Energy Reports 8 (2022) 10168–10182

Contents lists available at ScienceDirect

Energy Reports

journal homepage: www.elsevier.com/locate/egyr

Research paper

On the protection of power system: Transmission line fault analysis based on an optimal machine learning approach

Md. Sihab Uddin [a], Md. Zahid Hossain [a], Shahriar Rahman Fahim [b], Subrata K. Sarker [a], Erphan Ahmmad Bhuiyan [a], S.M. Muyeen [c,*], Sajal K. Das [a]

[a] Department of Mechatronics Engineering, Rajshahi University of Engineering & Technology, Rajshahi 6204, Bangladesh
[b] Department of Electrical and Electronic Engineering, American International University-Bangladesh, Dhaka 1229, Bangladesh
[c] Department of Electrical Engineering, Qatar University, Doha 2713, Qatar

ARTICLE INFO

Article history:
Received 5 February 2022
Received in revised form 21 July 2022
Accepted 28 July 2022
Available online 19 August 2022

Keywords:
Transmission line
Supervised learning
Fault diagnosis
Wavelet transform
Optimized model
Noise immunity

ABSTRACT

Transmission lines (TLs) of power networks are often encountered with a number of faults. To continue normal operation and reduce the damage due to the TL faults, it is a must to identify and classify faults as early as possible. In this paper, the design and development of an intelligent machine learning framework is presented to identify and classify faults in a power TL. The design of the proposed framework is done with the goal of reducing computational load and ensuring resilience against source noise, source impedance, fault strength, and sampling frequency variation. The design is carried out based on the selection of the optimal model parameters using a search optimization algorithm called GridSearchCV. The effectiveness of the proposed model is verified by testing the model on the IEC standard microgrid model in a MATLAB environment. The results show that the proposed model has more than ninety-nine per cent overall accuracy in the identification and classification of the TL faults. The results are also compared with some state-of-the-art approaches such as LSTM, RNN, DBN, DRL, and CNF to further examine the performance of the proposed framework. The comparison demonstrates that the proposed model outperforms other existing techniques in terms of accuracy, computational cost, and response speed.

1. Introduction

The transportation of electricity from generating stations to substations is referred to as electric power transmission. Transmission lines (TLs), an important component of a power grid, are exclusively responsible for carrying electricity from a producing station to substations and consumers (Godse and Bhat, 2020). Every day, new TLs are being installed to meet the intemperance demand for electricity (Godse and Bhat, 2020). Nevertheless, transmission lines are prone to a variety of faults. The faults in TLs are primarily categorized as open-circuit faults and short-circuit faults (Reddy and Mohanta, 2008; Fahim et al., 2021). A short-circuit fault, the severe one, follows when a phase to ground or a phase to a phase joint connection arises. When such faults occur, they not only disrupt power delivery but also raise the risk of damage to home properties as well as the entire electrical infrastructure. As a result, to seize control over the faults, a fast diagnosis of the faults is essential.

Over the last few years, there have been continuous research trends in the identification and classification of TLs faults. Several approaches are being attempted to get over the identification and classification of the TL faults. One of the earlier attempts was the measurement of the symmetrical components of reactive power for single-circuit TLs (Mahamedi and Zhu, 2013). However, this method suffers due to the presence of a propagation delay that affects the accuracy (Chen et al., 2016b). While the integration of phasor measurement units provides a solution, the uncertainty of dealing with missing data is a concern (Jiang et al., 2000). Another technique used in TL fault diagnostics is the traveling wave technology. This technique is only used for fault location finding (Chen et al., 2018). The fault classification problem necessitates a different approach integration. Another concern with this method is establishing higher-value sample rates, which is challenging (Tang et al., 2016).

The adapted techniques, as discussed earlier, suffer mostly because of the use of raw signals. The essential features of the original signal are not observable using these methodologies. Noises in the signal are common, and the features can be overly complex at times (Dutta et al., 2014). Therefore, researchers adopt techniques that are built on top of the special signal processing algorithm, and Fourier transform (FT) comes first on the list (Taheri et al., 2018). FT shows its proficiency, where both the time and

* Corresponding author.
E-mail address: sm.muyeen@qu.edu.qa (S.M. Muyeen).

https://doi.org/10.1016/j.egyr.2022.07.163

frequency of a signal are constant, along with a stationary signal analysis (Abdollahi and Seyedtabaii, 2010). On the contrary, voltage and current signals could be non-stationary which means the frequency could vary. Short time Fourier transform (STFT) is found to overcome the limitations of FT, but it is not adaptive for a different state of a signal, because of its fixed window size. Comparatively, s-transform (ST) is a good alternative, as its window size varies with the frequency of the signal, and it can be used for both finding the location (Ahmadimanesh and Shahrtash, 2013), and the type of fault (Dash et al., 2007). However, fault location estimation using ST is more common, and the integration of pattern recognition algorithms is observed for the faulty phase detection (Roy and Bhattacharya, 2015). ST is a viable solution, but the computation power that it requires and the processing time are things to be considered (Dash et al., 2015).

Wavelet transform (WT) is an eminent feature extraction technique (FET) (Livani and Evrenosoglu, 2013). In addition to providing time-domain information, it also provides frequency data of a signal. It correspondingly reduces noises in a signal by concentrating it into large magnitude coefficients of wavelet where the noises from coefficients of smaller magnitude are removed without compromising the quality of the main signal (Costa et al., 2012). Furthermore, it reduces the size of the dataset by converting a sequence to its energy value, therefore, making the computation simpler (Fahim et al., 2021). The availability of a wide range of wavelet functions makes it challenging for researchers to select the right version of mother wavelet for a given signal. Authors in Safavian et al. (2005) report that there are some basic properties that need to be considered while choosing a mother wavelet which determines the correlation and stability of a given signal with the wavelet function. These properties are the similarity in waveform between the wavelet function and the signal to be analyzed, vanishing moments, and computational complexity. Daubechies 4 (DB4) matches those properties for power system transients and most of the literature presents the implementation of this wavelet function for TL fault analysis. WT overcomes the limitations of each of the FETs mentioned earlier, and FIE accuracy implementing WT is promising (Jurado and Saenz, 2002).

FIE approaches that use machine learning (ML) algorithms provide a great advantage as they learn from the data and find useful features. A variety of machine learning methods have been used to detect and classify transmission line faults. Two algorithms, support vector machine (SVM) and decision tree (DT) are found to be of utmost use in the field of TL fault analysis (Mishra and Ray, 2018). SVM is a statistical learning theory-based algorithm that works with large data and efficiently computes (Dash et al., 2006). However, SVM is sensitive to its parameter selection (Duan et al., 2003). The DT solves tree-like graphs to make decisions and increases the complexity in computation for multi-class problems (Raza et al., 2020).

Artificial neural network (ANN) is known for recognizing patterns in data (Jain et al., 2009). In Fathabadi (2016), an FIE approach is proposed that utilizes an FIR filter for signal processing and ANN to detect and classify faults. However, it combines SVM and SVR models, which may increase the computation cost, and the sample size of the signal seems to vary. The use of microcontroller based approach for fault detecting and application of ANN for fault classifying is proposed in Koley et al. (2016). It uses current and voltage signals sampled at 1 kHz to identify fault and no feature extraction technique is used.

The use of combined neural net architectures, such as CNN-LSTM (Wang et al., 2022), R-CNN (Zhai et al., 2021), GSV-CNN (Han et al., 2021) is found in the paradigms of TL fault analysis. CNN is powerful and reliable when the input data is image or videos. The TL fault data obtained numerically is required to reprocess an image to work with CNN for best results. In case of LSTM, sequential data is required. Additionally, the performance of these algorithms is dependent on the layer design. More the layers more the computational load (Rai et al., 2021).

The above explanation clearly demonstrates that machine learning techniques, in particular deep learning techniques, are often used in the identification and classification of TL faults. The challenges of most deep learning algorithms are that they require a dataset containing a large amount of data, high computational power, and suffer from problems with a set of rules to select the architectural model. Transmission line fault data is often limited to small numbers and is computed with noise in real-time. The identification and classification of such faults are required to be done as soon as possible to rescue the normal operation.

This paper proposes a simple machine learning architecture that automatically selects its model parameters irrespective of the low to the high variation of data in a dataset and guarantees robustness against unwanted noise. The proposed method does not compromise with the state-of-the art methods such as ANN, CNN, and DBN in terms of accuracy and computational power. The proposed solution is important in the context of the existing literature because of its simplicity, low computational power, and high accuracy.

1.1. Contributions

The innovation of this research is to introduce a simple but self-optimized effective machine learning approach for transmission line fault diagnosis which can guarantee the high level of performances for any given dataset with consuming and utilizing the less computational time and hardware. The main findings of this work are as follows:

- Design and development of a statistical machine learning method that does not require a large quantity of data which is less computationally complex and utilizes minimal hardware.
- Integrate a wavelet-based feature extraction technique to obtain specific features of the signals by filtering the signal and converting to a sequence of data into a single feature.
- Develop a model parameter selection technique with proposed machine learning method in order to auto-tune the hyper parameter of the proposed ML model each time new data is input. The proposed model combined parameter selection technique offers the same level of performances when dealing with the uncertainties data presented in the power system.

The rest of the paper is organized as follows: After the introductory section, a System Modeling section is presented that was used to generate the dataset and validate the classifier performance. The Signal Processing, The Proposed Method, Result and Discussion, and the Conclusion sections are presented, respectively.

2. System modeling

Simulink software was used for designing and simulation of the power system model. Combination of two–three-phase source-based system, and line length of 100 km constitutes the model, as shown in Fig. 1. In the initiation of the sources, positive and zero sequence resistances are considered and assigned a value of 0.01274 Ω/km and 0.3863 Ω/km, respectively. The system is capable of generating 220 kV phase to ground voltage at 50 Hz frequency. We can see in Fig. 1 that T-1 and T-2 transformers are used for the step-down purpose. The system could be designed with a wide variety of parameter settings.

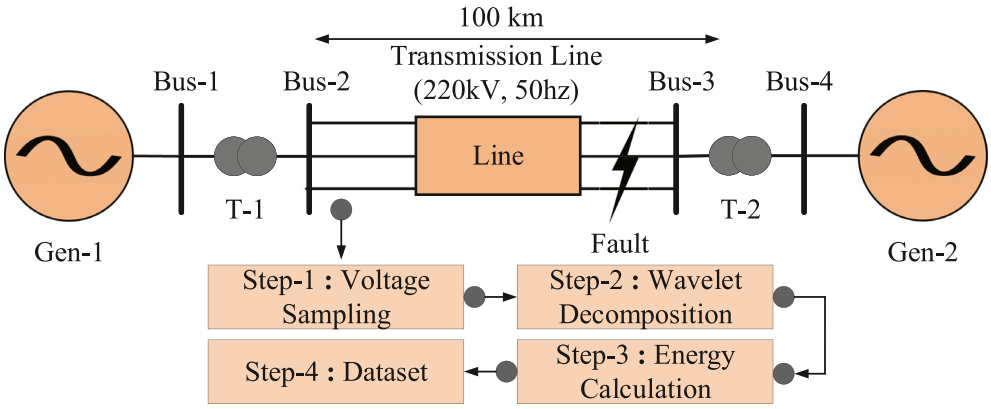


Fig. 1. The power system model designed in Simulink.

