

ORIGINAL RESEARCH

Open access

Comparative Assessment of Stacked Autoencoder Networks and Multi-Layer Extreme Learning Machines for Transmission Line Fault Detection and Classification Using WAMS Data

Anna Svensson^{1*}, Erik Lindberg¹

Abstract

Modern smart grids aim to deliver exceptional reliability and operational flexibility in power delivery. Contemporary electrical networks rely on synchronized wide-area monitoring tools, notably Phasor Measurement Units (PMUs), to supply instantaneous system status updates. These devices support prompt problem identification and corrective actions while ensuring uninterrupted grid operation. The surge in available measurement data has accelerated the adoption of machine learning techniques for tasks such as power system protection and state estimation. This work presents an intelligent fault detection and classification (FD&C) framework for transmission lines that employs an auto-encoder neural network architecture. The study evaluates and contrasts the capabilities of a Multi-Layer Extreme Learning Machine (ML-ELM) with those of a Stacked Auto-Encoder (SAE) neural network. These models are further benchmarked against established classifier approaches that rely on wavelet-based feature extraction, as well as against various other deep learning strategies. The proposed auto-encoder solutions attain perfect (100%) fault detection rates and achieve fault classification accuracies of 99.92% and 99.79%, all while demanding far less computation time during the testing phase. The ML-ELM variant stands out for its efficiency, maintaining excellent classification performance with both training and testing durations below 50 milliseconds. To reflect practical field conditions, the models were rigorously tested on datasets incorporating additive noise levels corresponding to signal-to-noise ratios (SNR) between 10 dB and 40 dB. Validation was performed using simulated data from the IEEE 39-bus benchmark system.

Keywords Smart grid, Machine learning, Fault detection, Fault classification, Auto-encoder, Transmission line

*Correspondence:

Anna Svensson
anna.svensson@angstrom.uu.se

¹ Department of Intelligent Energy Systems, Faculty of Engineering, Uppsala University, Uppsala, Sweden

Introduction

Advancements in digital and intelligent technologies are steadily modernizing electricity networks worldwide, pushing them toward greater intelligence and adaptability. In essence, a smart grid constitutes an advanced, robust, durable, and environmentally conscious electricity infrastructure [1]. Phasor Measurement Units (PMUs), which operate with precise GPS time synchronization, are now commonly installed throughout today's power systems. The expansion of these units has led to the accumulation of enormous volumes of wide-area measurement system (WAMS) data. Mining and interpreting this data using data analytics or machine learning methods enables operators to uncover critical insights into grid status and ongoing incidents [2].

The rising integration of power electronic devices and distributed energy resources has increased the incidence of relay misoperations [3]. Protective relays assigned to transmission networks play a pivotal role in safeguarding system integrity. Any fault that remains undetected or uncleared for an extended duration poses substantial risks to overall grid stability. For this reason, relay thresholds are deliberately set with high sensitivity to capture even subtle anomalies. Nevertheless, such tuning can occasionally trigger unwarranted activations during distant disturbances or during heavy grid loading.

Power network configurations evolve constantly as new lines are commissioned and distributed generation units are added. Consequently, a relay calibrated for one specific topology may inadvertently trip under altered operating scenarios. A large proportion of widespread outages stems from cascading sequences initiated by failures or inappropriate responses in primary protection schemes [4].

Effective protection of transmission corridors is fundamental to realizing a self-restoring and resilient smart grid [5]. Sophisticated diagnostic tools powered by artificial intelligence are indispensable for constructing dependable smart grid architectures. Broad-area situational awareness enhances both the resilience and operational trustworthiness of the grid [1]. Techniques drawn from data mining, when applied to WAMS datasets, facilitate the identification, pinpointing, and typing of short-circuit events on transmission pathways [2].

Machine learning approaches for fault detection and classification that incorporate a dedicated feature extraction phase

The success of any machine learning solution in this domain hinges primarily on its ability to identify and exploit informative patterns in the raw inputs. Well-designed feature extraction procedures are therefore indispensable for building fast and resource-efficient models. System reliability, in turn, depends heavily on the rapidity with which faults are recognized, categorized, and isolated so that service can be promptly restored.

Existing machine learning solutions for power system fault detection and classification (FD&C) typically rely on either supervised or unsupervised learning. One study [6] introduces a diagnostic method for high-voltage lines that combines traveling-wave analysis with Discrete Wavelet Transform Multi-Resolution Analysis (DWT-MRA). Three-phase currents captured by current transformers are decomposed, followed by the d-q-axis transformation of the detail coefficients. A fault is declared once the transformed coefficient surpasses a fixed threshold associated with a particular bus. This strategy requires simultaneous readings from both line terminals and relies on static threshold settings, which can become invalid when variable distributed generation alters system behavior.

Another contribution [7] applies the DB4 wavelet to derive entropy measures from voltage waveforms. Specific entropy values derived from detail and approximation coefficients at selected decomposition levels are fed into a probabilistic neural network, yielding 98.27% classification and localization accuracy. Yet the method addresses only individual lines equipped with dual-end metering.

Further work [8] decomposes three-phase currents via wavelet packet analysis using the db6 wavelet. It employs Shannon entropy from early frequency sub-bands as inputs to a radial basis function neural network for diagnosis. Costa *et al.* [9] explore Maximum Overlap Discrete Wavelet Transform (MODWT) energy features computed with the Db4 wavelet, enabling fault plane analysis for both detection and coordinate estimation.

In Anand and Affijulla [10], the Hilbert-Huang transformation converts measured currents and voltages into analytical forms, with fault presence determined via a relay-based energy threshold. Unfortunately, the authors omit quantitative metrics for accuracy or timing, and the threshold must be recalculated after any network reconfiguration—a clear drawback in dynamic grids with distributed sources.

Independent Component Analysis on voltage signals is proposed for location purposes in de Alencar *et al.* [11]. Ekici [12] merges Support Vector Machines with Wavelet Transform applied to single-ended transient recordings for simultaneous

type identification and distance estimation. A purely data-driven decision tree scheme is presented by Jamehbozorg and Shahrtash [13], while He *et al.* [14] process one-terminal measurements via wavelet decomposition of currents and deploy multiple Rough Membership Neural Network classifiers. Timing performance is not disclosed.

AsghariGovar *et al.* [15] tackle high-impedance and cross-country faults using dual-end current data, wavelet packet features, and an Extreme Learning Machine classifier.

Transmission lines remain among the most vital yet vulnerable elements of any power network; swift fault handling is therefore non-negotiable for sustained reliability and robustness. Growing distributed generation leads to frequent topological shifts, rendering traditional methods that depend on fixed parameters or known configurations increasingly impractical. Measurement-centric, data-driven intelligent algorithms emerge as the most viable option for achieving near-instantaneous responses.

Nevertheless, the bulk of published data-driven FD&C models still depend on manually engineered features generated by classical signal processing tools, including the Fourier, Stockwell, Wavelet, and Hilbert-Huang transforms. Such preprocessing demands specialized expertise and considerable processing overhead. In time-critical protection applications, it is far more advantageous to allow the model itself to discover pertinent representations straight from lightly conditioned measurement data. This bypasses the labor-intensive process of feature crafting and reduces both development effort and runtime complexity.

Machine learning approaches for fault detection and classification with built-in automatic feature learning

Solutions that operate effectively on minimally processed or raw signals are especially attractive for achieving minimal-latency diagnostics. A Hidden Markov Model is utilized for classification by Arouche Freire *et al.* [16], whereas Chen *et al.* [17] combine fault typing and localization within a single Extreme Learning Machine framework. Both investigations, however, limit evaluation to simplified single-line test systems; scalability to realistic multi-machine networks is not addressed.

An unsupervised convolutional sparse auto-encoder (CSAE) is explored by Chen *et al.* [18] for automatic feature discovery. In that setup, voltage and current time series from a basic two-source line configuration are first rendered as grayscale images before being presented to the network. The resulting high-dimensional input naturally inflates computational demands.

Liang *et al.* [19] developed a deep belief network to recognize overhead line faults and assign them to one of nine causal categories, trained on historical tripping logs and traveling-wave records. Separately, Omar *et al.* [20] investigate a Long Short-Term Memory (LSTM) recurrent network for classification, drawing its training cases from detailed simulations of faults placed along a 400 kV, 100 km line.

Fault detection and classification using WAMS data

With the rising deployment of Phasor Measurement Units (PMUs) throughout power networks, Wide-Area Measurement System (WAMS) data has become a valuable resource for fault detection and classification (FD&C) research. One study [21] developed an approach that processes data from a single PMU placed at a generator bus. It applies Park's transformation to the three-phase voltage and current phasors to derive equivalent phase angles, then employs a Support Vector Machine (SVM) for classification. However, the authors did not evaluate the method's computational performance, particularly the time needed for training, testing, and feature extraction.

In another work [22], features consisting of equivalent current phase angles (ECPA) derived from phase currents, along with their phaselet coefficients, were used to train a Gaussian Naive Bayes (GNB) classifier. This technique required PMU

measurements from both terminals of a transmission line. In practice, installing PMUs at both ends of every line is not economically viable, as utilities typically place them only at selected optimal buses [23].

Zhang *et al.* [24] transformed PMU data into polar plots and utilized a Convolutional Neural Network (CNN) for fault classification, with a combined preprocessing and classification time of 10 seconds. A different PMU-based backup protection scheme [25] calculated equivalent admittances between PMU buses in regions exhibiting the highest voltage distortion and compared them to a predefined threshold. This method, however, omitted fault-type classification and did not assess fault-location performance.

A wide-area backup protection framework based on WAMS data was introduced in Harish *et al.* [26], using a Particle Swarm Optimized Weighted Extreme Learning Machine (PSO-WELM) to handle an ensemble of features. Fault detection via three variants of CNN models was examined by Alqudah *et al.* [27], yielding excellent AUC and accuracy results. Yet, the detection speed was not reported, and the required data preprocessing could add a significant delay. Rezaei Jegarluei *et al.* [28] presented a fault-location algorithm that operates on both synchronized and unsynchronized PMU data for the IEEE 39-bus system, focusing on wide-area backup protection while leaving fault-type classification unaddressed.

Existing literature methods that integrate feature extraction are generally trained using fault scenarios limited to a single transmission line. To provide system operators with enhanced situational awareness and enable faster fault resolution, a true wide-area approach is necessary. Accordingly, this study adopts a wide-area viewpoint and develops auto-encoder neural network models specifically for fault detection and classification. Auto-encoders are particularly advantageous because they learn features automatically during training, eliminating the need for a separate, labor-intensive feature engineering step. They can also capture complex, nonlinear patterns in the data that are difficult to define manually. The proposed models were trained and validated using WAMS data generated from a 39-bus test power system. The observed training/testing times and accuracy levels confirm the computational efficiency of this methodology.

