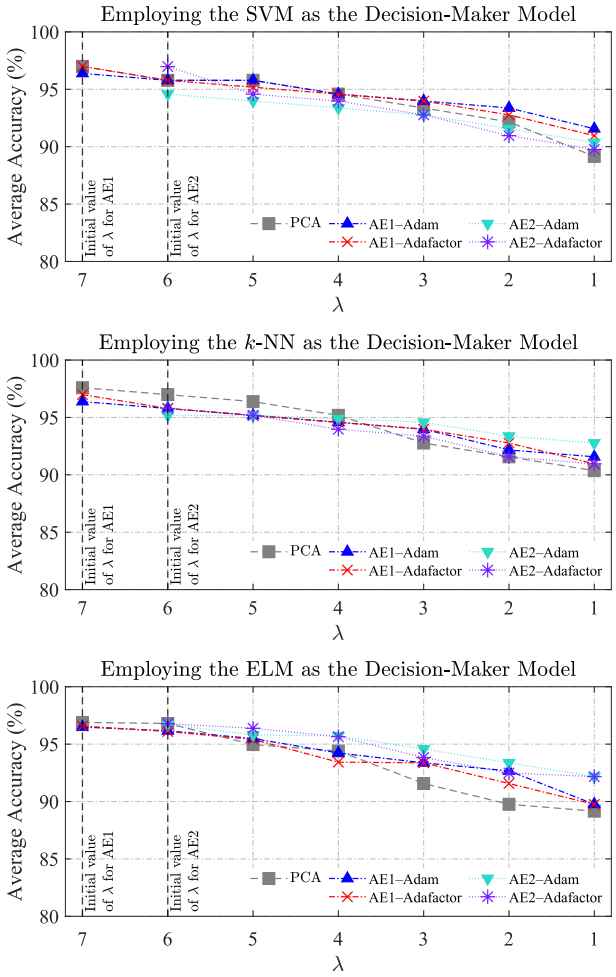

Fig. 4. Average accuracy for SVM, k-NN, and ELM models.

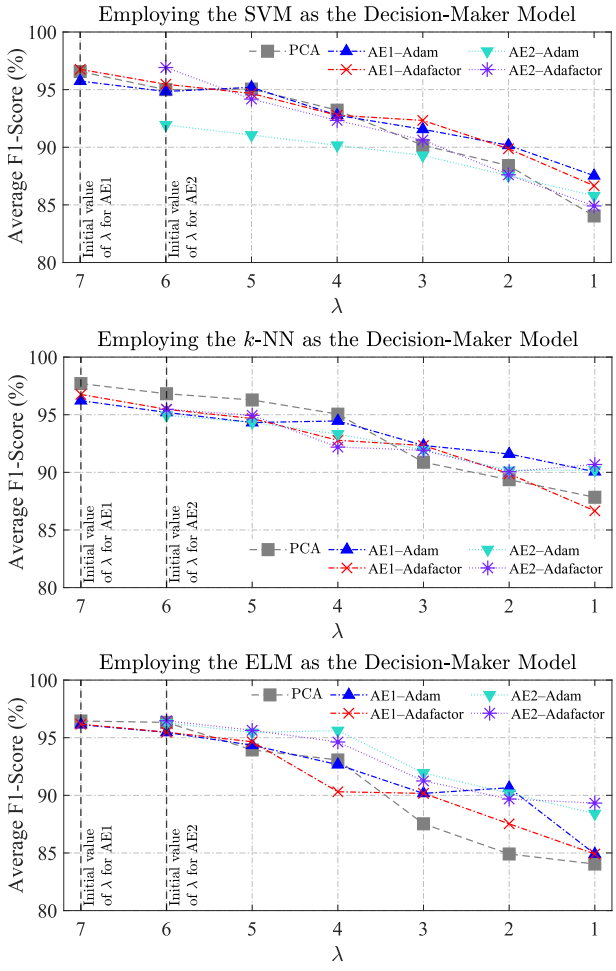

Fig. 5. Average F1-Score for SVM, k-NN, and ELM models.

brevity, we present the optimal hyperparameters corresponding to the winning models, i.e., the optimal hyperparameters whose models reach the best results (the highest percentage points advantages). From Algorithm 2, we find $h^{\text{SVM}}_{(\text{optimal})} = 1$, $h^{k\text{-NN}}_{(\text{optimal})} = 5$, and $h^{\text{ELM}}_{(\text{optimal})} = 100$. In addition, for both autoencoders using both optimizers, we employ a learning rate according to [10] and 50 training epochs, where each epoch takes approximately 7 milliseconds. It is also worth noting that all results derive from 50 iterations, with the average evaluation metrics being shown in Figs. 4, 5, and 6. In interpreting these findings, we equate a variation of under 0.5 percentage point between the methodologies as a tie. In this study, all the decision-maker models presented an average training time of less than 0.004 s.