Table 1
Simulink system parameter used for simulation and data generation.

Attributes	Values
Different types of fault	AG, BG, CG, ABG, BCG, CAG AB, BC, CA, ABCG
Fault resistance (Ω)	0.1, 0.2, 0.3, 0.4, 0.5
Fault distance (km)	1 to 99
Gen-1 Gen-2	220 kV, 600 MVA, 50 Hz

Different literates utilize different configurations, so, it can be a 25 kV 50 Hz or 440 kV 60 Hz source-based system. However, the focus is on designing a machine learning-based intelligent fault identification and estimation system; therefore, a benchmark topology (Chen et al., 2016a) is adopted in this research. Higher sampling frequency can provide more detailed information about a signal. A detailed signal contains more data points and precisely presents the impact on the signal due to the fault event or noise or any abrupt changes. Accordingly, the generated voltage signal is sampled at a sampling rate of 20 kHz that gives exactly 400 samples of the respective signal for one complete cycle. Different observation using the benchmark topology is discussed in the result section and the generated signal is presented in Fig. 12.

The incident of fault into the system is performed using Simulink three-phase fault block utility. It can generate all the fault types, as presented in Table 1. Some other system parameters considered when designing the system and switched during simulations are also presented in detail Table 1.

3. Data processing

The real world power system data are often noisy. The design of the FIE framework allow it to accurately classify fault types, even though the data are noisy. As a part of the scheme, a special signal processing scheme is deployed that can reduce noise from the signal and calculate the energy of the signal.

3.1. Signal processing

WT is initially used for working with voice and image data. However, its specialty in localizing frequency components of different range made researchers think of using it in other fields also. Consequently, the output provided is far better compared to other signal processing techniques, both in terms of usefulness and computational efficiency. There is a basic wavelet function at the core of wavelet transform called the mother wavelet which stretches a signal by translating and dilating to generate a wavelet based representation. The equation for mother wavelet is obtained as,

$$\phi_{(d,t_r)}(t) = \frac{1}{\sqrt{d}}\phi\left(\frac{t-t_r}{d}\right) \tag{1}$$

d represents the dilation factor that can stretch a signal or shrink it, and helps to capture the abrupt changes in the signal, where t_r is the position factor that moves the wavelet through the signal. If a signal $v(t)$ is continuous in time, then the equation of its Continuous Wavelet Transform (CWT) can be derived as,

$$CWT(d,t_r) = \frac{1}{\sqrt{d}}\int_{-\infty}^{\infty} v(t)\phi\left(\frac{t-t_r}{d}\right)dt \tag{2}$$

The result produced by CWT generates too many coefficients which are functions of d and t_r. It calculates the coefficients for every possible dilation that makes the analysis even harder. In contrast to CWT, the discrete wavelet transform (DWT) uses discrete dilation and position factors which helps in a compressed representation of the signal and provide output in a more understandable manner. The DWT of a signal can be performed by using the following equation,

$$DWT(v,p,q) = \frac{1}{\sqrt{d_0^p}}\sum_{j=1}^{\infty} v(j)\phi\left(\frac{q-jd_0^p}{d_0^p}\right) \tag{3}$$

in the above expression, d is replaced by d_0^p and t_r by the term $jd_0^p t_0$ where $d_0 > 1, t_0 > 1$ and j, p belongs to the set of positive integers.

The DWT uses high and low pass filters to decompose a given signal in terms of two different frequency ranges. The detail coefficients contains the high-frequency components while approximation contains the low-frequency components. Fig. 2 shows the decomposition process for this particular research. A real world is system is more likely to have noises in the signal. Some random noise is also introduced into the signal to verify that the noisy signals can also be handled using the proposed FIE approach. The figure shows the signal gets totally noise free at the end of the decomposition.

Literatures show using Daubechies wavelets with level 3 or 4 decomposition works better with voltage and current signals (Bhowmik et al., 2009). We found that the decomposition of level 4 gets the best result in this case. Although, level finding is a trail and error process, sparsity plot (Bekerman and Srivastava, 2021) is now being used to determine the exact level of decomposition for a particular signal. Fig. 3 presents the design process of wavelet based feature extraction for this research. The sparsity plot sends a signal by showing abrupt changes in the curve after the optimal decomposition level is reached. In Fig. 3(a), sparsity plot for noise free signal is presented. An abrupt change in sparsity value is observed after level 2. Therefore, level 2 is the optimal decomposition level for this signal.

For the noisy signal in Fig. 3(b), the change in sparsity value is very insignificant till level 4. After that a huge variation is

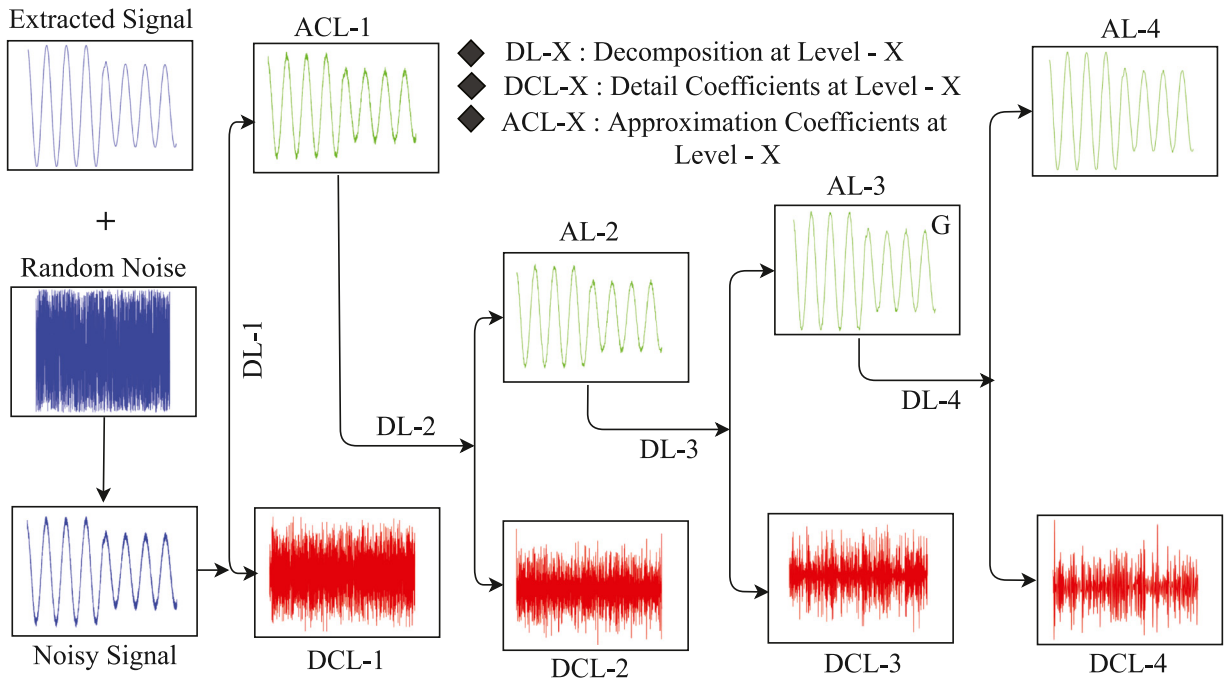


Fig. 2. Signal denoising and feature extraction using wavelet.

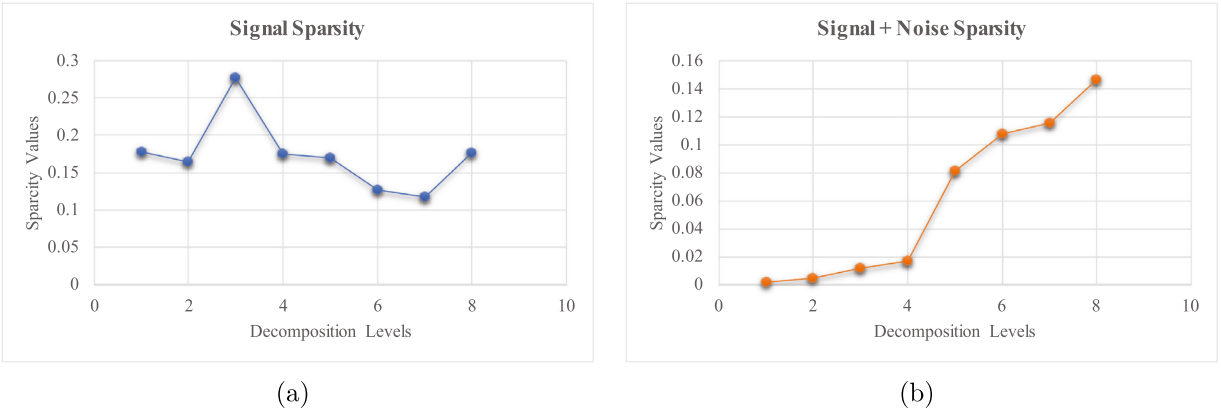


Fig. 3. Design process of wavelet based feature extraction (a) Signal Sparsity Plot (b) Signal + Noise Sparsity Plot.

observed. Thus, level 4 is selected as the optimal decomposition level for this research. When a fault occurs, the resulting curve of voltages and currents also varies. The useful features are extracted from those signals, in other words, we calculate the energy of the resulting curves to generate our dataset. Three different values of energy is obtained for three different phases and this combination varies for each fault type. We use the following equation to calculate the energy,

$$E_v = \sum_{j=1}^{n} |c_j|^2 \tag{4}$$

where, E_v is the energy of a voltage signal, c_j is the coefficient vector and n is the population of the detail coefficients.

3.2. Data generation

The values of fault resistance (FR) at which the dataset is generated are 0.1, 0.2, 0.3, 0.4, and 0.5, as can be seen from Table 1. The line length is set in the range 1 to 99 km for each of previously mentioned values of FR. The fault type is introduced manually into the power system designed in Simulink environment. A MATLAB script varies the Simulink parameter programmatically, runs the simulation, calculates energy from the measured voltage data, and stores them into tabular data format until it meets the dataset requirement. There are 495 data samples for each fault type and a total of 4950 data samples for all the classes present in that data table. A quantitative analysis of the relation among fault resistance, distance and signal energy is Table 2.

