Received: 11 December 2023 | Revised: 9 May 2024 | Accepted: 22 May 2024

The Journal of Engineering

DOI: 10.1049/tje2.12412

The Institution of Engineering and Technology WILEY

CASE STUDY

A comparative analysis of different transmission line fault detectors and classifiers during normal conditions and cyber-attacks

Animesh Sarkar Tusher[1] | **M. A. Rahman**[1,2] | **Md. Rashidul Islam**[1] | **M. J. Hossain**[3]

[1]Department of Electrical & Electronic Engineering, Rajshahi University of Engineering & Technology, Rajshahi, Bangladesh

[2]Department of Electronic & Electrical Engineering, Hongik University, Seoul, Republic of Korea

[3]School of Electrical and Data Engineering, University of Technology Sydney, Sydney, New South Wales, Australia

Correspondence
Md. Rashidul Islam, Department of Electrical & Electronic Engineering, Rajshahi University of Engineering & Technology, Rajshahi 6204, Bangladesh.
Email: rashidul@eee.ruet.ac.bd

Abstract

Transmission lines, the core part of the transmission and distribution system in the smart grid, require effective, efficient, and reliable protective measures against faults to avoid severe damage to physical infrastructure and financial losses. Due to their growing popularity, machine learning models are used in fault detection and classification, whose performances can be severely affected by cyber-attacks due to their data dependency, posing a critical concern. Hence, this paper introduces false data injection attacks to address the vulnerability of machine learning-based fault detectors and classifiers. A comparative study of 9 detection models and 6 classification models under normal conditions and during a combination of two models of false data injection attacks is presented to evaluate the severity of cyber-attacks. Experimental results show that highly accurate models in normal conditions are more susceptible to cyber-attacks, with up to 69% and 28% degradations in accuracy for fault detectors and classifiers, respectively. Furthermore, the detection models are found to be more vulnerable to cyber-attacks than the classification models. With no robust detectors and classifiers being found, this work addresses the importance of developing attack-resilient fault detection and classification schemes considering their academic and industrial significance.

1 | INTRODUCTION

The power grid, one of the critical infrastructures, is evolving towards the smart grid with a complete framework for control and automation according to the IEEE Grid Vision 2050, as addressed in [1]. With the integration of information and communication technologies (ICT), the smart grid allows the progressive penetrations of distributed renewable resources while enabling power flows in both directions among energy retailers and consumers through two-way communication for providing efficient and economical services. Although the two-way power flow is advantageous for both ends, they require reliable and effective transmission–distribution systems for transmitting electrical energy across long distances. As a core part of the power transmission network, transmission lines (TLs) play a critical role in energy transmission and are being monitored by various smart sensors and Internet of Things (IoT) based devices.

Due to the significance of TLs, their failures can severely affect the stability and uninterrupted operations of the power grids. Extreme atmospheric and climatic conditions, including lightning, ice, tree interference, bird nesting, breaking, hurricanes, aging, human activities, and a lack of preservation, are some of the notable causes for the failures of TLs, as discussed in [2]. Also, faults can be caused by other factors like damaged insulation, faulty equipment, or outside interference. These failures of TLs can cause voltage fluctuations, power supply interruptions, blackouts, damage to physical infrastructure, and financial losses. Hence, the detection and prompt identification of TL faults are crucial to preserve the reliability and integrity of power transmission networks. Several research works [3, 4] have discussed efficient fault detection systems and strategies for precise diagnosis and localization, enabling quick restoration while minimizing the operation of the power grid. Among different approaches, machine learning algorithms have found popularity in this application [5, 6]. These data-dependent methods

are greatly facilitated by data collected from various monitoring units through two-way communications using ICT systems.

Although the cyberinfrastructure offers various notable advantages in monitoring and control, it comes with a wide surface of vulnerabilities to security lapses and illegal access by cybercriminals. Cyber-attacks can take advantage of weaknesses in data management, control, and communication networks, posing severe risks to the power grid's dependability, availability, and resilience. Due to cyber-attacks on the smart grid, power transmission can be disrupted, and physical infrastructure can be compromised with the possibility of massive outages. For example, a significant outage lasting several hours affected Ukraine in December 2015, where three main distribution firms and over 225,000 consumers were affected due to an attack on the supervisory control and data acquisition (SCADA) system, resulting in manual reset of the circuit breakers [1]. This event has emphasized the importance of strong cyber security procedures, leading to the development of various protective measures, as addressed in [7–10].