Fig. 4 depicts the average accuracy results. When employing ELM, AEs tie with PCA for dimensionalities λ = 7 and 6. However, the AE approach prevails for the remaining λ values, 5, 4, 3, 2, and 1. Moreover, for ELM, our proposed method outperforms PCA significantly in scenarios with severe dimensionality reduction (λ = 3, 2, and 1), achieving the best result for λ = 2, where AE2-Adam exhibits the highest advantage, with 3.61 percentage points. Utilizing k-NN, the AE approach outperforms PCA once again in scenarios with severe dimensionality reduction, λ = 3, 2, and 1, and ties when λ = 4 and 7. Here, for k-NN, we achieve the best result in the extreme case, λ = 1, where AE2-Adam shows the highest advantage, with 2.41 percentage points. When using SVM, the AE approach outperforms PCA for λ = 6, 2, and 1. The AE aproach ties for λ = 7, 5, 4, and 3. In this case, the AE1-Adam approach achieves the best result for λ = 1, with a 2.41 percentage point advantage.

Fig. 5 displays the results for the average F1-Score. In this context, when employing ELM, AEs tie with PCA for dimensionalities λ = 7 and 6, while outperforming it for λ = 5, 4, 3, 2, and 1. Our approach achieves the best result for λ = 2, where AE1-Adam exhibits the highest advantage, with 5.75 percentage points. For k-NN, the AE approach ties with PCA for λ = 4 but prevails for λ = 3, 2, and 1, where the best result, via AE2-Adafactor and λ = 1, shows an advantage of 2.84 percentage points. Regarding SVM, the proposed approach ties with PCA for λ = 7, 5, and 4, but has superior performance for λ = 6, 3, 2, and 1, where the best result, via AE1-Adam with λ = 1, exhibits an advantage of 3.49 percentage points.

Fig. 6 illustrates the results for the average Jaccard index. Through ELM, the proposed method prevails for λ = 5, 4, 3, 2, and 1, where our approach achieves the best result for λ = 2 via AE2-Adam, with a notable 6.15 percentage point advantage over PCA. For k-NN, our proposed method demon-

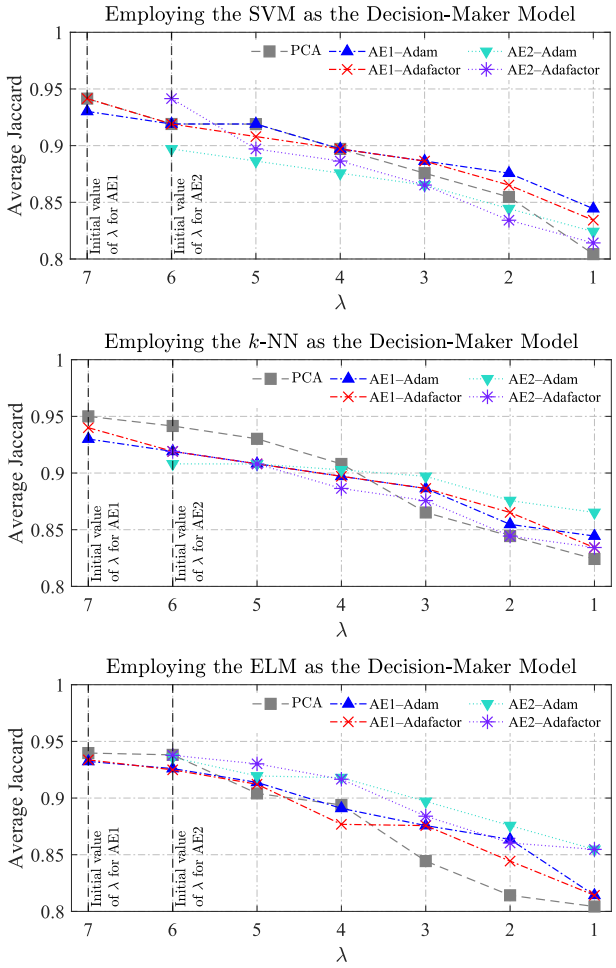

Fig. 6. Average Jaccard index for SVM, k-NN, and ELM models.

strates superiority for λ values of 3, 2, and 1, showcasing a 4.1 percentage point advantage when employing AE2-Adam. Employing SVM, our approach wins once again for λ = 3, 2, 1, under severe dimensionality reduction. In this case, for $\lambda = 1$, AE1-Adam achieves a 4.01 percentage point advantage over PCA.

Finally, when evaluating the extreme case (λ = 1), among all the results, the approach via AE with k-NN appears with the highest accuracy, F_1-Score and Jaccard values: 92.77% with AE2-Adam, 90.68% with AE2-Adafactor, and 0.8652 with AE2-Adam, respectively.