The table is a tiny slice of the main dataset that presents the variation in energy with respect to fault resistance and fault distance. The selected fault type is Phase B to the ground (BG). However, the effect of fault is same and depending on the fault type the energy of the corresponding phases are affected. In this case, the energy of the faulty phase is significantly higher than that of the other two phases. The table presents an analysis at two instances of fault distance. They are at 1kilometer (km) and 3 km. When the distance is constant and the fault resistance is increasing, a proportional increase in energy of phase B is observed. However, the increasing distance shows an inverse proportional relationship with the phase energy. The energy is

Table 2
Quantitative analysis of fault resistance and fault distance.

Sl. No.	Fault type	Fault resistance	Fault distance	Energy of Phase A	Energy of Phase B	Energy of Phase C
01	BG	0.1	1	98227.29	179271.78	99051.32
02	BG	0.3	1	98219.84	179307.17	99054.27
03	BG	0.5	1	98212.49	179343.14	99057.24
04	BG	0.1	3	97963.69	172525.19	98593.24
05	BG	0.3	3	97957.65	172569.76	98595.56
06	BG	0.5	3	97951.76	172615.11	98597.96

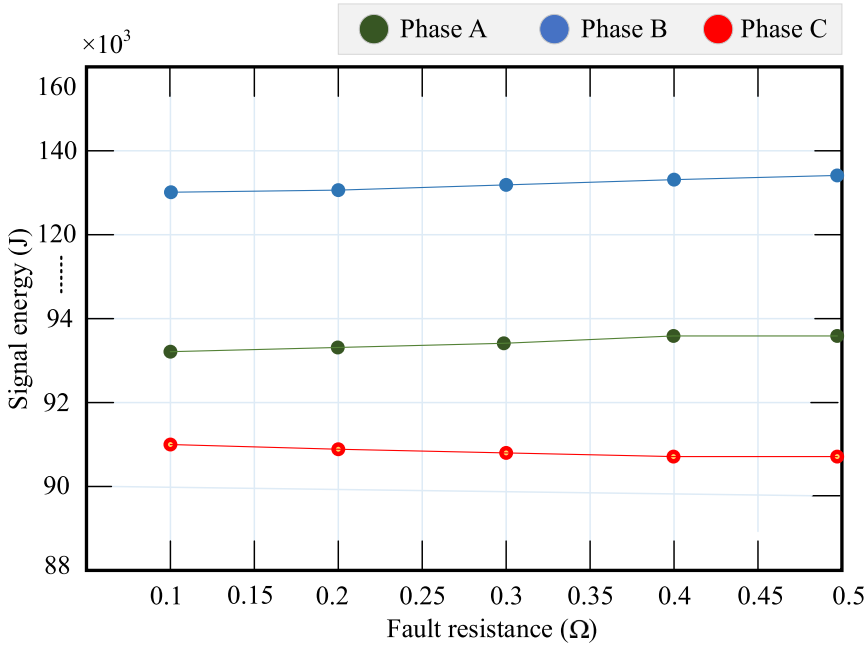


Fig. 4. Variation in signal energy due to the change in fault resistance.

much lower at 3 km distance than energy at 1 km distance. In detail analysis of these two factors is presented in the following sub-sections.

3.2.1. Fault resistance-energy relation

The graph in Fig. 4 expresses how the signal energy is affected by the fault resistance variation. Phase B is the faulty phase in this case, and it can be observed from the graph. The energy levels of phase A and phase C are close, while phase B is at a great distance. Due to the change in fault resistance, the value of energy increases for phase B and phase A. Therefore, the amplitude of signal energy of the faulty phase has a proportional relation to fault resistance . However, a very small fluctuation is observed in phase C. Additionally, the observation exhibits that the fault in a particular phase affects the energy of the other phases as well.

3.2.2. Fault distance energy relation

The magnitude of faulty phase energy is also affected by the location where the fault has occurred. The dataset is generated considering the total line length is 100 km, and the data samples are taken at every 1 km distance. It can be seen that, from Fig. 5, the signal energy of the faulty phase, in this case, phase B, changes inversely with gradual increment of the distance between the fault location and the fault measuring point. Each of the phases' signal energy reaches an identical level when the fault distance is between 40 km to 80 km. The phase B signal energy even decreases more than the level phase A and phase C maintaining while the distance goes over 80 km. The effect of fault distance on the signal energy is prominent, as it can make the data very difficult to distinguish the faulty phase. The energy fluctuation for the other two phases is very little.

4. Proposed method

4.1. Model derivation

The proposed FIE approach that incorporates an ML algorithm to learn from the data and predict fault type is based on a theory called statistical learning. It takes training data as input and maps them into an alternate dimension that is called a high-dimensional feature space, where it can find a separator between two classes, if we consider it a binary classification problem. The separator is called a hyperplane. A margin denotes the distance between the nearest data points of a particular class and a hyperplane. The degree of a good fit hyperplane is depended on the expansion of the margin. The equation of the hyperplane is given by the following equation,

$$h(z) = w^T z + l = 0 \tag{5}$$

where, w stands for a n-dimensional vector, c is sets the position of the hyperplane, and $h(z)$ represents the hyperplane itself, which is also termed the decision boundary. If the input data is $z_1, z_2, \ldots, z_N$ in this case, N represents the number of input data samples belonging to two classes $c_j = 1$ for class-1 and $c_j = -1$ for class-2 then $h(z) \geq 1$ refers to class-1 and $h(z) \geq -1$ refers otherwise.

This leads us to a general conclusion that could be derived as,

$$h(z)c_j = c_j(w^T z + l) \geq 1 \tag{6}$$

If the algorithm could set the hyperplane in a way that the margin is maximum then it is known as a optimal hyperplane,

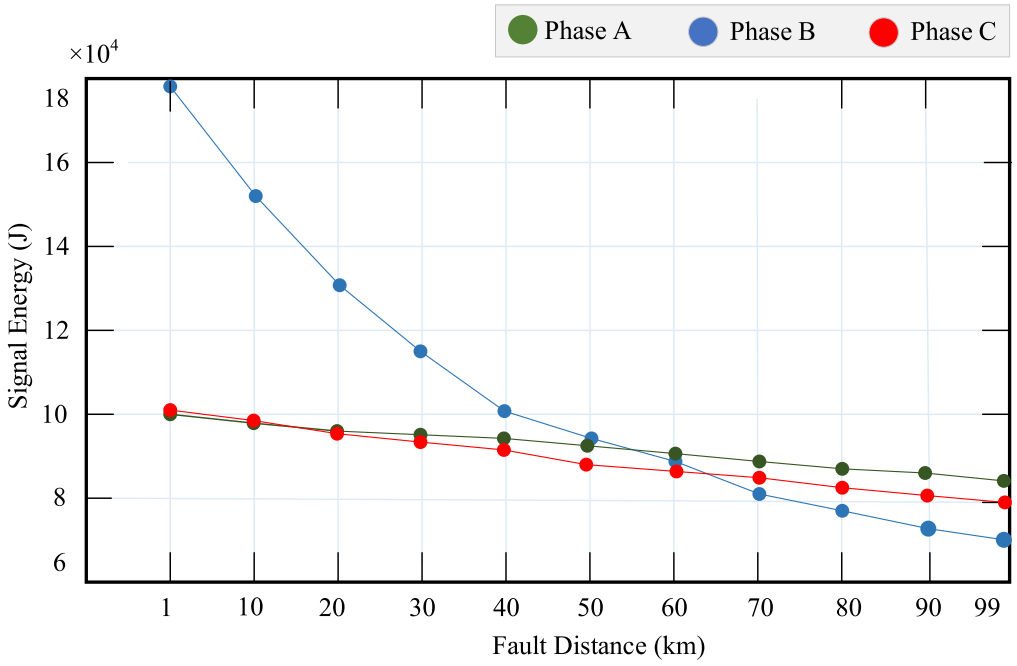


Fig. 5. Variation in signal energy due to the change in fault distance.

and in that case, the mathematical representation of the geometric margin becomes $\|w\|^{-2}$. The following equations are used to find a optimal hyperplane, and we take the help of convex quadratic-optimization to solve for it. Minimize,

$$\phi(w,\zeta) = \frac{1}{2}\|w\|^{-}2 + c\sum_{j=1}^{N} \tag{7}$$

so that, $h(z)c_j \geq 1-\zeta_j, \zeta_j \geq 0, j = 1, 2, \ldots, N$. If a training sample x_j is placed on the wrong size of the margin, then ζ_j denotes the distance, because it and C sets the penalty for errors. Kuhn–Tucker conditions and Lagrange dual problem can help simplify the above calculations; maximize,

$$W(\alpha) = \sum_{j=1}^{N}\alpha_j - \frac{1}{2}\sum_{j,k=0}^{N}\alpha_j\alpha_k c_j c_k z_j^T z_k \tag{8}$$

subject to, $\sum_{j=1}^{N} C_j\alpha_j = 0, C \geq \alpha_j \geq 0, j = 1, 2, \ldots, N$. If α^* and w^* are the optimal solutions of the dual problem and α^* is not equal to zero only, then the Karush–Kuhn–Tucker theorem holds the equality condition in Eq. (2). The training examples are considered as support vectors (SVs) in this scenario, and they are considerably lower than the training data points z_j. Now, the value of the horizontal bias could be derived as,

$$l^* = -\frac{1}{2}\sum_{S} VsC_j\alpha_j^*(s_1^T z_i + s_2^T z_i) \tag{9}$$

in the above equation, s1 and s2 are any two support vectors, belonging to class-1 and class-2 respectively. Considering only the samples that are associated with the SVs, the final decision function is derived as,

$$h(z) = \sum_{SVs}\alpha_j c_j z_j^T z + l^* \tag{10}$$

An unlabeled data sample x is now classified as class-1, *if*$h(z) \geq 0$, or class-2 otherwise. The equation above can only be used for linear problems. The proposed algorithm can also work with the non-linear ones, utilizing a special feature called the kernel function. The data samples are mapped into a high-dimensional feature space, but in this case, the feature vector used is also non-linear. Considering, $\phi(z) = \phi(z_1), \phi(z_2), \ldots, \phi(z)$ as the feature vector function, the equation for the decision boundary is given as,

$$h(z) = \sum_{SVs}\alpha_j c_j \phi^T(z_j)\phi(z) + l^* \tag{11}$$

The high dimensional feature space helps when working with complex functions, however, it is also the reason of computational problems. In addition, there is a chance that it could cause overfitting. This problem is solved by using the kernel function that returns the dot space feature mapping of the original data points. Now, we obtain the following equation by substituting the kernel function $K(z_j, z) = \phi^T(z_j)\phi(z)$ in Eq. (10),

$$h(z) = \sum_{SVs}\alpha_j^* c_j K(z_j, z) \tag{12}$$

Now, Eq. (11) is used to classify data samples as before, with the additional advantage of being able to handle non-linear problems. If $i, j = 1, 2, \ldots, N$, then $K(z_i, z_j)$ forms a kernel matrix, and it is the basic structure in kernel theory. The function for the polynomial kernel is given as,

$$K(z_i, z_j) = (z_i^T z_j + 1)^n \tag{13}$$

Additionally, the rbf kernel function is given as,

$$k(z_i, z_j) = exp\Big\{\frac{-|z_i - z_j|^2}{2\sigma^2}\Big\} \tag{14}$$

where, σ represents the Gaussian function width. Equations (10) and (11) both follows Mercer's kernel matrix theorem necessary for it to be used as the kernel (Lundgren et al., 2001). The polynomial or rbf kernels are used when the classes are not separable in lower dimensional feature space.