Contributions

The key contributions of this research are:

- Utilization of WAMS data from a wide-area standpoint to support fault diagnosis across the entire power system. This paper examines data-driven Auto-Encoder neural networks that automatically discover relevant features straight from raw inputs during training. These networks construct effective FD&C models for transmission lines using WAMS measurements, thereby eliminating the traditional, time-consuming manual feature-extraction process. Performance is evaluated based on training and testing durations and classification accuracy.
- Comparative analysis between a Stacked Auto-Encoder (SAE) model and a multi-layer Extreme Learning Machine (ML-ELM) model. These auto-encoder-based approaches are benchmarked against current state-of-the-art classifiers to demonstrate their superior effectiveness.
- Proposal of a fast and computationally efficient FD&C model for transmission lines in modern smart grids that fully exploits WAMS data.

Auto-Encoder Networks

Stacked auto-encoder neural network (SAE)

An auto-encoder is an unsupervised neural network architecture that learns compact representations of input data. It comprises an encoder that extracts latent features and a decoder that reconstructs the original input from them [29, 30]. A Stacked Auto-Encoder (SAE) is created by connecting multiple auto-encoders in series to form a deeper network.

Initially, the encoder maps the input vector X_i to a hidden representation Y_i :

$$Y_i = g(X_i) = A_f(W_{f1}^T + b_{f1}) \quad (1)$$

Where A_f denotes the activation function, $f1$ represents the weight matrix, and $bf1$ is the bias vector of the encoder layer. The decoder then reconstructs the output Y_o from the encoded features:

$$Y_o = g_o(Y_i) = A_o(W_{f2}^T + b_{f2}) \quad (2)$$

With A_o as the decoder activation function, W_{f2} as the decoder weights, and $bf2$ as the decoder biases. These parameters are optimized through back-propagation by minimizing the loss function defined in [Eq. 3](#):

$$\Phi(X_i, Y_o) = \Phi_r(X_i, Y_o) + 0.5\sigma(\|W_{f1}\|_2^2 + \|W_{f2}\|_2^2) \quad (3)$$

Where $\Phi_r(X_i, Y_o)$ is the reconstruction error, and σ is the weight decay regularization term. In an SAE, several such autoencoders are stacked, so the learned representation from one layer serves as input to the next, building a hierarchical deep network ([Figure 1](#)).

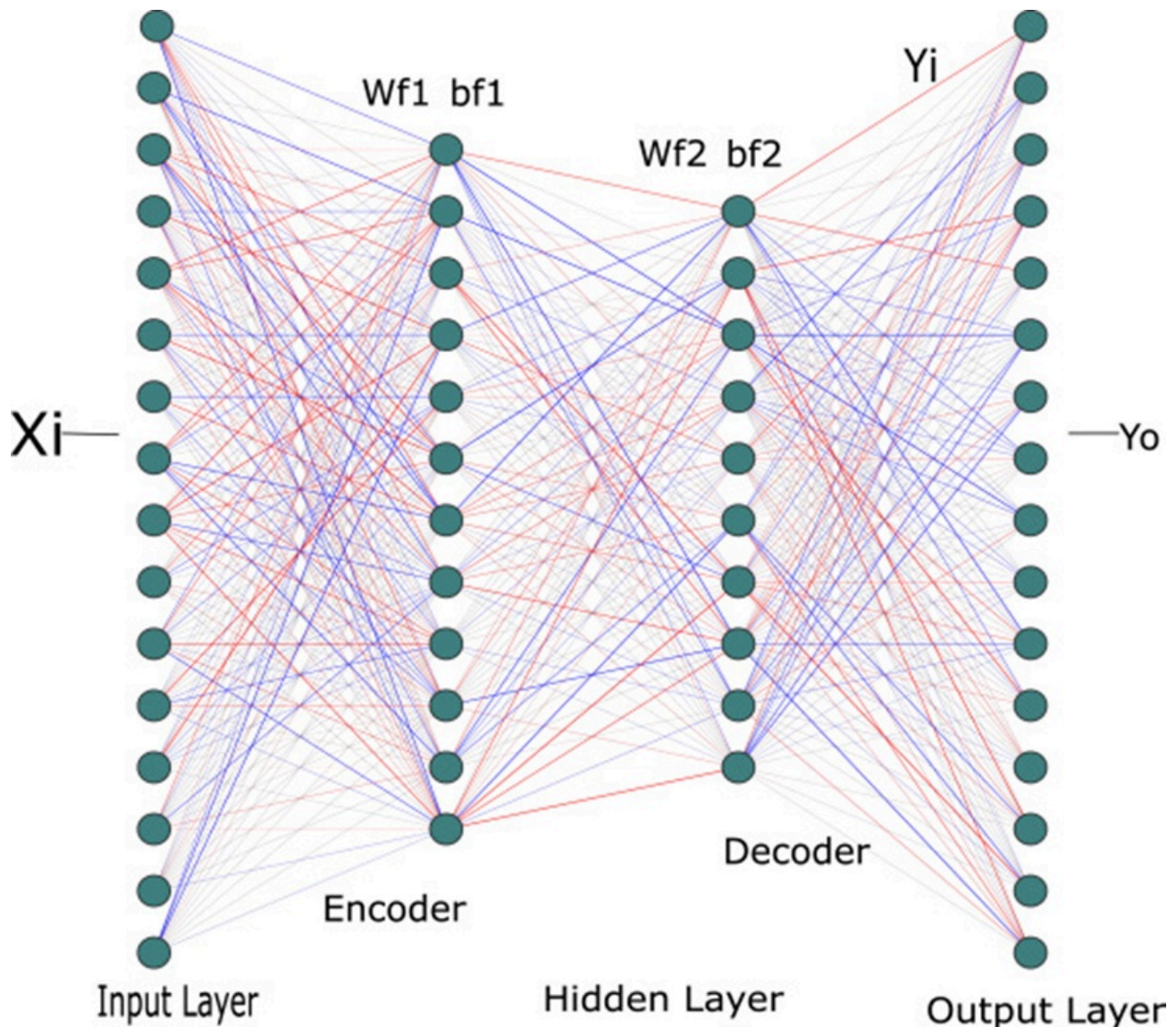


Figure 1. General layout of an auto-encoder neural network.

Multi-layer Extreme Learning Machine (ML-ELM)

Multi-layer Extreme Learning Machines (ML-ELM) are constructed by stacking several Extreme Learning Machine-based Auto-Encoders (ELM-AE), which act as the core building units of the overall architecture. The basic structure of one ELM-AE unit is presented in [Figure 2](#) [31, 32].

Consider an ELM-AE with n input variables and S neurons in the hidden layer. When the number of inputs n equals S , the hidden layer represents the same dimensionality. If n is smaller than S , the network creates a compressed feature set. Conversely, when n is larger than S , it results in a sparse representation.

In this ELM-AE setup, the hidden layer generates orthogonal features derived from the input data. These features are then passed through a nonlinear activation function, most commonly the sigmoid [33]. The mathematical expression for the hidden layer output h_i is provided in [Eq. 4](#):

$$\begin{aligned} h_i &= g(a_i x + b_i) \\ a_i^T a_i &= I, b_i^T b_i = I \end{aligned} \quad (4)$$

Where $a_i = [a_{i1}, a_{i2}, \dots, a_{iS}]$ are the randomly assigned orthogonal weights and $b_i = [b_{i1}, b_{i2}, \dots, b_{iS}]$ are the corresponding orthogonal biases linking the input and hidden layers. The output weights β_i map the hidden features back to the original inputs.

$$\beta_1 = \left(\frac{I}{C} + H_i^T H_i \right) H_i^T X_i \quad (5)$$

In the above, H_i denotes the hidden-layer activation matrix, and X_i denotes the input vectors. The final reconstruction of the input for both compressed and sparse cases is described by [Eq. 5](#). When the input and hidden layer sizes are equal, the output is given by [Eq. 6](#).

$$\begin{aligned} \beta_1 &= H_i^{-1} X_i \\ \beta_i^T \beta_i &= I \end{aligned} \quad (6)$$

The learned features essentially reflect the singular value decomposition properties of the input data.

$$H_i^j = g(\beta_1^k)^T H_i^{j-1} \quad (7)$$

Here, j refers to the j -th hidden layer, and H_{ij} denotes its output. The weights connecting the final hidden layer to the output nodes are computed directly via regularized least-squares optimization [\[34\]](#).