Similar to other areas of the smart grid, transmission and distribution systems are also vulnerable to cyber-attacks. Phasor measurement units (PMU), responsible for collecting data at different points of transmission systems, are susceptible to cyber-attacks, as discussed in [11]. As addressed in [12], corrupting time stamps of data collected by PMUs can affect the conventional fault localization method. Also, machine learning (ML) algorithms used in smart grids are susceptible to various cyber-attacks [13], especially the data-driven nature of ML algorithms makes them susceptible to cyber-attacks. In other sectors, examples of adversaries manipulating data during real-time operations to mislead models have been discussed in [14, 15]. Although ML-based schemes have become important in fault detection and classification in recent years, their vulnerabilities to cyber-attacks are yet to be addressed. Hence, this study aims to address this research gap by thoroughly exploring the performance of various ML models used in fault detection and classification during normal and cyber-attack scenarios. The contributions of this work are listed as follows.

1. Introducing false data injection (FDI) based cyber-attacks in transmission line fault detection and classification using a mixed attack template,
2. Investigating the impacts of FDI attacks on ML-based fault detectors and classifiers,
3. Comparing performances of fault detection models under normal cases and cyber-attacks through a comparative study,
4. Comparing performances of fault classifiers under normal cases and cyber-attacks through a comparative study.

The remainder of the paper is laid out as follows. A review of the existing research works is presented in a summarized manner in Section 2. Then, the ML-based fault detection and classification models are discussed with performance evaluation metrics in Section 3. Also, Section 3 presents cyber-attack templates used in this work. After that, the experimental analysis of detectors and classifier models in different scenarios is described in Section 4, including the necessary experimental setup. Lastly, this work is concluded in Section 5.

2 | LITERATURE REVIEW

Traditional failures in TLs are frequently brought on by a number of variables, including weather and atmospheric conditions, lightning strikes, tree interference, insulation problems, defective equipment, and external disruptions from human activity or other external sources, as mentioned in Section 1. Artificial intelligence (AI) has shown the potential to develop protective measures against these TL flaws by providing efficient methods for fault detection, classification, and localization. For example, discrete Fourier transform (DFT), naive Bayes classifier (NBC), and NBC integrated with discrete wavelet transform (DWT) have shown high accuracy with fast response in fault detection and classification, as discussed in [16, 17]. As data can be noisy, considering their collection by physical infrastructure and their transmission through communication channels, fast discrete orthonormal s-transform (FDOST) has been used by combining with different methods for fast, accurate, and robust fault classification [18–20]. Also, due to fault localization being supportive of improving the efficiency and effectiveness of protective measures, several works have addressed this in addition to detection and classification. In [21], a decision tree classifier (DTC) combined with traveling wave analysis has been utilized for fault identification and localization in multi-terminal transmission lines (MTTLs) systems. In other works [22, 23], random forest (RF) has been applied to classify and locate TL faults with high accuracy in the presence of noisy data.

Similar to ML-based methods, neural networks have been found useful for such applications. For instance, artificial neural network (ANN), feedforward neural network (FNN), convolutional neural network (CNN), and deep neural network (DNN) have been used for fault detection and classification of TLs in [24–28]. The applications of ANN with wavelet-oriented methods have been discussed in [29–31] to achieve low-cost, fast, and reliable operation in real-time for fault detection, classification, and localization. Also, with the objectives of fault detection, classification, and localization, a CNN-based protective system has been presented in [32], demonstrating robust performance across diverse fault and power swing conditions during a real-time simulation on an OPAL-RT digital simulator. In another work [33], a CNN and long short-term memory (LSTM) based hybrid model has been used for improving performances in high impedance faults. The fault diagnosis efficiency in diverse length transmission lines has been improved by using a transfer learning framework based on a pre-trained LeNet-5 CNN, as addressed in [2]. Furthermore, YOLOv5 deep learning algorithm has been employed in [34] to improve TLs fault identification accuracy in complicated backgrounds. However, although the applications of ML approaches in data-driven power grids are susceptible to cyber-attacks [13, 35], all the above-mentioned research works are limited to normal cases, raising concerns about their robustness during cyber-attacks. Hence, the effectiveness of ML-based fault detection and

classification during cyber-attacks is yet to be addressed despite the transmission system being vulnerable to cyber-attacks.