The class separation process is shown in Fig. 6, which is plotted using the t-SNE algorithm. This algorithm is used to visualize data points located in higher dimensional feature space. It creates a 2D or 3D map of the higher dimensional data and projects the output, reducing its dimension. The data points in Fig. 6a state data analyzing starting scenario. It can be observed how

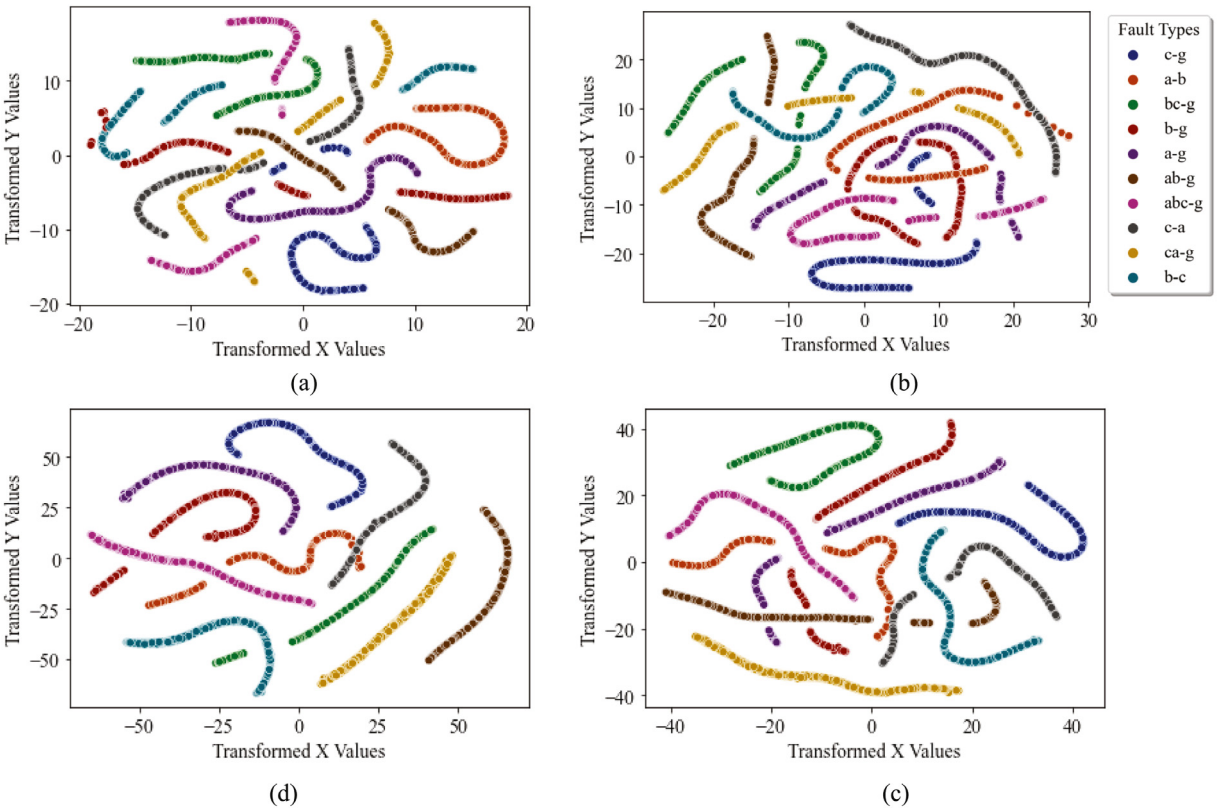


Fig. 6. Visual representation of the data points separation process.

each class is separated in each step. The classes are completely separated by a hyperplane when the model is completely trained at the final stage shown in Fig. 6d.

An ML model is said to be optimized when it provides optimal output using the resources that it has. In this case, the hyper-parameters are the resources that the proposed ML algorithm possesses that can be tuned in a manner that would result in fitting the data points properly. When defining the model, the hyper-parameters can be set; however, the problem is that there exists no defined set of values for the hyper-parameters. The values can be different for different problems. One way to find the values is to run the program in a loop until it reaches maximum accuracy. This iterative process takes a huge amount of time and causes computational inefficiency. GridSearchCV optimization (GSO) provides a solution to these concerns.

4.2. GridSearchCV optimization

If we be break down the term "GridSearchCV", we get GridSearch and CV. First, all the hyper-parameters are selected based on which GSO constructs the GridSearch space. In this research, the kernel function, C parameter, degree and gamma are the selected hyper-parameter for tuning the model accuracy. The influence of these hyper-parameters on the proposed model and the defined set of values are presented in Fig. 6. The GSO algorithm tries each of the hyper-parameter combinations found at the intersection of the grids, and then move forward to the next one.

Now, it is time to have a look at the term CV. The term CV stands for cross-validation. The GSO algorithm performs several tests on the dataset to check what kind of hyper-parameter combination gives the best result. However, if the tests are performed with the same data, the result might not be efficient. CV is a technique to spilt a dataset into k small sets of data. The variable k is manually defined. In this case $k = 5$, and let us assume the small sets of data are $y1, y2, y3, y4, y5$. Now, GSO selects $y2, y3, y4, y5$ as the training set while $y1$ is selected as test set and finds the accuracy of the model. In the next iteration, $y2$ is selected as the test data and the remaining are the training data and continues.

$$AverageAccuracy = \frac{\sum_{j=0}^{k} Accuracy_j}{k} \tag{15}$$

Though k-fold CV is applied, there is a possibility that in those sets, one class could be found in a higher proportion and in the other set; another could have the higher proportion. An advanced technique called stratified k-fold CV is used to ensure the right class distribution proportion in the small data sets. Fig. 8 shows the dataset fragment preparation process. In this method, first the main dataset is divided into 80% training data and 20% test data. After that, the first 20% of the 80% training data and the first 20% of 20% test data is selected and a dataset fragment is formed using which the model is trained and tested. The accuracy of the test result is stored into the memory. It completes Fold-1 process. The same process is repeated for Fold-2 except in this case the 2nd 20% from each of the set is selected. The process keeps executing for k number of folds. At the end of k folds the average accuracy of the model is calculated using Eq. (15) and stored into the memory. After that, another combination of hyper-parameters is selected and average is accuracy of the k folds testing is calculated. The algorithms keeps comparing the average accuracy to find the highest accuracy among the test results. This process keeps running until it reaches the end of the all the possible combination of hyper-parameters.

At the end of the process, the optimizer provides the hyper-parameters for which maximum average accuracy was obtained.

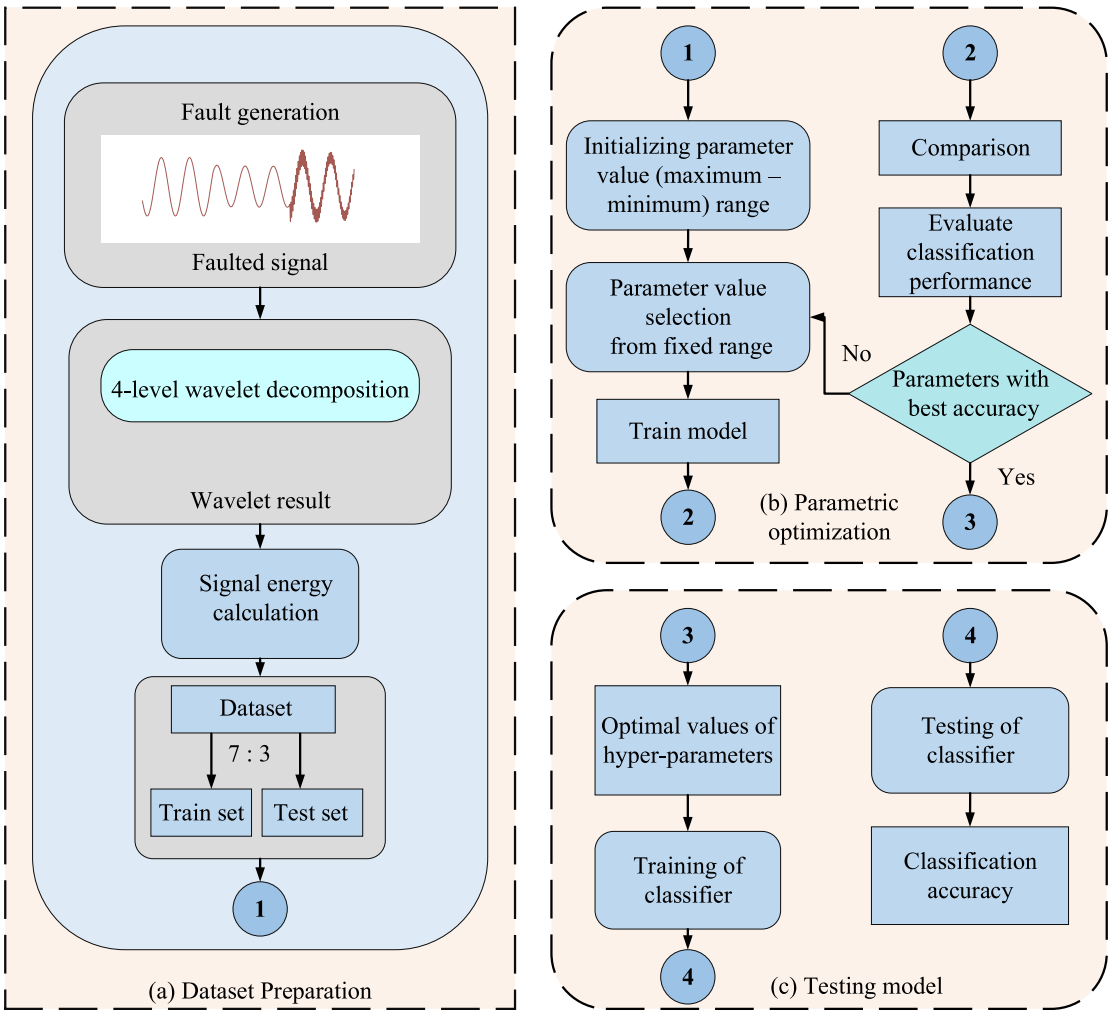


Fig. 7. Schematic representations of the model combined with GSO.