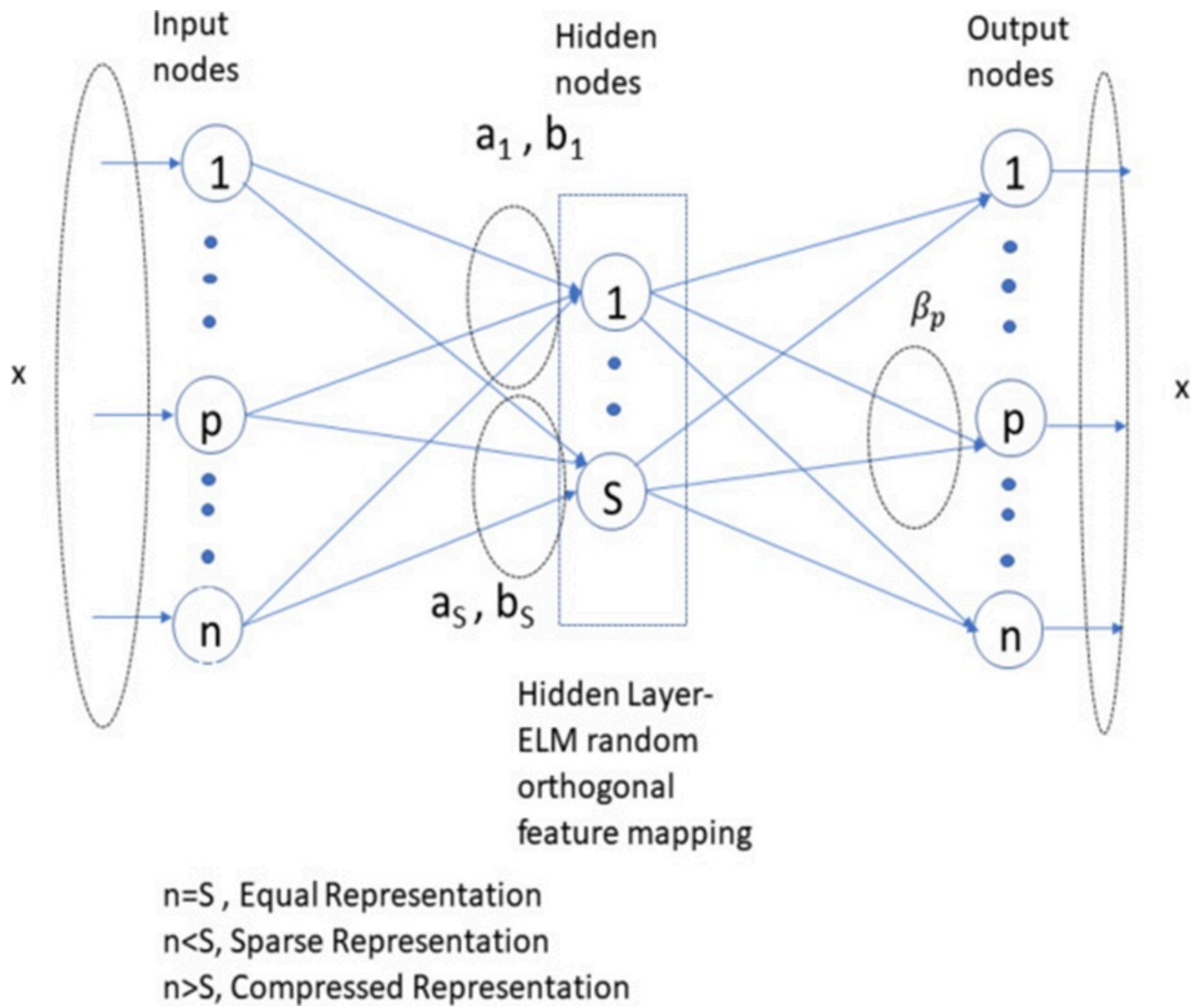


Figure 2. Structure of an ELM auto-encoder neural network.

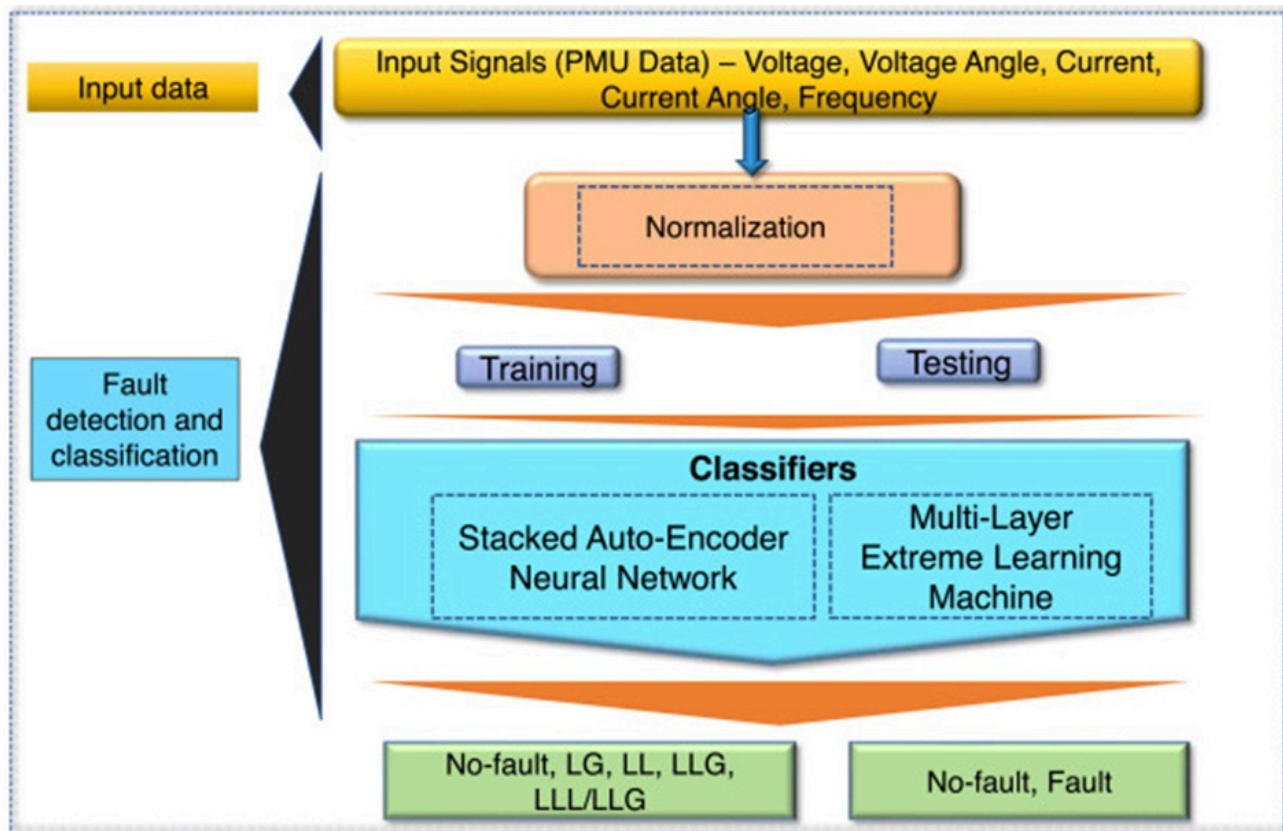


Figure 3. Methodology for transmission line fault detection and classification using auto-encoder network models.

Data Acquisition, Fault Detection, and Classification

The complete procedure for performing fault detection and classification (FD&C) based on Stacked Auto-Encoder (SAE) and Multi-layer Extreme Learning Machine (ML-ELM) approaches is outlined in [Figure 3](#).

Generating the dataset via simulation

Researchers modeled the IEEE 39-bus power system using the PowerWorld Simulator software [35]. This benchmark network comprises 10 synchronous generators, 39 buses, 34 transmission lines, and 12 power transformers. It operates at a base voltage of 345 kV and a frequency of 60 Hz. The one-line diagram of this test system, including all model parameters, is displayed in [Figure 4](#) [35].

Fault conditions were created and analyzed using the Transient Stability analysis tool within the simulator. Studies covered both the normal (no-fault) operating state and several common fault types: single-line-to-ground (LG), line-to-line (LL), double-line-to-ground (LLG), and three-phase (LLL/LLG).

During each simulation run, measurements of voltage magnitude, voltage phase angle, and line current were collected. Fault resistance was varied between 1 Ω and 50 Ω in 5 Ω increments, while fault positions along the lines were adjusted from 10% to 90% of the total line length in 10% steps. A full list of simulation parameters for the different fault scenarios is provided in [Table 1](#).

As specific examples: an LG fault was applied on the transmission line linking bus 2 to bus 25, at 70% distance from bus 2. The fault occurred at $t = 1$ s, was cleared at $t = 1.03$ s, and had a resistance of 10 Ω . A line-to-line fault was introduced between buses 3 and 18 at 80 percent of the line length, starting at 1 s and cleared at 1.03 s with 50 Ω resistance.

Similarly, an LLG fault was placed on the line between buses 21 and 22, 40% from bus 21, initiated at 1 s and removed after 1.03 s using 50 Ω resistance. A symmetrical three-phase fault was also simulated on the same line section (buses 21–22) at the 40% location from bus 21, starting at 1 s and cleared at 1.03 s with 50 Ω resistance. Each simulation case was executed for 5 seconds.

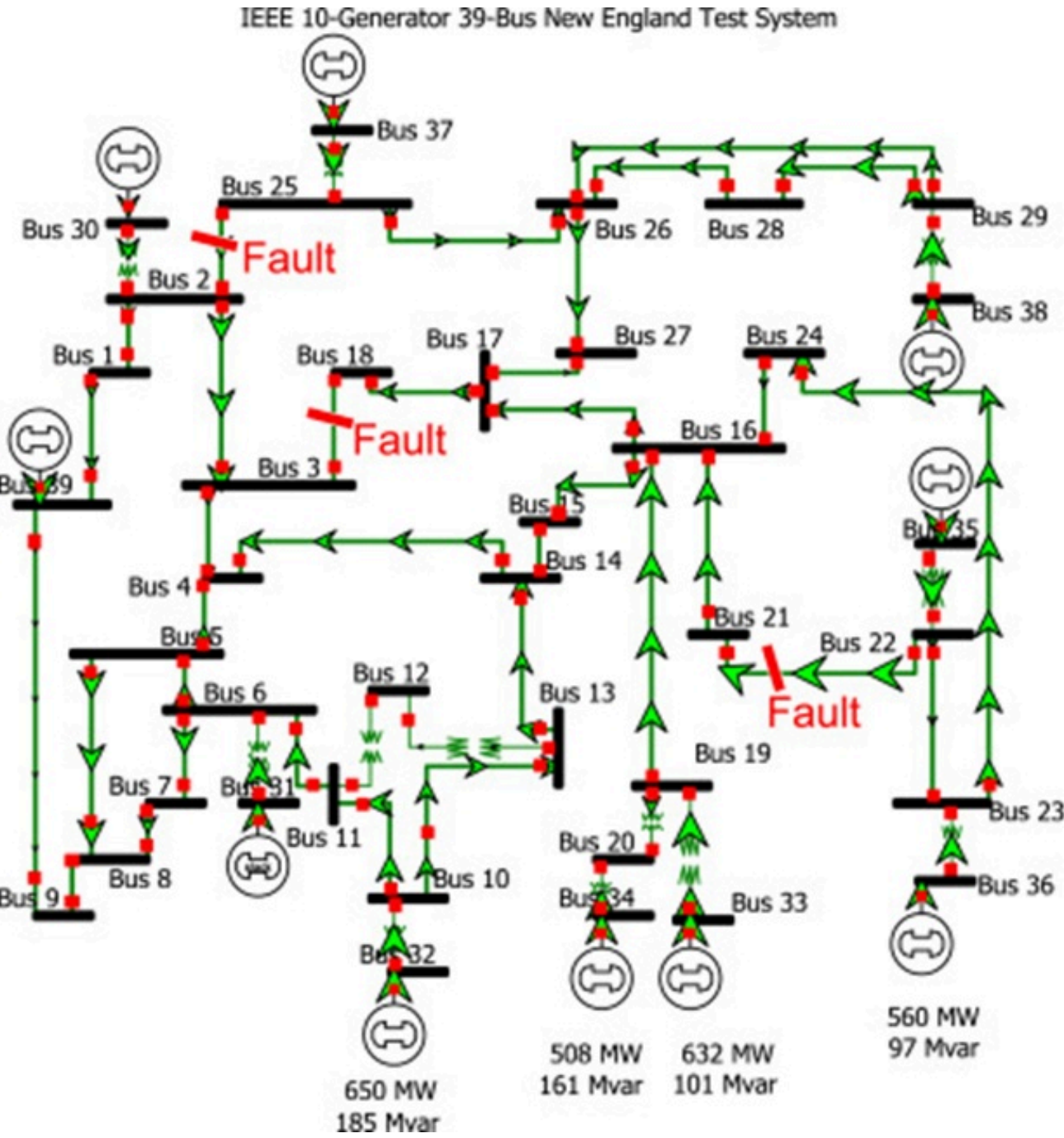


Figure 4. Single-line representation of the IEEE 39-bus benchmark system.