IV. Conclusion

In this manuscript, we presented an approach for diagnosing faults in power transformers via DGA, combining dimensionality reduction using AEs with ML models. We introduced two AE structures, conducted a pioneering test of the Adafactor optimizer in DGA using AEs, and compared our strategy with PCA. As validated in Section III, the AE methodology demonstrates robustness, particularly in scenarios of severe dimensionality reduction (λ = 1, 2, and 3), where it consistently exhibits the most substantial percentage point advantages across all metrics compared to PCA. Specifically, AE1 presents these advantages over PCA for accuracy and Jaccard index through Adam optimizer, SVM, and λ = 1, and for F1-Score via ELM with λ = 2. In turn, AE2 shows such advantages for accuracy and Jaccard index through Adam, ELM, and λ = 2, and for F1-Score via Adafactor, kNN, and λ = 1. In the extreme case (λ = 1), our AE-based approach with k-NN achieves the highest values in accuracy, F_1-Score and Jaccard index: 92.77% accuracy with AE2-Adam, 90.68% F_1-Score with AE2-Adafactor, and 0.8652 Jaccard index with AE2-Adam. Overall, we demonstrated that autoencoders beat PCA under severe dimensionality reduction while exhibiting comparable performance under slight to moderate dimensionality reduction. These findings underscore the potential of autoencoders as versatile tools for enhancing the accuracy of DGA-based fault diagnosis in power transformers.

References

[1] K. Bacha, S. Souahlia, and M. Gossa, "Power transformer fault diagnosis based on dissolved gas analysis by support vector machine," *Electric power systems research*, vol. 83, no. 1, pp. 73–79, 2012.

[2] F. M. Laburú, T. W. Cabral, F. V. Gomes, E. R. de Lima, J. C. Filho, and L. G. Meloni, "New insights into gas-in-oil-based fault diagnosis of power transformers," *Energies*, vol. 17, no. 12, p. 2889, 2024.

[3] L. Ganyun, C. Haozhong, Z. Haibao, and D. Lixin, "Fault diagnosis of power transformer based on multi-layer svm classifier," *Electric Power Systems Research*, vol. 74, no. 1, pp. 1–7, 2005.

[4] M. Duval and A. DePabla, "Interpretation of gas-in-oil analysis using new IEC publication 60599 and iec tc 10 databases," *IEEE Electrical Insulation Magazine*, vol. 17, no. 2, pp. 31–41, 2001.

[5] X. Zhang, G. Zhang, P. Paul, J. Zhang, T. Wu, S. Fan, and X. Xiong, "Dissolved gas analysis for transformer fault based on learning spiking neural p system with belief adaboost.," *International Journal of Unconventional Computing*, vol. 16, 2021.

[6] P. Mirowski and Y. LeCun, "Statistical machine learning and dissolved gas analysis: a review," *IEEE Transactions on Power Delivery*, vol. 27, no. 4, pp. 1791–1799, 2012.

[7] M. E. A. Senoussaoui, M. Brahami, and I. Fofana, "Combining and comparing various machine-learning algorithms to improve dissolved gas analysis interpretation," *IET Generation, Transmission & Distribution*, vol. 12, no. 15, pp. 3673–3679, 2018.

[8] J. Zhai, S. Zhang, J. Chen, and Q. He, "Autoencoder and its various variants," in *2018 IEEE international conference on systems, man, and cybernetics (SMC)*, pp. 415–419, IEEE, 2018.

[9] T. W. Cabral, F. B. Neto, E. R. d. Lima, G. Fraidenraich, and L. G. Meloni, "Load recognition in home energy management systems based on neighborhood components analysis and regularized extreme learning machine," *Sensors*, vol. 24, no. 7, p. 2274, 2024.

[10] N. Shazeer and M. Stern, "Adafactor: Adaptive learning rates with sublinear memory cost," in *International Conference on Machine Learning*, pp. 4596–4604, PMLR, 2018.

[11] 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.

[12] A. A. Taha and A. Hanbury, "Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool," *BMC medical imaging*, vol. 15, pp. 1–28, 2015.

[13] S. S. Ghoneim and I. B. Taha, "A new approach of dga interpretation technique for transformer fault diagnosis," *International Journal of Electrical Power & Energy Systems*, vol. 81, pp. 265–274, 2016.

[14] S. I. Ibrahim, S. S. Ghoneim, and I. B. Taha, "Dgalab: an extensible software implementation for dga," *IET Generation, Transmission & Distribution*, vol. 12, no. 18, pp. 4117–4124, 2018.

[15] O. E. Gouda, S. Salem, and S. H. El-Hoshy, "Power transformer incipient faults diagnosis based on dissolved gas analysis," *TELKOMNIKA Indonesian J. Electr. Eng*, vol. 17, no. 1, pp. 10–16, 2016.

[16] M. Duval, "A review of faults detectable by gas-in-oil analysis in transformers," *IEEE electrical Insulation magazine*, vol. 18, no. 3, pp. 8–17, 2002.

[17] M.-H. Wang, "A novel extension method for transformer fault diagnosis," *IEEE transactions on power delivery*, vol. 18, no. 1, pp. 164–169, 2003.