The optimization process is only necessary when a completely new dataset is provided to the model which helps the model optimize its parameters automatically. Fig. 7(a) presents the feature extraction and data generation process, while Fig. 7(b) and (c) summarizes the hyper parameter selection process.

5. Results and discussion

The FIE framework is tested with the remaining data that was kept separate earlier for performance evaluation purposes. The result will not be reliable if the model is tested with the same data it was trained with because in that case the model will provide maximum accuracy as the data is already known to it. As a preventive approach, the dataset is divided into a training set which contains 70% of the total data and a test set with 30% data. The model generates a classification report (CR) as an output when the test data is applied as an input. The CR is used to plot the confusion matrix (CM) that summarizes the classification results and helps visualize the performance of the ML model. The size of the CM depends on the number of the output classes in a dataset, in this case, it is 10.

The tables in Fig. 9 are the graphically enhanced version of the CM that was generated by the model. Fig. 9 a shows how well the model classifies each type of fault. If we consider it as a square matrix, for illustration, then the primary diagonal represents the number of each fault type that the model can correctly classify. However, eight of the ten classes were predicted accurately, “b-g” and “c-g" faults confuse the model. The CM shows 20 of the “b-g" faults are labeled as “a-g", 31 as “c-g" and 15 as “b-c" fault types. Additionally, six of the “c-g" faults are predicted as “c-a" faults. The obtained accuracy of “b-g" and “c-g" fault class is 59.5% and 95.68% respectively, and for other classes, it is 100% . The overall prediction accuracy is about 95.93%.

The accuracy after using the optimized algorithm is shown in Fig. 9b . The algorithm changed the classification result drastically from 95.93% to 99.46% and reduced the computation time to a great extent. As can be seen from the CM in Fig. 9b, that misclassified results decrease to a great extent. There exists no fault prediction for "c-g" anymore and "b-g" accuracy improved to 93.87%. Fig. 10 might help understand what is going on in the background that is causing this drastic change in accuracy. The choice of appropriate hyper-parameters is crucial for the proposed model. The optimizer is automatically choosing this for us.

If we look closely at Fig. 10(a) and (b), we see how the data is fit into the model. In this case, the used kernel is linear so, the model cannot properly separate two of the classes using the decision boundary as it can be seen it Fig. 10(a). The process of fitting the data into the hyperspace is observed though objective function model shown in Fig. 10(b)

In the second case, the kernel used is polynomial and we can see in Fig. 10(c) the decision boundary is shaped to support the data points properly which leads in higher classification accuracy. The 3D visualization presented in Fig. 10(b) support twisting

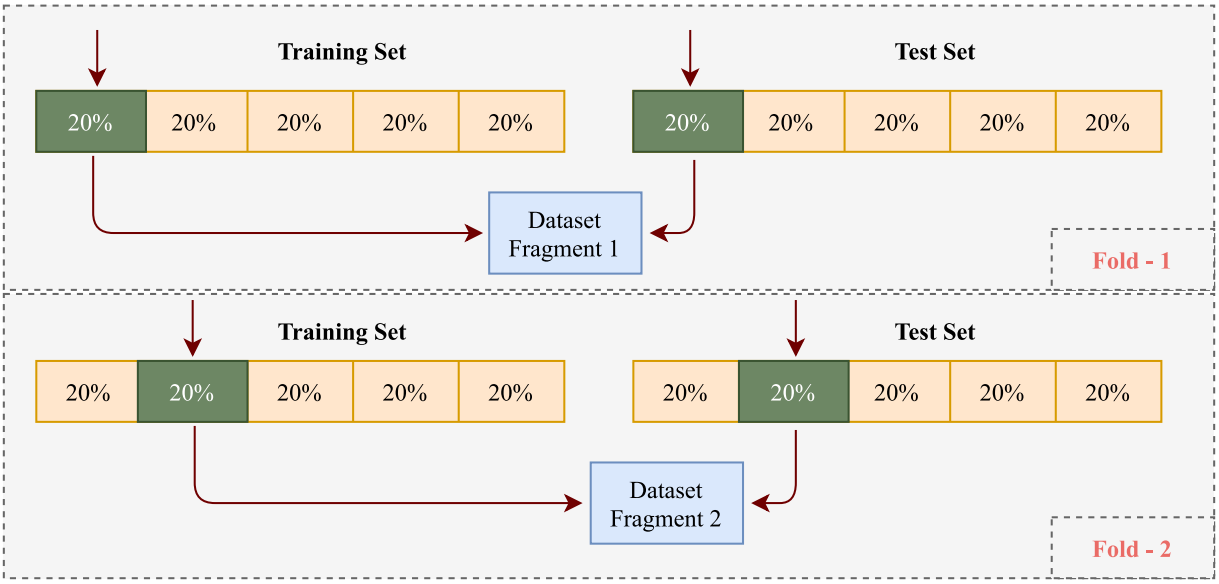


Fig. 8. Dataset fragment preparation process.

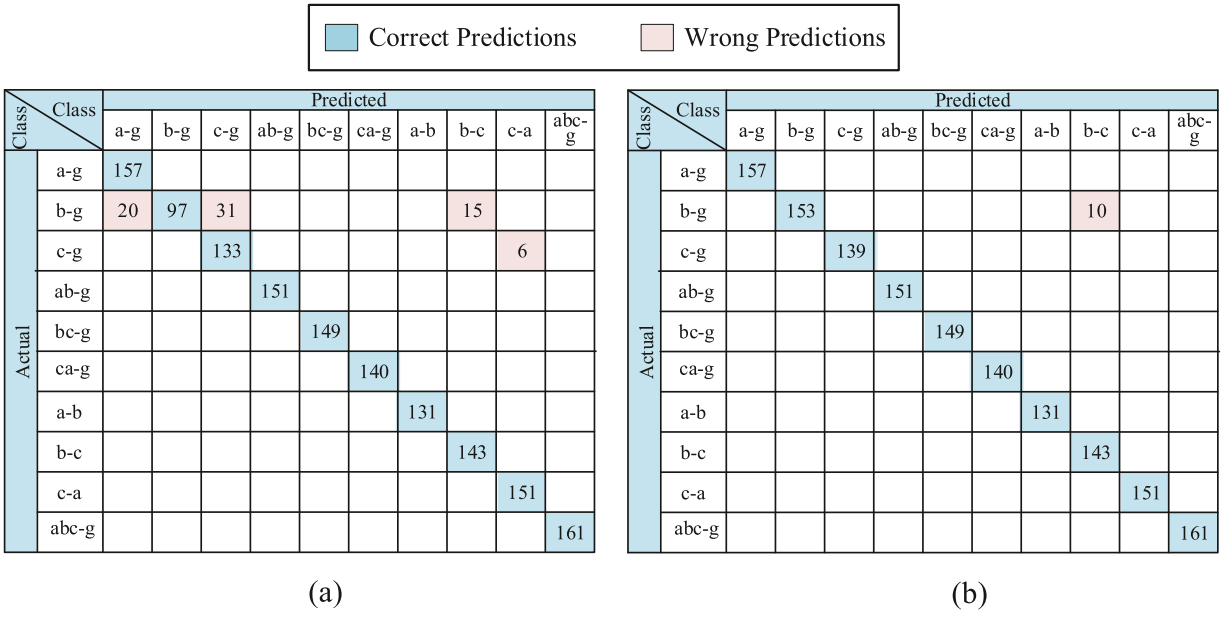


Fig. 9. Confusion matrix generated by the proposed model; (**a**) Before optimization, and (**b**) After optimization.

Table 3
Performance Metrics Obtained for the Designed Model.

Performance metrics	Formula	Before optimization	After optimization
Overall Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$	95.93%	99.46%
Mean Precision	$\frac{TP}{TP+FP}$	0.96	0.99
Mean Recall	$\frac{TP}{TP+FN}$	0.96	0.99
Mean F1_Score	$\frac{2TP}{2TP+FP+FN}$	0.95	0.99

of decision boundary. In an addition to the accuracy, different metrics obtained from the classification report of the given model are presented in the table in Table 3, which helps to provide a better idea of how the system performance is affected by the optimization algorithm.

The graphic in Fig. 11 exhibits which fault condition affects the model prediction. The prediction result was divided into five categories, before and after applying the GSO is observed. The model can predict double line-to-ground, line-to-line, and all the lines-to-ground fault with 100% accuracy. Single line-to-ground (LG) is the case in which the model generally makes the wrong choice, in effect, only 84.31% accuracy is achieved. The model gets to set the decision boundary in a more precise way when the GridSearchCV Optimizer (GSO) is deployed. The model