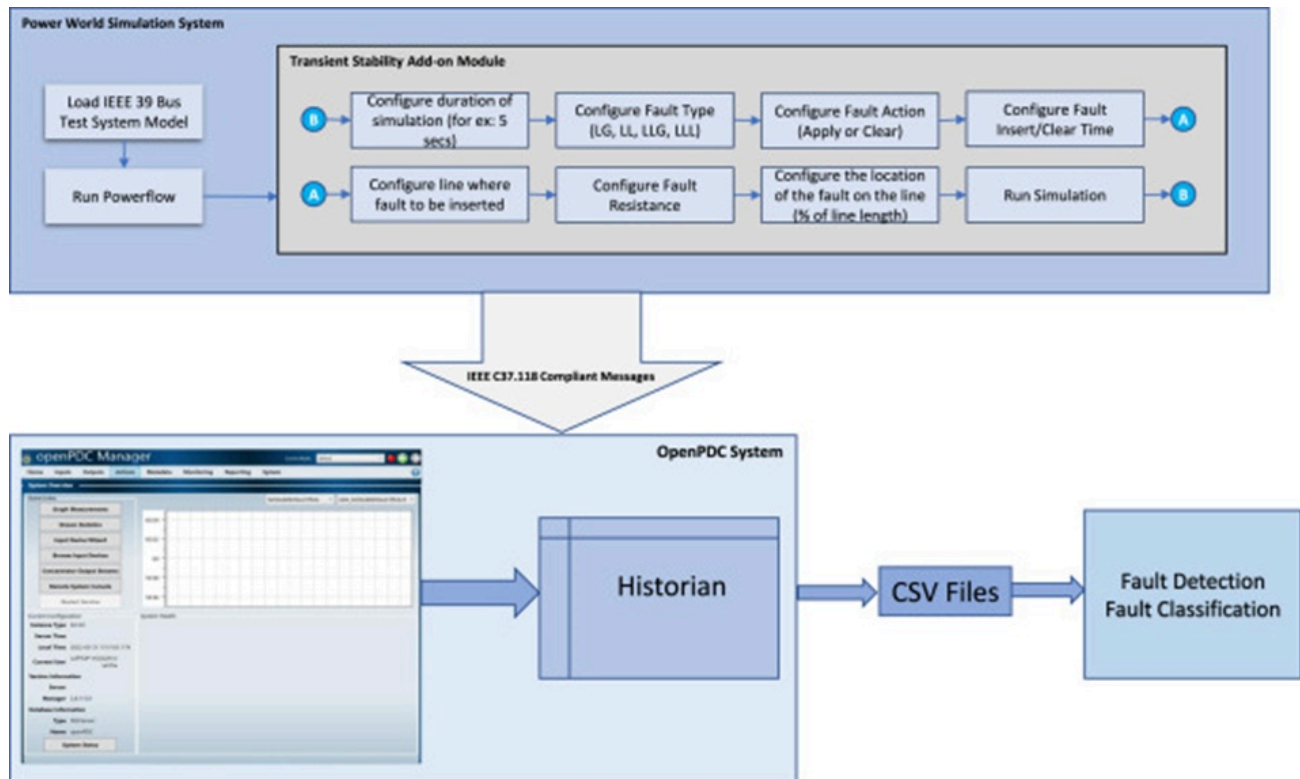


Figure 5. Overview of the simulation environment used for creating the dataset.

In each simulation, the PowerWorld Simulator dynamic simulation engine produced data streams in the form of IEEE C37.118.2011-compliant synchrophasor messages. These messages were captured and recorded using OpenPDC software. The collected information was then archived in OpenPDC's Historian database. Later, the stored data was exported from the Historian as CSV files. The exported files underwent thorough validation to detect any inconsistencies, followed by cleaning, proper labeling, and final import into MATLAB for further processing. The complete Wide-Area Measurement System (WAMS) data acquisition workflow is depicted in [Figure 5](#).

Table 1. Parameters for fault dataset generation.

Parameters	Values
Fault type	No-fault, LG, LL, LLG, LLL/LLL
Fault Resistance Ω	1, 5, 10, 15, 20, 25, 35, 40, 50
Fault location (% of the length of the line)	10, 20, 30, 40, 50, 60, 70, 80, 90

The final compiled dataset consisted of 24,654 samples of voltage, voltage angle, current, current angle, and frequency measurements collected across all 39 buses. This dataset was randomly split into two parts at a 70:30 ratio, yielding 15,317 samples for training and 7,134 for testing. The breakdown of samples among the various categories for both fault detection and fault classification tasks is presented in [Tables 2 and 3](#), respectively.

Table 2. Composition of the training and testing dataset for detecting faults.

Dataset	No-fault (1)	Fault (2)
Training	3868	11449
Testing	2229	4905

Table 3. Distribution of data for different fault types in the dataset for classifying the faults.

Dataset	NF (1)	LG (2)	LL (3)	LLG (4)	LLL/LLLG (5)
Training	3868	2694	2694	2694	3367
Testing	2229	1154	1154	1154	1443

Fault Detection and Classification

Fault detection and classification with Stacked Auto-Encoder (SAE) model

A Stacked Auto-Encoder (SAE) architecture was specifically built to identify and categorize faults occurring on transmission lines. The complete layout of this SAE-based network is displayed in [Figure 6](#). The main purpose of the auto-encoder components is to significantly reduce the dimensionality of the raw input measurements by extracting compact and meaningful representations.

The model consists of three auto-encoder layers stacked together. These layers are trained in two separate phases: an initial unsupervised pre-training phase, followed by a supervised fine-tuning phase.

During the unsupervised pre-training, the first auto-encoder compresses the original high-dimensional input data. The resulting feature set from this layer is then passed to the second auto-encoder for further compression. In the same way, the features produced by the second layer serve as input to the third auto-encoder. The internal configurations of these three auto-encoder stages are illustrated in [Figure 7](#).

Once the unsupervised pre-training is complete, a Softmax layer is attached on top and trained using supervised learning. This final layer uses the compressed features along with their corresponding labels to perform classification.

In the case of fault detection, the task is framed as a simple binary classification problem distinguishing between normal operating conditions (No-fault) and the presence of any fault (Fault). For fault classification, the model handles a five-class problem: No-fault (class 1), LG (class 2), LL (class 3), LLG (class 4), and LLL/LLLG (class 5). The detailed structure of the Softmax output layer for both detection and classification is presented in [Figure 8](#).

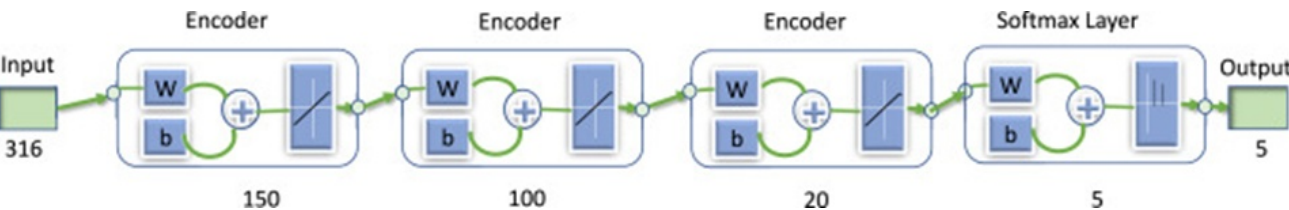


Figure 6. Stacked Auto-Encoder neural network for fault classification.

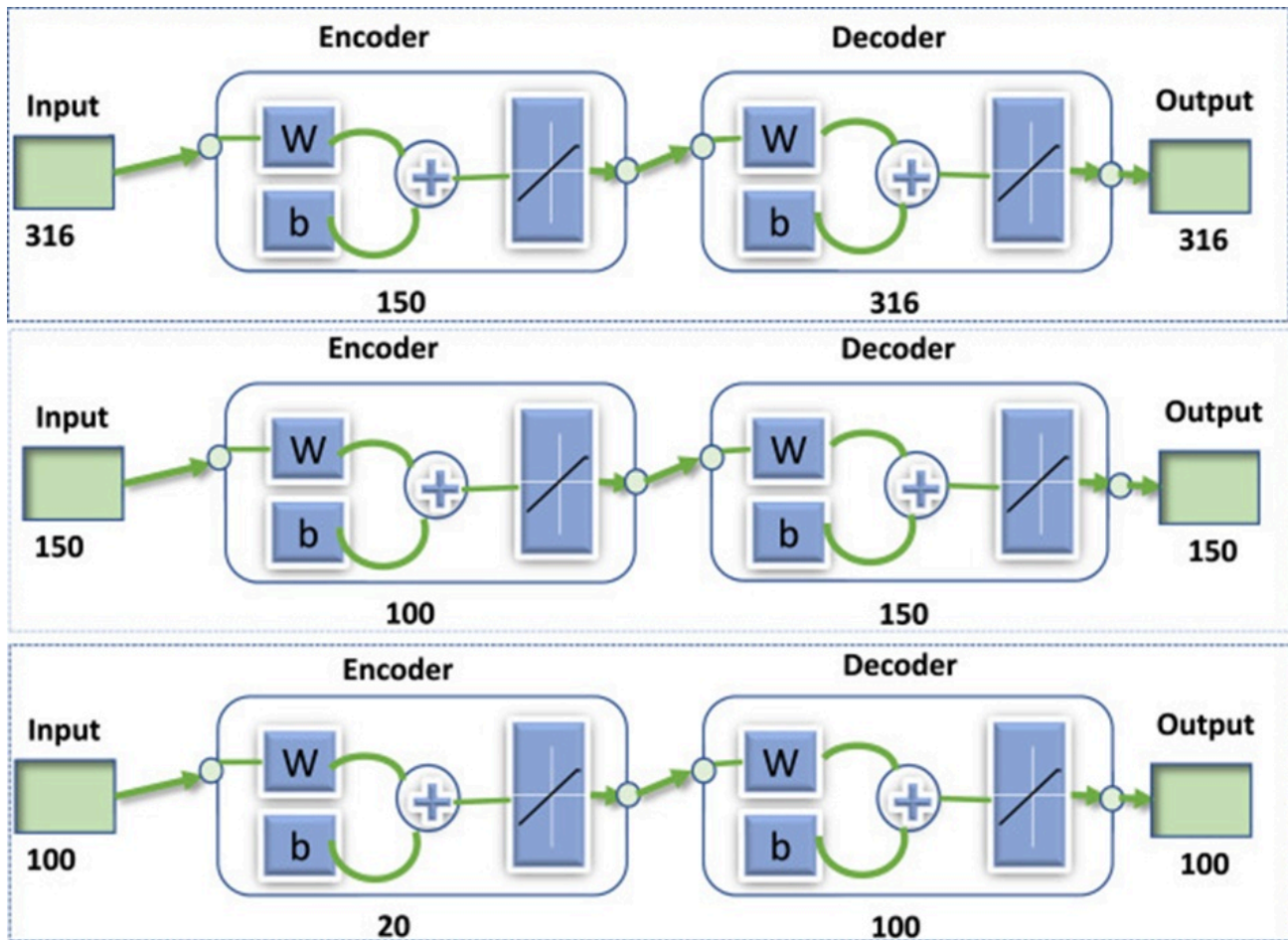


Figure 7. Auto-encoders trained for fault classification.