As the transmission system is responsible for the efficient, reliable, and economical transmission of electrical energy, the attackers may intend to hamper its normal operations. Hence, the vulnerability of the transmission system is discussed from a broad perspective with a component interdependency graph-based attack strategy in [36]. The impacts on mathematical approach-based fault detection and localization are analysed in the presence of corrupted time stamps of the measured values collected from PMUs in [12]. Also, the attacks can be intended to affect the proper functioning of the protective units in the transmission system. The possibility of cyber-attacks to cause malfunction of fault current limiter is addressed with countermeasures in [37]. In [38], an approach for improving the robustness of line current differential relays is discussed with the ability to identify actual faults in the presence of attacks for tripping lines appropriately and resiliently. In another work [39], to enable resilient PMU data-based protection applications, a new scheme for a wide area measurement system is discussed, considering the severity of cyber-attacks. Although these studies are notable for addressing the vulnerability of transmission systems while presenting countermeasures, they have yet to analyse the performances of fault detectors and classifiers during cyber-attacks. Hence, this work is intended to introduce cyber-attacks for ML-based fault detectors as well as classifiers and to analyse the resiliency of the ML-based approaches.

3 | METHODOLOGIES

One of the critical activities in monitoring and control of the power grid is to monitor faults and initiate protective protocols upon successful fault identification. An overview of a simplified ML-based fault monitoring unit for transmission lines is presented in Figure 1. As illustrated in Figure 1, measured data (e.g. voltages, currents, and others) are transmitted using cyber-infrastructure, which is used as input for fault detection. Later, the fault classification is performed using those data in case of a positive response from the fault detectors. For the proper functioning of the fault monitoring unit, the employment of a classification unit along with the detection unit is crucial. While the protective protocol might be activated upon any fault detection, the fault classification module can help to identify the specific type of fault. The fault type identification allows for more targeted and appropriate protection measures to be implemented, resulting in damage, downtime, and cost minimization while improving the system efficiency, as addressed in [40, 41].

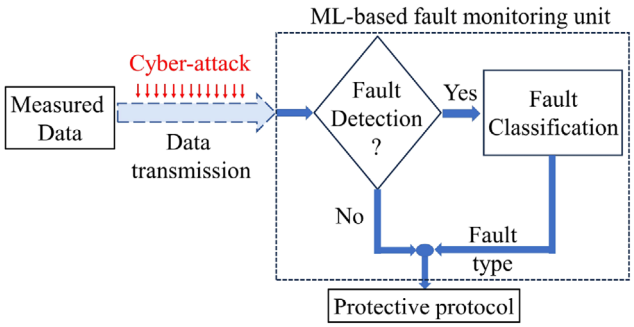

FIGURE 1 An overview of an ML-based fault monitoring unit and its vulnerability.

The protective protocols are activated depending on the response of the fault monitoring unit. As illustrated in Figure 1, the proper functioning of the fault monitoring unit and the implementation of the protection protocol depends on the measured data shared using ICT. During the data transmission process, there is a possibility of cyber-attacks corrupting the transmitted data, as illustrated in Figure 1. Hence, this area is identified as the vulnerability of the fault monitoring unit in this work to perform a comparative analysis of different fault detectors and classifiers during cyber-attacks. In this section, a summarized discussion of the fault detection and classification models is provided while presenting attack templates for FDI attacks and performance evaluator metrics.

3.1 | Fault detectors and classifiers models

Fault detectors and classifiers perform classification tasks in general, where the former identifies the presence of faults, and the latter determines the types of the identified faults. Different ML models have been used in this work as fault detectors and classifiers. As the objective of this work is to evaluate the impacts of cyber-attacks on fault detection and classification systems, few models with good performances have been utilized. For fault detection, ANN [25], adaptive boosting (AdaBoost) [42], k-nearest neighbours (k-NN) [43], DTC [43], RF [43], extra trees classifier (ETsC) [44], bagging classifier (BC) [45], extreme gradient boosting (XGBoost) [46], and light gradient-boosting machine (LGBM) [47] have been used. On the other hand, quadratic discriminant analysis classifier (QDAC) [48], DTC [43], RF [43], ETsC [44], BC [45], and XGBoost [46] models have been utilized for fault classifications. Among these models, ANN is developed for this comparative study, whereas the other models are based on the existing literature. A brief description of these models is provided in this section.

3.1.1 | ANN

A modified architecture for ANN-based fault detection is presented in this work for comparative analysis with different ML-based approaches during normal conditions and cyber-attacks. Due to the notable performance of back-propagation networks with different combinations of hidden layers and varying numbers of neurons during fault detection in [25], an ANN featuring a 6-16-10-5-1 structure is implemented. As shown in Figure 2, the modified ANN consists of three hidden layers with 16, 10, and 5 neurons in each respective layer. The first three layers use rectified linear unit (ReLu) as transfer functions, whereas the fourth layer uses sigmoid. This specific network

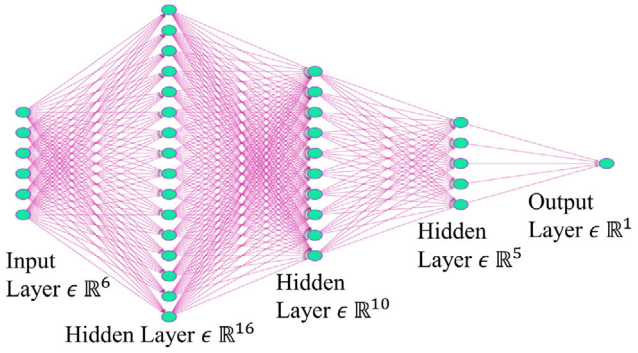

FIGURE 2 The architecture of the modified ANN model.

architecture has been selected after careful evaluation of various configurations, prioritizing high accuracy at normal conditions.