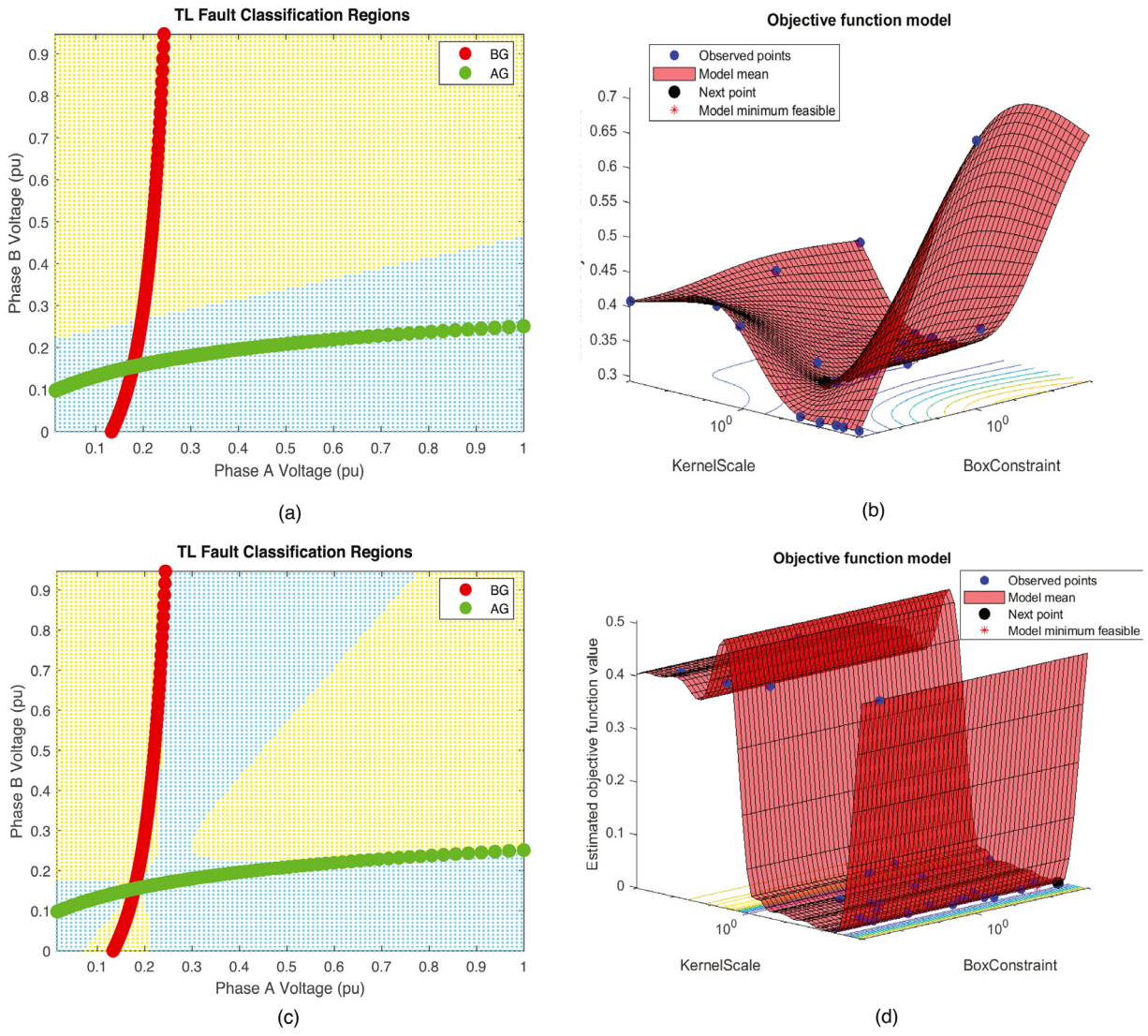


Fig. 10. (a) Plot of the decision boundary using linear kernel, (b) The objective function model for liner kernel, (c) Plot of decision boundary using polynomial kernel (d) The objective function model for polynomial kernel.

hyper-parameters have been tuned accordingly. The increase in accuracy for the LG fault type was significant. Consequently, the overall prediction accuracy gets a raise.

The robustness of a system depends on the consistency of its performance under various possible system parameter variations, and dealing with uncertainty. Accordingly, the system parameter variations and uncertainty normally considered in the model efficiency tests are presented below, leaving other variables such as X/R ratio, line parameter etc. out of the scope of this study.

5.1. Effect of sampling frequency variation

Sampling frequency sets the number of data samples for one complete cycle of voltage signal for a particular phase. The data points count is 12,000 at 20 kHz sampling rate, while the number is half at 10 kHz frequency. Fig. 12a, b present those signals respectively. The graph shown in Fig. 13a expresses the effect of sampling frequency variation on accuracy. Data samples at 2 kHz, 5 kHz, 10 kHz, 15 kHz, and 20 kHz sampling frequency are taken for the test. The system frequency is 50 kHz constant, and the number of data points in one cycle and for each phase varies from 40 at 2 kHz to a maximum of 400 at 20 kHz sampling rate. Rapid change in accuracy is observed from 2 kHz to 15 kHz frequency. However, the accuracy does not seem to vary much from 15 kHz to 20 kHz frequency. However, the overall deflection in accuracy was 3.18% which proves the consistency in performance under varying sampling rates.

5.2. Varying fault resistance condition

The FIE accuracy could face significant disturbances due to changing fault resistance. The amplitude of the faulty signal is directly affected by the variation in fault resistance, and the images in Fig. 12c,d agree in this context. The first one is captured at a fault resistance of 0.2 Ω, while the other is at 30 Ω. The fact is, if the distance between two value of fault resistance taken is too small, then the amplitude variation is not clearly noticeable by the human eye.

The effect in classification performance due to variation in fault resistance is shown in Fig. 13b. The accuracy at fault resistance 0.1–0.5 Ω seems to vary by 0.03% maximum. However, there is an abrupt change with higher values of fault resistance, as can be seen in the figure.

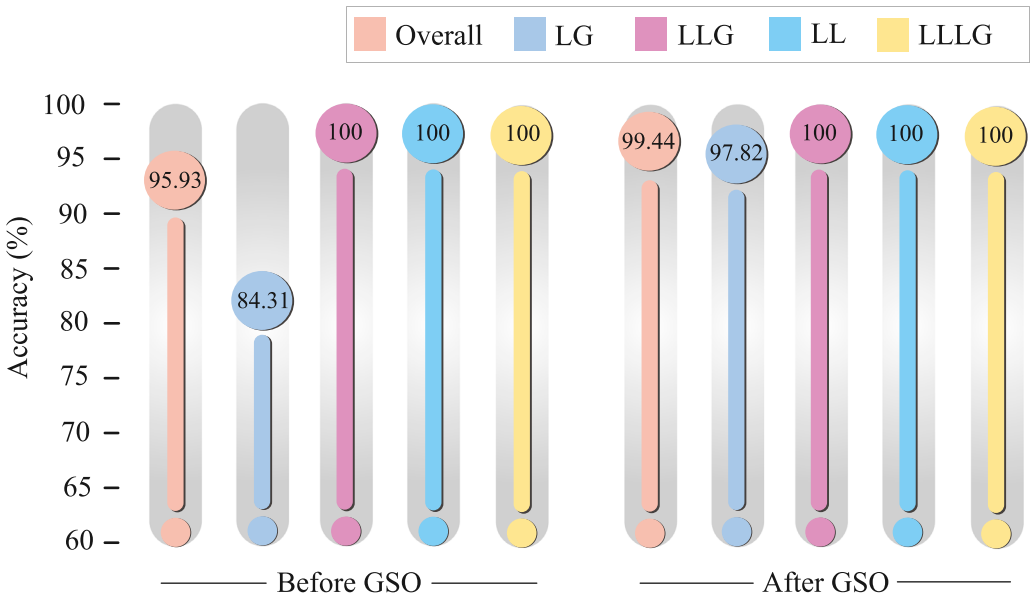


Fig. 11. Overall accuracy for LL, LLG, LL, and LLLG fault classes.

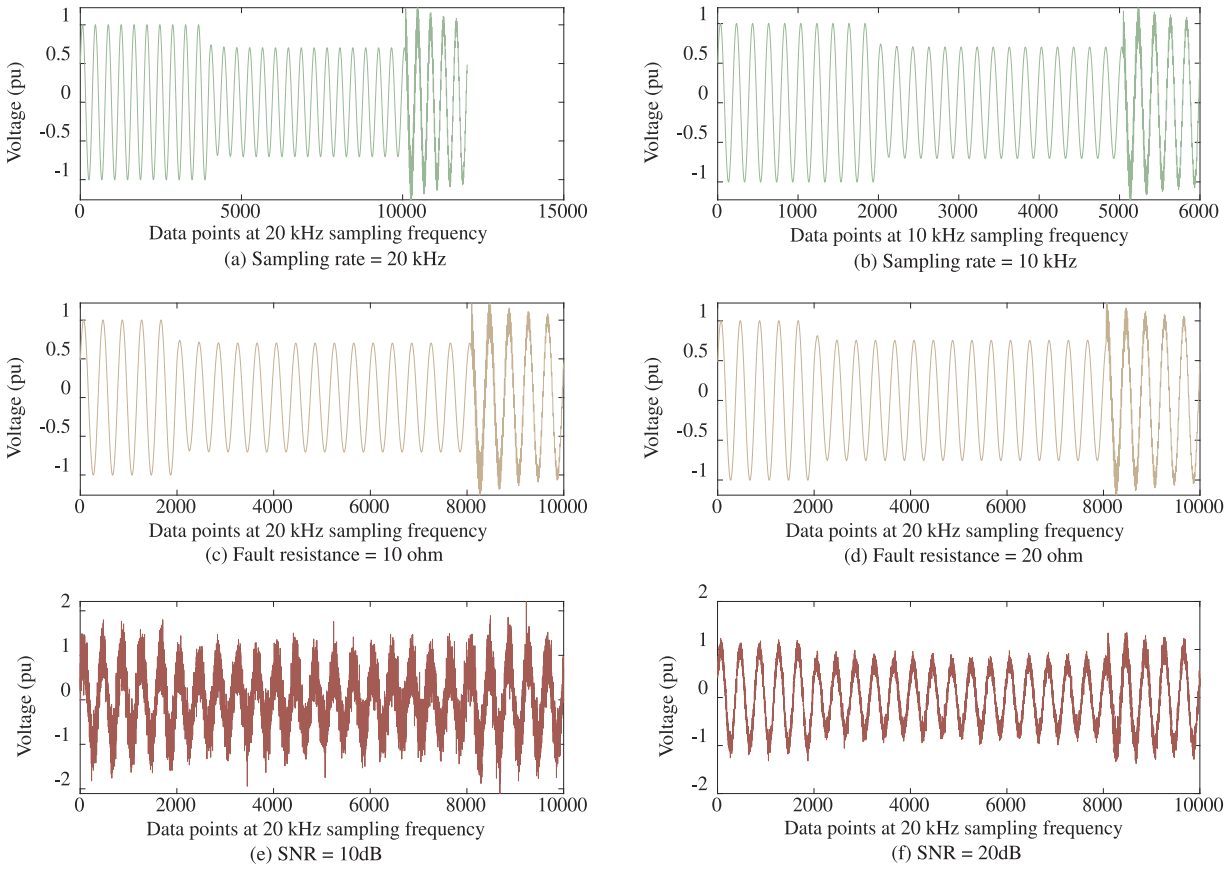
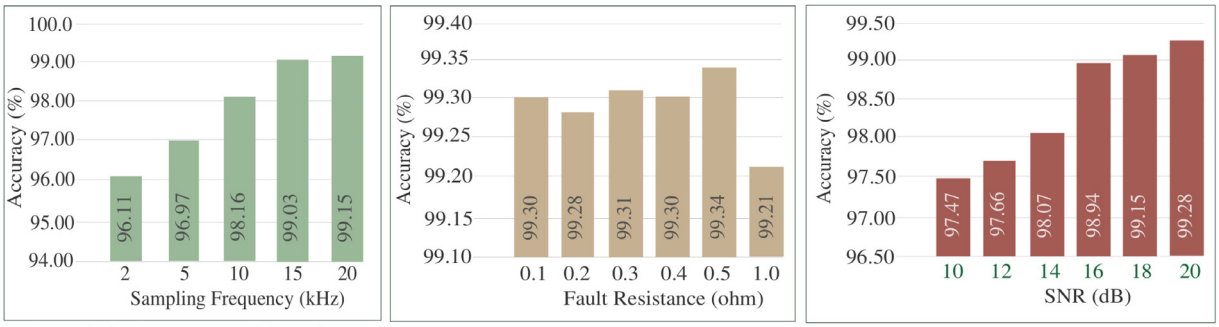


Fig. 12. Exported signals at different system conditions.