In designing the Artificial Neural Network, several critical factors required careful adjustment, including the number of hidden layers, the number of neurons per layer, the type of activation function, and the strength of L2 regularization. A very small number of neurons tended to cause underfitting, while excessively large layers often led to overfitting of the training data.

An iterative trial-and-error approach was used to identify the most effective settings. Multiple configurations were tested, and each was evaluated based on its performance on a held-out validation dataset. The set of parameters yielding the strongest validation results was chosen as the final.

Three activation functions — ReLU, tanh, and sigmoid — were tested. After comparing their outcomes, sigmoid consistently delivered superior performance and was therefore adopted across all hidden layers.

The tuning process was carried out in two phases. Initially, the neuron counts in each layer were adjusted while holding the L2 regularization value constant. Once the best neuron configuration was locked in, the L2 regularization strength was incrementally adjusted for each layer.

Extensive experiments revealed that the optimal SAE configuration for processing WAMS data consisted of three hidden layers with 150, 100, and 20 neurons, respectively. The L2 regularization coefficients were fixed at 0.001 for the first layer, 0.002 for the second, and 0.003 for the third.

This resulted in a stacked architecture of 150-100-20 neurons. For the fault classification task, the output layer was set to 5 neurons to match the 5 fault categories. For fault detection, the output layer was reduced to 2 neurons to handle the binary decision. The complete SAE network developed for fault detection is depicted in Figure 9.

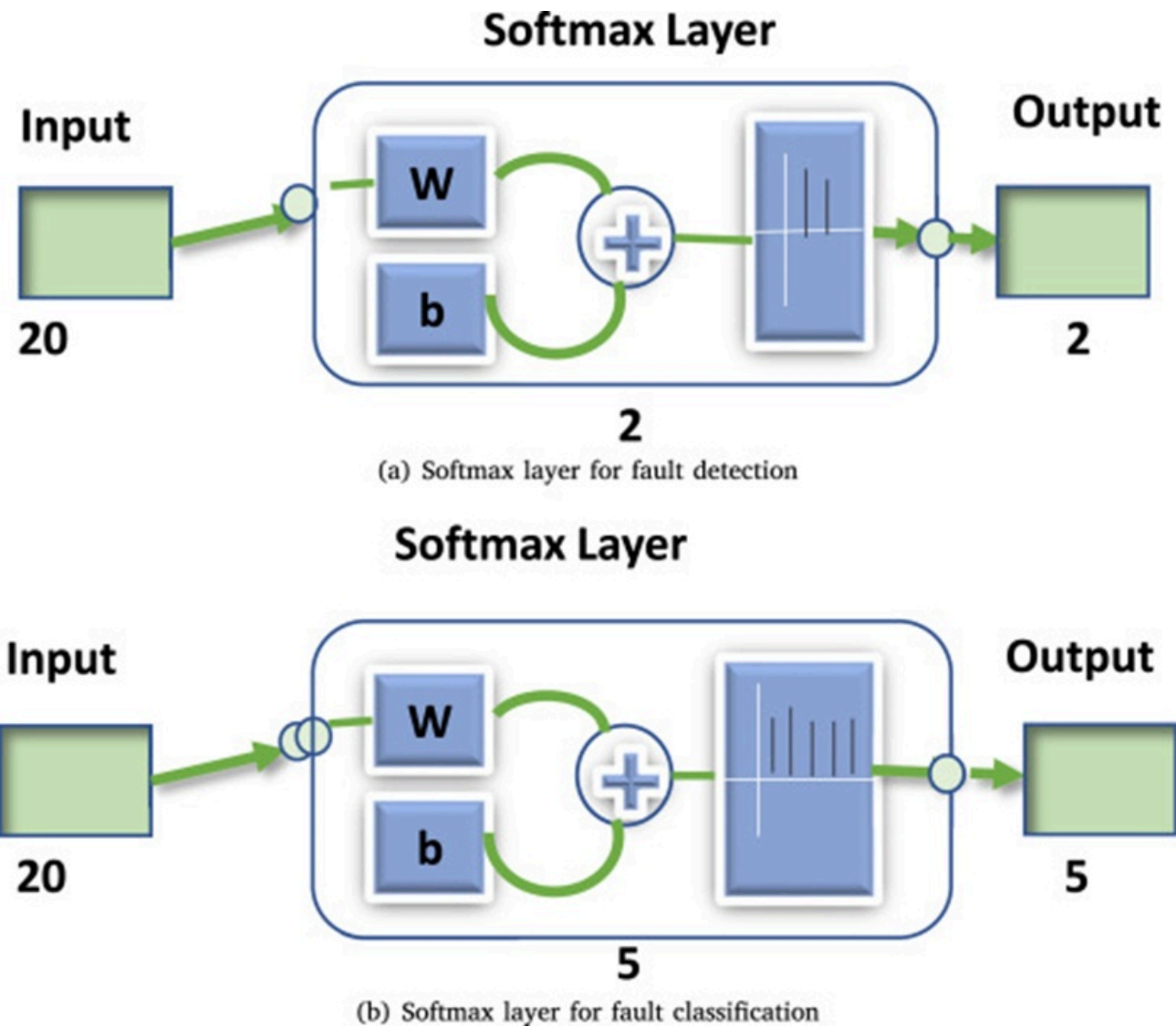


Figure 8. Softmax layer for fault detection and classification.

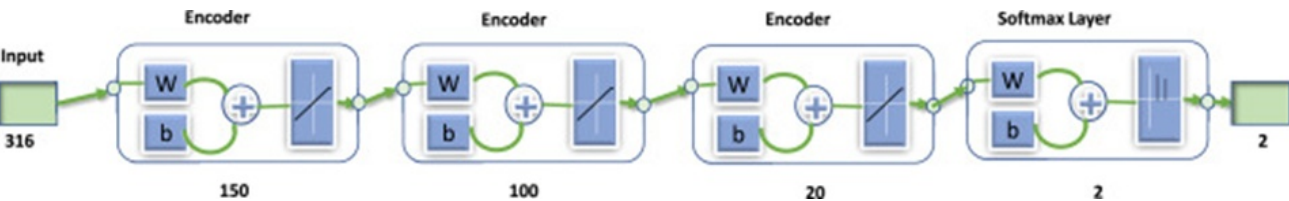


Figure 9. Stacked auto-encoder neural network for fault detection.

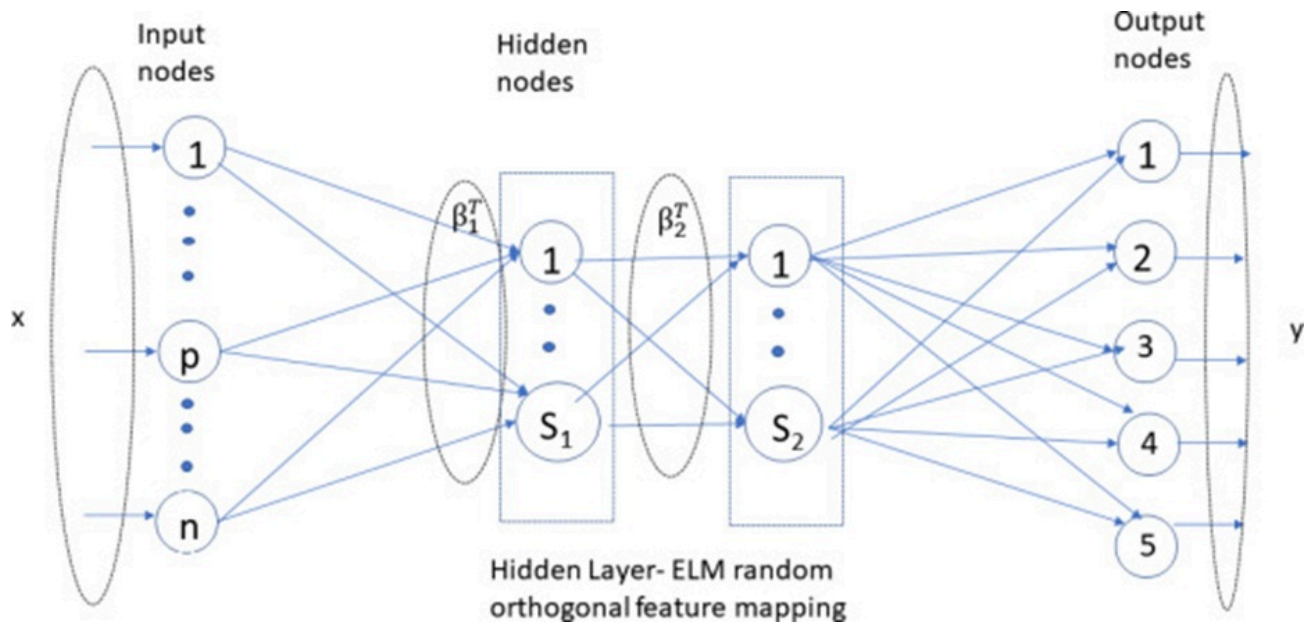


Figure 10. ML-ELM for fault classification.