5.3. Noisy signal condition

The graph in Fig. 13c presents the variation in accuracy at five different SNR levels. The SNR basically denotes the ratio of the amplitude of a particular signal to the ratio of the amplitude of noise. The value of SNR is small when the amplitude of the noise is more, and vice-versa. The faulty signal in Fig. 12e is generated at 10 dB SNR and Fig. 12f at 20 dB SNR . The noise is maximum at 10 dB and classification accuracy is 97.47%, which is minimum in this case compared to other SNR levels. The maximum accuracy, 99.28%, is achieved at 20 dB. The accuracy seems to increase faster, from SNR 10 dB to 16 dB. After that, it seems to stabilize around 98.94% to 99.28%.



(a) Variation in accuracy with varying sampling frequency

(b) Variation in accuracy with varying fault resistance

(c) Variation in accuracy with varying SNR

Fig. 13. Effect on accuracy due to the system condition changes.

Table 4
Change in Accuracy with Varying Impedance.

Selected source	Impedance variation	Obtained accuracy
Source-1	(+) 50%	99.12%
	(-) 50%	99.15%
Source-2	(+) 50%	99.16%
	(-) 50%	99.19%
Source-3	(+) 50%	98.99%
	(-) 50%	99.05%

5.4. Varying impedance condition

The source impedance is another parameter that affects the accuracy of the classifier. In a practical scenario, there is a good chance that the source impedance will not always be constant. Therefore, an FIE scheme needs to be able to handle those changes and perform consistently. The power system model developed in simulink provides the convenience to vary the impedance if necessary. Table 4 shows three observations, at the first case, the impedance of source-2 is kept constant and source-1 is increased by 50%, and after that, it is decreased by 50%. The obtained accuracy depicts that the accuracy of our model is affected to a minimum extent, even though the variation in impedance is significant. The same goes for source-2; only a 0.03% change in accuracy is observed. When both sources go through the change, the system experiences only 0.06% deflection in accuracy.

5.5. Comparative analysis

A comparison with previously developed ML-based FIE method of the same background as well as state-of the art methods is presented in Table 5. It depicts the effectiveness of the proposed scheme over the mentioned methods. The SVM is used in Ray and Mishra (2016) to classify the fault types that generally occur in power transmission lines. Particle swarm optimization (PSO) algorithm is utilized to find most effective hyper parameters and obtain a better performance. Its ability to handle noise and system parameter variation is also presented in the table. Compared to the deep belief network (DBN) based method proposed in Zhang et al. (2018), the utilization of the proposed solution can provide better result. The implementation of convolutional neuro fuzzy (CNF) (Eboule et al., 2022) and deep reinforcement learning (DRL) (Teimourzadeh et al., 2021) are proposed to solve power system fault analyzing problems. Although, the CNF and DRL provides better result than the previously discussed DBN based method, the proposed approach is still superior to these. The reason for such statement is its ability to deal with the changes in signal and system parameters. Ref. Shadi et al. (2022) shows the use of Rate of change of frequency (ROCOF) signal to analyze disturbances in power system and authors in Rafique et al. (2021) implements LSTM for FIE. For the ROCOF based method, the accuracy is decent but observations under system uncertainties are not present. The LSTM based method deals with most of parameter variation but not all and the proposed method still exceeds this method in terms of accuracy and system implementation simplicity.

In recent times there are more powerful algorithms, for instance, Auto tuning ANN (Ferreira et al., 2020), or self attention cnn (Fahim et al., 2020) or the gradient similarity visualization based CNN (Han et al., 2021) gives comparatively better results than the earlier mentioned model. However, directly comparison of performance of these models with the proposed one is not logical as the paradigms for comparison are not same. These models have been developed for solving specific sets of problems and hence response to the problem is very specific. Although, it is quite evident from Table 5 that the proposed models excels the previously developed model which shares the same dimensions of application.

5.6. Computational effectiveness analysis

One of the highlighted focus of the proposed model is its ability to use less time and power in computation. Keeping that in mind, an analysis is presented in Table 6 which shows the computational effectiveness of our model in terms of training time, response time, accuracy, memory and CPU usage over ANN, CNN, and DBN. The most basic architecture of ANN presented in Abdullah (2018) and CNN, found in Liu et al. (2020), is used for this experiment and the outcomes are there in the table. Although the ANN takes less time in training than the proposed model, in reality more complex structure of ANN is used which will increase the number of trainable parameters. The time complexity (TC) for neural networks is formulated as $O(nxtx(ij + jk + kl))$ where i, j, k are the nodes, t is the size of training sample and n is the number of epochs. TC can also be regarded as the ratio of accuracy to the elapsed training time (Lee and Chen, 2020). It gives us a TC value of 1.69. The value of the time complexity will be higher when the accuracy will be higher and training time will be smaller. As for the proposed method, the TC value obtained is 1.74. The experimental architecture of the DBN is a 4 layer NN (Zhang et al., 2018) which provided 96.62% accuracy and took 30 ms (ms) to response.

In order to use CNN for classification, the architecture implemented is presented in Liu et al. (2020). The numerical data needed to be converted into image format using imaging techniques such as gramian angular field or the markov transition field. It takes more time for the dataset to be prepared and the

Table 5
Comparison between the proposed scheme and the methods developed prior to it.

Parameters	SVM + PSO (Ray and Mishra, 2016)	CNF (Eboule et al., 2022)	DRL (Teimourzadeh et al., 2021)	DBN (Zhang et al., 2018)	RNN (Shadi et al., 2022)	LSTM (Rafique et al., 2021)	The proposed scheme
Signal Type	Current	Current & Voltage	Current & Voltage	Voltage	ROCOF	Current or voltage	Voltage
Overall Accuracy	99.21	98.41	98.04	96.70	99.2	99.45	99.46
Noise Tolerant	No	No	No	No	No	Yes	Yes
Varying Fault Resistance Tolerant	Yes	Yes	No	No	No	Yes	Yes
Varying Sampling Frequency Tolerant	Yes	No	No	No	No	No	Yes
Varying Source Impedance Tolerant	No	No	No	No	No	Yes	Yes

Table 6
Comparison of computation effectiveness with existing models.

Dimensions	Proposed method	ANN (Abdullah, 2018)	CNN (Liu et al., 2020)	DBN (Zhang et al., 2018)
Layers	2	3	6	4
Training time (s)	57.12	56.26	320.56	270.42
Response time (ms)	1.99	43.83	36.42	30
Accuracy (%)	99.46	95.15	98.15	96.62
Memory usage (MB)	260.19	350.27	372.24	365.2
CPU usage (%)	12	22	51	26
Time complexity $= \frac{Accuracy}{TrainingTime}$ (Lee and Chen, 2020)	1.74	1.69	0.31	0.36

model to be trained for classification. Some other deep learning algorithms used in this field are the faster R-CNN and LSTM. These algorithms uses more complex architecture and results in costly computation and dataset prepared in a specific manner to be used. The response time of the proposed model is really impressive which is as low as 1.99 ms, however, the determined TC value is 0.31. To summarize, in all the dimensions, for all the methods discuses in Table 6 the proposed method represent itself as the computationally effective approach.

5.7. Proposed model validation on IEC microgrid system

The designed scheme is adaptable to the change of power system model scale variation. A test is performed on the International Electrotechnical Commission (IEC) standard microgrid system model to verify its adaptability. The specification of the system parameters and description of the model designing is described (Rahman Fahim et al., 2020). The selected mode of operation is islanded radial mode and line 1–3 provides highest accuracy that is 99.34%. The classification accuracy obtained for other line are respectively, line 1–2 - 99.24%, line 3–4 - 99.31%, line 3–5 - 99.19%, line 5–6 - 99.27%. Although, the focus of this paper is to propose an optimized framework that is computationally efficient designed for a prototype power system model, its performance on the IEC system is satisfactory. The confusion matrix on Fig. 14 shows details of classification for line 1–3.

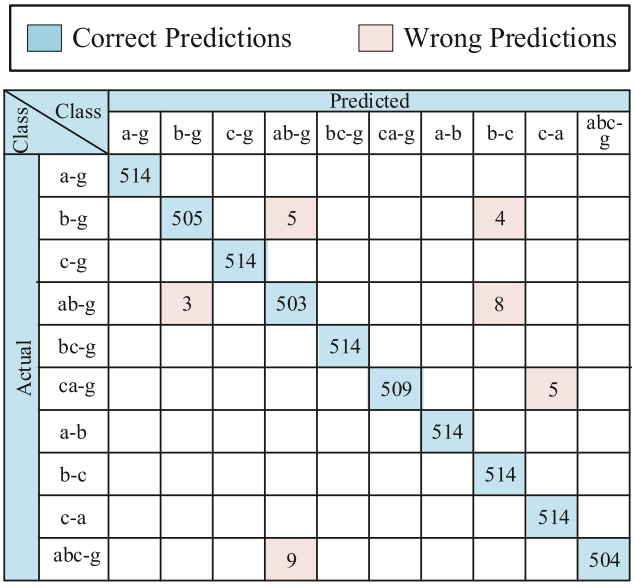


Fig. 14. Confusion matrix for line 1–3 obtained after applying the proposed scheme on the IEC system model.

6. Conclusions

An intelligent framework for identifying and estimating TL faults is presented in this paper that puts more focus on fault analyzing ability enhancement with less computational complexity. The design of the FIE includes several layers of proficiency enhancement techniques that provide an excellent accuracy of 99.46% and deal with the uncertainties that could affect its performance. It can effectively train with a smaller dataset, less computation power while providing higher response speed and ability to handle variations in dataset. The obtained results using the proposed ML model provides a clearer view of its behavior and the ability to maintain its performance under different challenges such as sampling frequency variation, fault resistance changes, and source impedance variation. In all the cases, the accuracy does not seem to fall below 96%. Additionally, a comparative analysis with some of the recently developed methods is included to guarantee the competitive performance of the proposed ML solution. A study about the computational effectiveness of the proposed ML model over the recent DL based methods is done to show the proposed method's supremacy in terms of

training time, response time and computation, and time complexity. Also, the scalability of the proposed model is tested under the experiment of IEC standard microgrid system. The results show that the proposed method is capable enough to effectively classify the faults with more than 99% accuracy for microgrid system. The outcomes mentioned above from this research depicts the importance of the proposed ML solution. Although this research covers most of the disturbances an FIE generally faces, the effect of fault angle variation and voltage–current inversion is kept out of the scope of this paper. The system can also be studied for non-linear load condition and effect of load variation. Furthermore, the microgrid system is tested for the islanded mode only. The other mode of operation and observations using system parameter variation could be a scope for future research.