Fault detection and classification using ML-ELM based on ELM-AE

The overall design of the Multi-layer Extreme Learning Machine (ML-ELM) developed for fault classification is illustrated in **Figure 10**. This network uses two hidden layers: the first with 150 nodes and the second with 300 nodes.

For ELM-based networks, the main adjustable settings include the number of hidden layers and the number of neurons per layer, the activation function, and the ridge parameter (C). Having too few neurons usually leads to poor learning (underfitting), while having too many can cause the model to memorize noise rather than generalize (overfitting). The optimal number of neurons was determined using a stepwise experimental search method [36]. This involved testing many different hyperparameter combinations and picking the ones that performed best on a validation dataset.

Input weights (w) and biases (b) were randomly initialized while satisfying the orthogonality constraints ($w^T w = I$ and $b^T b = I$). The sigmoid function served as the activation function for both hidden layers. Each hidden layer produces an orthogonal feature representation by applying the sigmoid function to the transformed inputs.

In this multi-layer setup, for a given dataset $\{(x_i, y_i)\}$ from $i=1$ to N , the output of each successive hidden layer l is calculated using the recursive formulation involving the sigmoid function $g(\cdot)$ as described by Yang and Wu [34]. These orthogonal features are further refined using singular value decomposition (SVD) of the matrix $C = \text{Input}^T H$, yielding the final orthogonal matrix $Q = UV^T$ [37].

The best hyperparameter values were identified in two phases: first, by adjusting the number of neurons while holding C constant; then, by tuning C across layers. The performance achieved with different neuron counts and C values is summarized in **Table 4**.

After thorough testing, the optimal ML-ELM configuration for WAMS data was 150 neurons in the first hidden layer and 300 neurons in the second. The ridge parameter C was set to 0.02 for the layer leading into 150 nodes, and 2×10^8 for both the 150-to-300 layer and the final 300-to-5 layer.

For fault detection, the same network architecture is used except for the output layer. Since fault detection is a binary problem, the output layer has only 2 nodes, corresponding to the No-fault (normal operation, class 1) and Fault (faulty condition, class 2) classes.

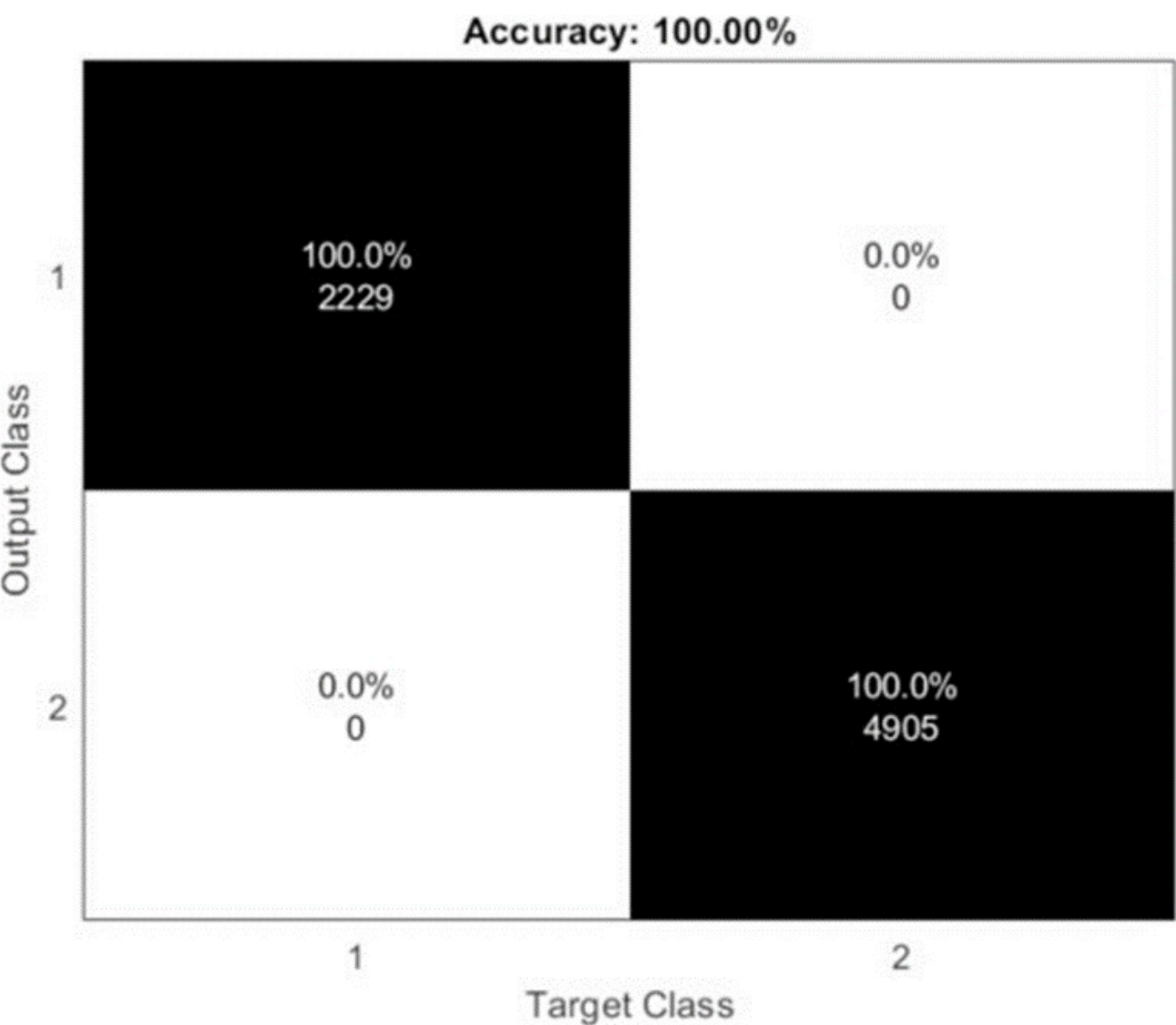


Figure 11. Confusion matrix illustrating the fault detection performance of the ML-ELM model.

Table 4. ML-ELM network parameters determination.

Testing	Number of neurons	Number of neurons	Parameter
Accuracy (%)	(First layer)	(Second layer)	C
99.36	50	100	[2e−2, 2e4,2e8]
99.72	75	150	[2e−2, 2e4,2e8]
99.78	100	200	[2e−2, 2e4,2e8]
99.85	150	250	[2e−2, 2e4,2e8]
99.88	150	300	[2e−2, 2e4,2e8]
99.88	200	300	[2e−2, 2e4,2e8]
99.70	150	300	[2e2, 2e2,2e2]

99.53	150	300	[2e4, 2e2,2e2]
99.62	150	300	[2e2, 2e4,2e2]
99.71	150	300	[2e2, 2e4,2e4]
99.59	150	300	[2e2, 2e4,2e6]
99.87	150	300	[2e−2, 2e4,2e8]
99.89	150	300	[2e−2, 2e6,2e8]
99.92	150	300	[2e−2, 2e8,2e8]

Results and Discussion

The three individual auto-encoder networks were first trained separately and then combined into a single stacked model. This final Stacked Auto-Encoder (SAE) system successfully detected all faults with 100% accuracy and achieved 97.99% classification accuracy across the different fault types.

In comparison, the Multi-layer Extreme Learning Machine (ML-ELM) model also achieved 100% fault detection accuracy and 99.92% classification accuracy.

A detailed summary of the classification performance metrics for both the SAE and ML-ELM models is provided in [Table 5](#). Similarly, the detection performance results for both approaches are presented in [Table 6](#).

Table 5. Fault classification using auto-encoder networks.

Classifier	Testing accuracy (%)	Training accuracy (%)	Testing time (s)	Training time (s)
SAE	97.99	97.97	0.091	809.15
ML-ELM	99.92	99.90	0.029	0.0381

Table 6. Fault detection using auto-encoder networks.

Classifier	Testing accuracy (%)	Training accuracy (%)	Testing time(s)	Training time(s)
SAE	100	100	0.0427	806.54
ML-ELM	100	100	0.0495	0.4626

The experiments were carried out on a system equipped with an Intel Core i7-9750H processor operating at 2.60 GHz, 32 GB of RAM, and MATLAB 2023a.

For fault detection, the SAE neural network required 806.5 seconds for training and 43 ms for testing. In contrast, the ML-ELM model was much faster to train, needing only 463 ms, while its testing time was 50 ms.

Regarding fault classification, training the SAE network took 809 seconds, and the model completed testing in 91 ms. This testing duration remains comfortably below the recommended threshold of 100 ms for power system fault diagnosis techniques [38]. The ML-ELM approach proved highly computationally efficient, with a training time of just 38 ms and a testing time of 29 ms. After fine-tuning the SAE network using back-propagation, it successfully achieved 100%

classification accuracy. The confusion matrices illustrating the performance of the ML-ELM model for both fault detection and classification are presented in **Figures 11 and 12**, respectively.

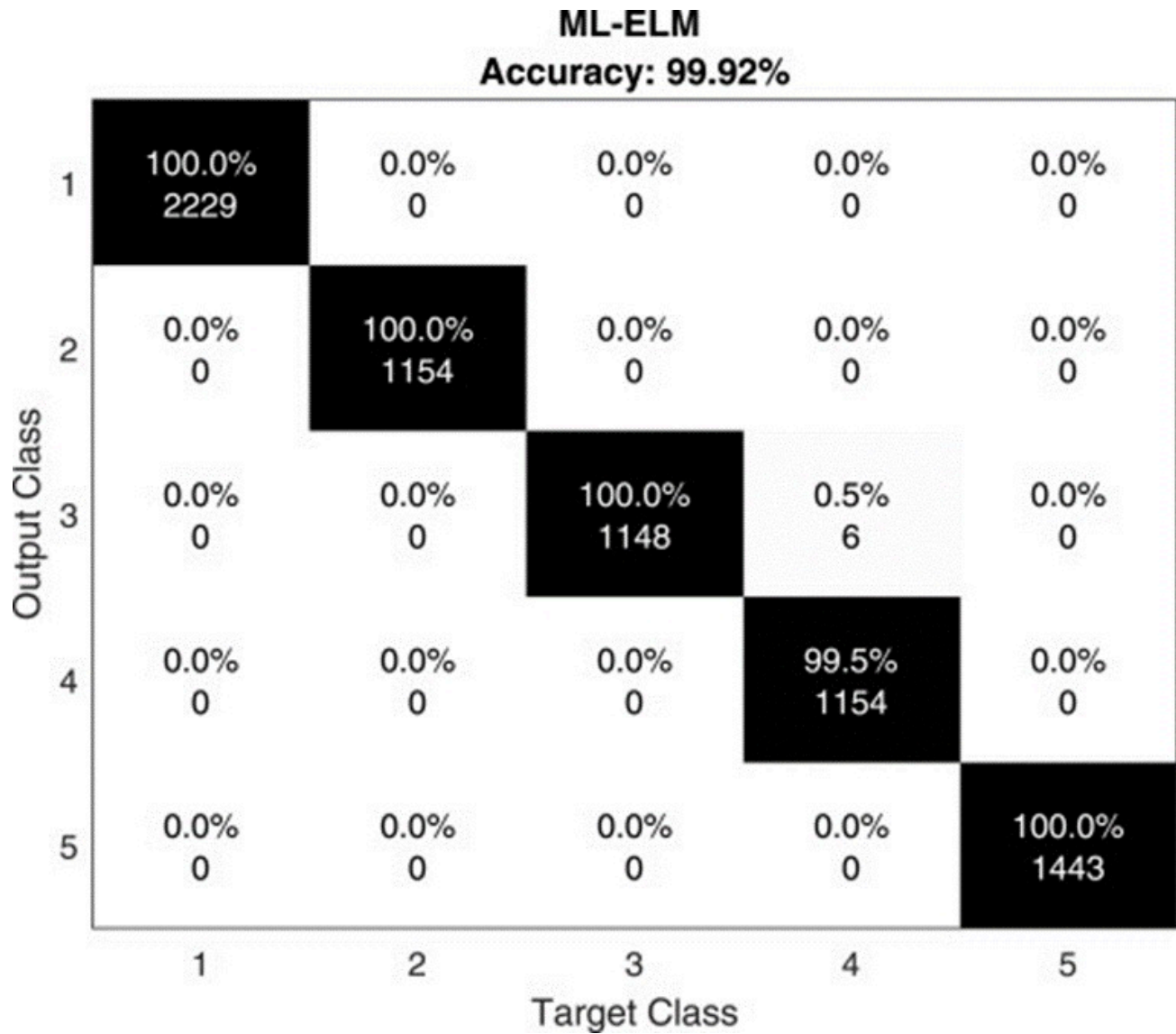


Figure 12. Confusion matrix of the ML-ELM model for fault classification.

Performance under noisy conditions

The studies in Brown *et al.* [39] and Bhandari *et al.* [40] investigated the types and characteristics of noise present in practical PMU measurements. According to Brown *et al.* [39], field PMU data typically exhibits white Gaussian noise with a zero mean and a signal-to-noise ratio (SNR) of 45 dB or higher.

To evaluate the models' robustness under noisy conditions, white Gaussian noise was added to the input data, with SNR levels ranging from 10 dB to 40 dB. The resulting classifier performance on these noisy datasets is illustrated in **Figure 13**.

When the input data had an SNR of 10 dB, the SAE network achieved a fault classification accuracy of 95.54%, whereas the ML-ELM model reached 98.72%. This indicates that the ML-ELM model maintains high performance even in the presence of substantial noise. On noise-free datasets, the ML-ELM model achieved 99.92% accuracy, while the SAE

network achieved 97.99%. The reduction in accuracy for the ML-ELM model when moving from noise-free data to the noisiest dataset (SNR 10 dB) was only 1.2%. In contrast, the SAE network experienced a slightly larger drop of approximately 2.45%.

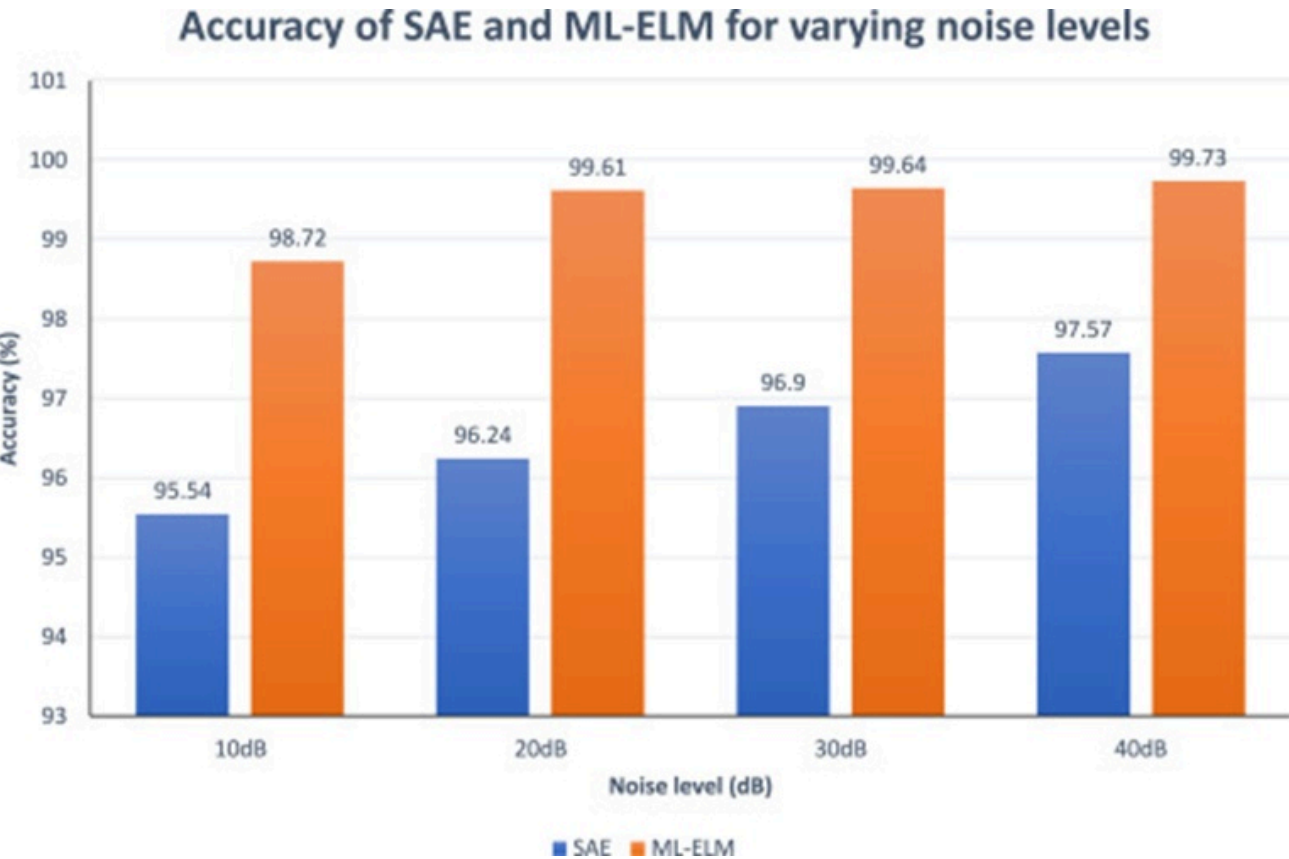


Figure 13. Accuracy of SAE and ML-ELM models for fault classification under noisy conditions.

Evaluation against contemporary supervised learning approaches

The classification of transmission line faults was further assessed by comparing SAE and ML-ELM models with several modern supervised learning techniques, including ANN, DT, SVM, PSO-WELM, Ensemble Tree, and KNN, using feature sets derived from wavelet transforms [41]. This study specifically examined the relative performance of SVM, KNN, and PSO-WELM models compared with SAE and ML-ELM architectures. The input features for all models were extracted using wavelet-based techniques as detailed in Jnaneswar *et al.* [6], Dasgupta *et al.* [7], Patel *et al.* [8], and Costa *et al.* [9]: Jnaneswar *et al.* [6] used energy from wavelet approximation coefficients of current signals; Dasgupta *et al.* [7] extracted Wavelet Entropy from voltage signals; Patel *et al.* [8] applied Wavelet Packet decomposition on three-phase fault currents; Costa *et al.* [9] employed Maximum Overlap Discrete Wavelet Transform (MODWT) with energy coefficients as features. Additionally, Harish *et al.* [26] created a composite feature set combining statistical, energy, and entropy measures from wavelet coefficients. The resulting classifier performances are summarized in Table 7.

Reported testing times include both the feature extraction process and model evaluation. For instance, SVM achieved 99.92% accuracy on Wavelet Packet Entropy features with a testing duration of 5.67 s, while PSO-WELM reached 99.85% accuracy on the ensemble feature set in 0.78 s. In comparison, SAE and ML-ELM models operating on normalized input data achieved accuracies of 97.99% and 99.92%, respectively, with minimal testing times of 0.091 s and 0.0283 s, highlighting their efficiency for rapid fault classification.

Table 7. Testing accuracy and testing time of state-of-the-art classifiers for fault classification using WAMS data from the IEEE-39 bus test system.

Reference	Input	Classifier	Testing time (s)	Testing accuracy (%)
[6]	Energy values of Wavelet Coefficients	KNN	0.122	98.95
[7]	Entropy values of Wavelet Coefficients	SVM	0.588	66.13
[8]	Wavelet Packet Entropy	SVM	5.699	99.92
[9]	MODWT coefficients Energy values	SVM	5.562	62.41
[26]	Ensemble features	PSO-WELM	0.78	99.85
Proposed method	Normalized raw data	SAE	0.091	97.99
Proposed method	Normalized raw data	ML-ELM	0.029	99.92

Table 8. Performance measures of deep learning methods for fault classification using WAMS data from the IEEE-39 bus test system.

Reference	Classifier	Testing accuracy (%)	Training accuracy (%)	Testing time (s)	Training time (s)
[27]	CNN	98.95	99.12	0.7700	527.74
[20]	LSTM	96.78	98.09	1.879	798.88
Proposed method	SAE	97.99	97.97	0.091	809.15
Proposed method	ML-ELM	99.92	99.90	0.029	0.0381

Table 9. Accuracies of cross-validated models with five folds for ANOVA and t-test.

Method	Testing accuracy (%)
SAE	99.22
ML-ELM	99.93
SAE	98.01
ML-ELM	99.93
SAE	98.24
ML-ELM	99.83
SAE	95.49
ML-ELM	99.97
SAE	98.99
ML-ELM	99.90

Comparison with alternative deep learning approaches

The effectiveness of the SAE and ML-ELM networks was benchmarked against deep learning models such as LSTM [20] and CNN [27], trained using WAMS data. The LSTM model consisted of a Sequence-Input layer with a single channel corresponding to the input signal, followed by three LSTM layers containing 52, 40, and 25 hidden neurons, each separated by Dropout layers. The final layers included a fully connected layer matching the number of fault classes, a Softmax layer, and a classification layer. The CNN architecture started with a Sequence-Input layer (minimum filter length 25), followed by a 1D Convolutional layer with 25 filters, Batch Normalization, ReLU activation, Dropout, Global Max Pooling, a Fully Connected layer with outputs equal to the number of classes, Softmax, and a Classification layer. Table 8 summarizes the performance metrics for these deep learning models, alongside those of the auto-encoder models.