CRediT authorship contribution statement

Md. Sihab Uddin: Conception and design of study, Acquisition of data, Analysis and/or interpretation of data, Writing – original draft. **Md. Zahid Hossain:** Conception and design of study, Acquisition of data, Writing – original draft. **Shahriar Rahman Fahim:** Conception and design of study, Acquisition of data, Writing – original draft. **Subrata K. Sarker:** Conception and design of study, Analysis and/or interpretation of data, Writing – original draft, Writing – review & editing. **Erphan Ahmmad Bhuiyan:** Analysis and/or interpretation of data, Writing – original draft. **S.M. Muyeen:** Analysis and/or interpretation of data, Writing – review & editing. **Sajal K. Das:** Analysis and/or interpretation of data, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Acknowledgment

The publication of this article was funded by Qatar National Library. All authors approved the version of the manuscript to be published.

References

Abdollahi, A., Seyedtabaii, S., 2010. Comparison of fourier amp; wavelet transform methods for transmission line fault classification. In: 2010 4th International Power Engineering and Optimization Conference. PEOCO, pp. 579–584. http://dx.doi.org/10.1109/PEOCO.2010.5559232.

Abdullah, A., 2018. Ultrafast transmission line fault detection using a DWT-based ANN. IEEE Trans. Ind. Appl. 54 (2), 1182–1193. http://dx.doi.org/10.1109/TIA.2017.2774202.

Ahmadimanesh, A., Shahrtash, S.M., 2013. Transient-based fault-location method for multiterminal lines employing S-transform. IEEE Trans. Power Deliv. 28 (3), 1373–1380.

Bekerman, W., Srivastava, M., 2021. Determining decomposition levels for wavelet denoising using sparsity plot. IEEE Access 9, 110582–110591.

Bhowmik, P., Purkait, P., Bhattacharya, K., 2009. A novel wavelet transform aided neural network based transmission line fault analysis method. Int. J. Electr. Power Energy Syst. 31 (5), 213–219.

Chen, Y.Q., Fink, O., Sansavini, G., 2018. Combined fault location and classification for power transmission lines fault diagnosis with integrated feature extraction. IEEE Trans. Ind. Electron. 65 (1), 561–569. http://dx.doi.org/10.1109/TIE.2017.2721922.

Chen, K., Hu, J., He, J., 2016a. Detection and classification of transmission line faults based on unsupervised feature learning and convolutional sparse autoencoder. IEEE Trans. Smart Grid 9 (3), 1748–1758.

Chen, K., Huang, C., He, J., 2016b. Fault detection, classification and location for transmission lines and distribution systems: a review on the methods. High Voltage 1 (1), 25–33.

Costa, F., Souza, B., Brito, N., 2012. Real-time classification of transmission line faults based on maximal overlap discrete wavelet transform. In: PES T&D 2012. IEEE, pp. 1–8.

Dash, P.K., Das, S., Moirangthem, J., 2015. Distance protection of shunt compensated transmission line using a sparse S-transform. IET Gener. Transm. Distrib. 9 (12), 1264–1274.

Dash, P., Samantaray, S., Panda, G., 2006. Fault classification and section identification of an advanced series-compensated transmission line using support vector machine. IEEE Trans. Power Deliv. 22 (1), 67–73.

Dash, P., Samantaray, S., Panda, G., Panigrahi, B., 2007. Time–frequency transform approach for protection of parallel transmission lines. IET Gener. Transm. Distrib. 1 (1), 30–38.

Duan, K., Keerthi, S.S., Poo, A.N., 2003. Evaluation of simple performance measures for tuning SVM hyperparameters. Neurocomputing 51, 41–59.

Dutta, P., Esmaeilian, A., Kezunovic, M., 2014. Transmission-line fault analysis using synchronized sampling. IEEE Trans. Power Deliv. 29 (2), 942–950.

Eboule, P.S.P., Pretorius, J.H.C., Mbuli, N., 2022. Machine learning techniques applied on a nine-phase transmission line for faults classification and location. Energy Rep. 8, 801–810.

Fahim, S.R., Sarker, S.K., Muyeen, S., Das, S.K., Kamwa, I., 2021. A deep learning based intelligent approach in detection and classification of transmission line faults. Int. J. Electr. Power Energy Syst. 133, 107102.

Fahim, S.R., Sarker, Y., Sarker, S.K., Sheikh, M.R.I., Das, S.K., 2020. Self attention convolutional neural network with time series imaging based feature extraction for transmission line fault detection and classification. Electr. Power Syst. Res. 187, 106437.

Fathabadi, H., 2016. Novel filter based ANN approach for short-circuit faults detection, classification and location in power transmission lines. Int. J. Electr. Power Energy Syst. 74, 374–383.

Ferreira, V.H., Zanghi, R., Fortes, M.Z., Gomes Jr., S., da Silva, A.P.A., 2020. Probabilistic transmission line fault diagnosis using autonomous neural models. Electr. Power Syst. Res. 185, 106360.

Godse, R., Bhat, S., 2020. Mathematical morphology-based feature-extraction technique for detection and classification of faults on power transmission line. IEEE Access 8, 38459–38471.

Han, J., Miao, S., Li, Y., Yang, W., Yin, H., 2021. Faulted-phase classification for transmission lines using gradient similarity visualization and cross-domain adaption-based convolutional neural network. Electr. Power Syst. Res. 191, 106876.

Jain, A., Thoke, A., Patel, R., 2009. Classification of single line to ground faults on double circuit transmission line using ANN. Int. J. Comput. Electr. Eng. 1 (2), 197–203.

Jiang, J.-A., Lin, Y.-H., Yang, J.-Z., Too, T.-M., Liu, C.-W., 2000. An adaptive PMU based fault detection/location technique for transmission lines. II. PMU implementation and performance evaluation. IEEE Trans. Power Deliv. 15 (4), 1136–1146. http://dx.doi.org/10.1109/61.891494.

Jurado, F., Saenz, J.R., 2002. Comparison between discrete STFT and wavelets for the analysis of power quality events. Electr. Power Syst. Res. 62 (3), 183–190.

Koley, E., Kumar, R., Ghosh, S., 2016. Low cost microcontroller based fault detector, classifier, zone identifier and locator for transmission lines using wavelet transform and artificial neural network: A hardware co-simulation approach. Int. J. Electr. Power Energy Syst. 81, 346–360.

Lee, R., Chen, I.-y., 2020. The time complexity analysis of neural network model configurations. In: 2020 International Conference on Mathematics and Computers in Science and Engineering. MACISE, IEEE, pp. 178–183.

Liu, Y., Zhu, Y., Wu, K., 2020. Cnn-based fault phase identification method of double circuit transmission lines. Electr. Power Compon. Syst. 48 (8), 833–843.

Livani, H., Evrenosoglu, C.Y., 2013. A machine learning and wavelet-based fault location method for hybrid transmission lines. IEEE Trans. Smart Grid 5 (1), 51–59.

Lundgren, J., Rönnqvist, M., Värbrand, P., 2001. Linjär Och Ickelinjär Optimering. Studentlitteratur.

Mahamedi, B., Zhu, J.G., 2013. Fault classification and faulted phase selection based on the symmetrical components of reactive power for single-circuit transmission lines. IEEE Trans. Power Deliv. 28 (4), 2326–2332.

Mishra, D.P., Ray, P., 2018. Fault detection, location and classification of a transmission line. Neural Comput. Appl. 30 (5), 1377–1424.

Rafique, F., Fu, L., Mai, R., 2021. End to end machine learning for fault detection and classification in power transmission lines. Electr. Power Syst. Res. 199, 107430.

Rahman Fahim, S., K Sarker, S., Muyeen, S., Sheikh, M., Islam, R., Das, S.K., 2020. Microgrid fault detection and classification: Machine learning based approach, comparison, and reviews. Energies 13 (13), 3460.

Rai, P., Londhe, N.D., Raj, R., 2021. Fault classification in power system distribution network integrated with distributed generators using CNN. Electr. Power Syst. Res. 192, 106914.

Ray, P., Mishra, D.P., 2016. Support vector machine based fault classification and location of a long transmission line. Eng. Sci. Technol. Int. J. 19 (3), 1368–1380.

Raza, A., Benrabah, A., Alquthami, T., Akmal, M., 2020. A review of fault diagnosing methods in power transmission systems. Appl. Sci. 10 (4), 1312.

Reddy, M.J., Mohanta, D.K., 2008. Adaptive-neuro-fuzzy inference system approach for transmission line fault classification and location incorporating effects of power swings. IET Gener. Transm. Distrib. 2 (2), 235–244.

Roy, N., Bhattacharya, K., 2015. Detection, classification, and estimation of fault location on an overhead transmission line using S-transform and neural network. Electr. Power Compon. Syst. 43 (4), 461–472.

Safavian, L., Kinsner, W., Turanli, H., 2005. Aquantitative comparison of different mother wavelets for characterizing transients in power systems. In: Canadian Conference on Electrical and Computer Engineering, 2005. pp. 1461–1464. http://dx.doi.org/10.1109/CCECE.2005.1557253.

Shadi, M.R., Ameli, M.-T., Azad, S., 2022. A real-time hierarchical framework for fault detection, classification, and location in power systems using PMUs data and deep learning. Int. J. Electr. Power Energy Syst. 134, 107399.

Taheri, M.M., Seyedi, H., Nojavan, M., Khoshbouy, M., Ivatloo, B.M., 2018. High-speed decision tree based series-compensated transmission lines protection using differential phase angle of superimposed current. IEEE Trans. Power Deliv. 33 (6), 3130–3138.

Tang, L., Dong, X., Luo, S., Shi, S., Wang, B., 2016. A new differential protection of transmission line based on equivalent travelling wave. IEEE Trans. Power Deliv. 32 (3), 1359–1369.

Teimourzadeh, H., Moradzadeh, A., Shoaran, M., Mohammadi-Ivatloo, B., Razzaghi, R., 2021. High impedance single-phase faults diagnosis in transmission lines via deep reinforcement learning of transfer functions. IEEE Access 9, 15796–15809. http://dx.doi.org/10.1109/ACCESS.2021.3051411.

Wang, X., Zhou, P., Peng, X., Wu, Z., Yuan, H., 2022. Fault location of transmission line based on CNN-LSTM double-ended combined model. Energy Rep. 8, 781–791.

Zhai, Y., Yang, X., Wang, Q., Zhao, Z., Zhao, W., 2021. Hybrid knowledge r-cnn for transmission line multifitting detection. IEEE Trans. Instrum. Meas. 70, 1–12.

Zhang, Y., Mei, W., Dong, G., Gao, J., Wang, P., Deng, J., Pan, H., et al., 2018. A cable fault recognition method based on a deep belief network. Comput. Electr. Eng. 71, 452–464.