The ML-ELM classifier required only 38 ms for training and 29 ms for testing, achieving the fastest evaluation time among the considered models and confirming its high suitability for transmission line fault detection and classification (FD&C).

Statistical evaluation of SAE versus ML-ELM performance

The training data were divided into five folds for cross-validation, and the test accuracies for each fold were collected for comparison (Table 9). ANOVA and two-tailed t-tests were conducted to assess model performance differences. The ANOVA test produced a p-value of 0.02, indicating a statistically significant difference between the two groups. A left-tailed, two-sample t-test returned $h = 1$, leading to rejection of the null hypothesis at the 5% significance level, confirming that the mean accuracy of the ML-ELM model is significantly higher than that of the SAE model.

Conclusion

In modern power grids with distributed generation, fault diagnosis methods that provide wide-area situational awareness are essential. Computational efficiency is critical when implementing intelligent fault-detection models. The ML-ELM model achieved 100% fault detection and 99.92% classification accuracy within 29 ms, whereas the SAE network detected faults at 100% but classified them at 97.99% accuracy within 91 ms. Training times for SAE models were considerably higher than for ML-ELM models, and while fine-tuning slightly improved SAE accuracy, ML-ELM models remain far superior in speed. Comparative evaluations with traditional classifiers, CNN, and LSTM models further validate the robustness and efficiency of the auto-encoder-based approaches.

Limitations and future work

The proposed methods rely on feature learning from input data, which requires training on datasets that cover all operational scenarios for optimal performance. Due to the limited availability of real-world data, the models were trained on simulated data from a 39-bus test system. Future improvements could include hybrid training with both synthetic and real data, extending the models to larger power systems for validation. The approaches can also be adapted to detect evolving faults, distinguish primary from secondary faults, and locate faults to accelerate system recovery. Additionally, these methods could be applied to distribution systems using micro-PMU data to enhance fault diagnosis capabilities.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 04 Oct 2024

Revised: 10 Nov 2024

Accepted: 19 Dec 2024

Published online: 20 March 2025

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at <https://creativecommons.org/licenses/by-nc-sa/4.0/>.

References

- Li F, Qiao W, Sun H, Wan H, Wang J, Xia Y, et al. Smart transmission grid: Vision and framework. *IEEE Trans Smart Grid*. 2010;1(2):168-77.
- Bhuiyan SM, Khan JF, Murphy GV. Big data analysis of the electric power PMU data from smart grid. In: *Conf Proc IEEE SOUTHEASTCON*. 2017. p. 1-5.
- Biswal M, Brahma SM, Cao H. Supervisory protection and automated event diagnosis using PMU data. *IEEE Trans Power Del*. 2016;31(4):1855-63.
- Phadke AG, Wall P, Ding L, Terzija V. Improving the performance of power system protection using wide area monitoring systems. *J Mod Power Syst Clean Energy*. 2016;4(3):319-31.
- Labrador Rivas AE, Abrão T. Faults in smart grid systems: Monitoring, detection and classification. *Electr Power Syst Res*. 2020;189:106602.
- Jnaneswar K, Mallikarjuna B, Devaraj S, Roy DS, Reddy MJB, Mohanta DK. A real-time DWT and traveling waves-based multi-functional scheme for transmission line protection reinforcement. *Electr Eng*. 2020.
<https://doi.org/10.1007/s00202-020-01117-0>.
- Dasgupta A, Nath S, Das A. Transmission line fault classification and location using wavelet entropy and neural network. *Electr Power Compon Syst*. 2012;40(15):1676-89.
- Patel B, Bera P, Saha B. Wavelet packet entropy and RBFNN based fault detection, classification and localization on HVAC transmission line. *Electr Power Compon Syst*. 2018;5008:1-12.
- Costa FB, Souza BA, Brito NS. Real-time classification of transmission line faults based on maximal overlap discrete wavelet transform. In: *Proc IEEE Power Eng Soc Transmiss Distrib Conf*. 2012. p. 1-8.
- Anand A, Affijulla S. Hilbert-Huang transform-based fault identification and classification technique for AC power transmission line protection. *Int Trans Electr Energy Syst*. 2020;30(10):1-15.

de Alencar GT, dos Santos RC, Neves A. A new robust approach for fault location in transmission lines using single channel independent component analysis. *Electr Power Syst Res*. 2023;220:109281.

Ekici S. Support vector machines for classification and locating faults on transmission lines. *Appl Soft Comput*. 2012;12(6):1650-8.

Jamehbozorg A, Shahrtash SM. A decision tree-based method for fault classification in double-circuit transmission lines. *IEEE Trans Power Del*. 2010;25(4):2184-9.

He Z, Lin S, Deng Y, Li X, Qian Q. A rough membership neural network approach for fault classification in transmission lines. *Int J Electr Power Energy Syst*. 2014;61:429-39.

AsghariGovar S, Pourghasem P, Seyedi H. High impedance fault protection scheme for smart grids based on WPT and ELM considering evolving and cross-country faults. *Int J Electr Power Energy Syst*. 2019;107:412-21.

Arouche Freire JC, Garcez Castro AR, Homci MS, Meiguins BS, De Moraes JM. Transmission line fault classification using hidden Markov models. *IEEE Access*. 2019;7:113499-510.

Chen YQ, Fink O, Sansavini G. Combined fault location and classification for power transmission lines fault diagnosis with integrated feature extraction. *IEEE Trans Ind Electron*. 2018;65(1):561-9.

Chen K, Hu J, He J. Detection and classification of transmission line faults based on unsupervised feature learning and convolutional sparse autoencoder. *IEEE Trans Smart Grid*. 2018;9(3):1748-58.

Liang H, Liu Y, Sheng G, Jiang X. Fault-cause identification method based on adaptive deep belief network and time-frequency characteristics of travelling wave. *IET Gener Transm Distrib*. 2019;13(5):724-32.

Omar AMS, Osman MK, Ibrahim MN, Hussain Z, Abidin AF. Fault classification on transmission line using LSTM network. *Indonesian J Electr Eng Comput Sci*. 2020;20(1):231-8.

Gopakumar P, Reddy MJB, Mohanta DK. Adaptive fault identification and classification methodology for smart power grids using synchronous phasor angle measurements. *IET Gener Transm Distrib*. 2015;9(2):133-45.

Swain KB, Mahato SS, Cherukuri M. Expeditionary situational awareness-based transmission line fault classification and prediction using synchronized phasor measurements. *IEEE Access*. 2019;7:168187-200.

Yuill W, Edwards A, Chowdhury S, Chowdhury SP. Optimal PMU placement: A comprehensive literature review. In: *IEEE Power Energy Soc Gen Meet*. 2011. p. 1-8.

Zhang X, Guo Z, Zheng Y, Liu J, Yan P, Zheng L. Power grid fault diagnosis using polar PMU data plots. *Int J Electr Power Energy Syst*. 2022;141:108148.

Shazdeh S, Golpîra H, Bevrani H. A PMU-based back-up protection scheme for fault detection considering uncertainties. *Int J Electr Power Energy Syst*. 2023;145:108592.

Harish A, Prince A, Jayan MV. Fault detection and classification for wide area backup protection of power transmission lines using weighted extreme learning machine. *IEEE Access*. 2022;10:82407-17.

Alqudah M, Pavlovski M, Dokic T, Kezunovic M, Hu Y, Obradovic Z. Fault detection utilizing convolution neural network on timeseries synchrophasor data from phasor measurement units. *IEEE Trans Power Syst*. 2022;37(5):3434-42.

Rezaei Jegarluei M, Aristidou P, Fernandes W, Azizi S. Wide-area backup protection using sparse synchronized and unsynchronized PMU measurements. *IEEE Trans Power Del*. 2023.

<https://doi.org/10.1109/TPWRD.2023.3249822>.

Wang L, You ZH, Chen X, Xia SX, Liu F, Yan X, et al. A computational-based method for predicting drug-target interactions by using stacked autoencoder deep neural network. *J Comput Biol*. 2018;25(3):361-73.

Chen K, Hu J, He J. Detection and classification of transmission line faults based on unsupervised feature learning and convolutional sparse autoencoder. *IEEE Trans Smart Grid*. 2018;9(3):1748-58.

Kasun LLC, Zhou H, Huang GB, Vong CM. Representational learning with ELMs for big data. *IEEE Intell Syst*. 2013;28(6):31-4.

Zhang N, Ding S, Zhang J. Multi-layer ELM-RBF for multi-label learning. *Appl Soft Comput*. 2016;43:535-45.

Zhu W, Miao J, Qing L, Huang GB. Hierarchical extreme learning machine for unsupervised representation learning. In: *Proc Int Joint Conf Neural Netw*. 2015.

Yang Y, Wu QM. Multi-layer extreme learning machine with subnetwork nodes for representation learning. *IEEE Trans Cybern*. 2016;46(11):2570-83.

Texas A&M University. IEEE 39-bus system. Electric Grid Test Cases Library. Available from: <https://electricgrids.engr.tamu.edu/electric-grid-test-cases/ieee-39-bus-system/>. [Accessed 17 Jun 2026].

Cao W, Gao J, Ming Z, Cai S. Some tricks in parameter selection for extreme learning machine. *IOP Conf Ser: Mater Sci Eng*. 2017;261(1):012002.

Golub GH, Van Loan CF. *Matrix Computations*. Baltimore: JHU Press; 2013.

Yu F, Booth C, Dysko A, Hong Q. Wide-area backup protection and protection performance analysis scheme using PMU data. *Electr Power Energy Syst*. 2019;110:630-41.

Brown M, Biswal M, Brahma S, Ranade SJ, Cao H. Characterizing and quantifying noise in PMU data. In: *2016 IEEE Power Energy Soc Gen Meet*. 2016. p. 1-5.

Bhandari A, Yin H, Liu Y, Yao W, Zhan L. Real-time signal-to-noise ratio estimation by universal grid analyzer. In: *2019 Int Conf Smart Grid Synchronized Meas Anal*. 2019.

Harish A, Asok P, Vasudevan JM. Evaluation of wavelet transform based feature extraction techniques for detection and classification of faults on transmission lines using WAMS data. *Adv Electr Electron Eng*. 2023;20(4):405-17.