3.1.2 | AdaBoost

AdaBoost is an ensemble learning algorithm known for its simple structure and parameter-free nature while having the resistance to overfitting and the ability to combine multiple weak classifiers into a strong classifier [49]. This algorithm is popular for classification tasks. Due to its sensitivity to outliers and/or noises, it requires well pre-processed data to perform effectively.

3.1.3 | k-NN

The k-NN algorithm is another popular approach for classification that uses proximity to make decisions. It selects the most suitable neighbour based on calculated distances using various methods, such as Euclidean, Cityblock, Minkowski, Chebyshev, and Manhattan distances. Its performance relies on the choice of distance matrix and the number of nearest neighbours (k) [43]. Computational expenses and high memory requirements can make it less desirable for some applications. Also, similar to AdaBoost, its performance depends on the quality of data.

3.1.4 | DTC

DTC is a classification and regression algorithm that mimics a tree structure, where each branch represents a class, and the dataset is recursively split based on features to make predictions [43]. Although this algorithm is prone to overfitting and sensitive to data quality, it has gained popularity due to its versatile nature, less data preparation requirement, and simplicity in understanding and interpretation.

3.1.5 | RF

RF is a supervised learning algorithm that combines multiple decision trees to improve prediction accuracy by utilizing random attribute selection and aggregating the results of individual trees for classification tasks [43]. It offers a reduction in overfitting with comparatively higher accuracy, but it is computationally expensive and has a comparatively long training time. Also, it has low interpretability.

3.1.6 | ETsC

The ETsC is an estimator that fits randomized decision trees on sub-samples of the dataset, utilizing averaging and random splits to improve accuracy and mitigate overfitting [50]. Also, this algorithm offers computational efficiency and feature selection flexibility while having less sensitivity to noise. Although it has some similarities with RF, its construction makes it easily distinguishable.

3.1.7 | BC

The BC is an ensemble method that creates a group of classifiers by training them on random subsets of the training set, combining their outputs through averaging or voting to classify test patterns, and it is particularly effective for unstable learning algorithms like neural networks and decision trees [51]. Although this approach is easy to implement, it is computationally expensive and has less interpretability and flexibility.

3.1.8 | XGBoost

With increased scalability and utilization of resources, XGBoost is a scalable and highly effective tree-boosting system that uses sparsity-aware algorithms, weighted quantile sketches, and optimization approaches to produce state-of-the-art results on machine learning problems [46]. This algorithm is suitable for applications requiring high accuracy and speed, but it needs comparatively large memory space while having computational complexities, lack of transparency, and the possibility of overfitting.

3.1.9 | LGBM

Both the XGBoost classifier and the LGBM classifier use ensemble tree approaches, but the LGBM classifier stands out for having a distributed, extremely effective gradient boosting structure based on decision tree algorithms and for using a leaf-wise method that generates better precision than other boosting techniques [52]. It requires low training time and less memory space while having higher efficiency and less interpretability.

3.1.10 | QDAC

QDAC is a machine learning method that employs a quadratic surface to effectively separate different classes in order to

generate a prediction model [53]. This algorithm is capable of modelling more complex relationships among variables while offering flexibility, but can be computationally expensive with the possibility of overfitting.

3.2 | Models of cyber–attacks

Data-driven methods are vulnerable to data contamination and data unavailability attacks. Data unavailability attacks, such as denial of service (DoS) attacks, can be severe, but they are comparatively easier to notice due to data flow discontinuous for several data points. On the other hand, data contamination attacks are stealthier due to maintaining data flow while corrupting data with small changes. Hence, this work focuses on data contamination attacks to introduce cyber-attacks to ML-based fault detectors and classifiers. Among different data contamination attacks, FDI and adversarial attacks are quite popular, where FDI attacks involve comparatively lower computational complexities and easier implementation processes without requiring any knowledge about the detection and classification models. As addressed in [54], FDI attacks are notable for their severity in different areas of smart grids. Consequently, this attack is often used to introduce cyber-attacks and investigate vulnerabilities even for emerging areas, as addressed in [55]. Hence, FDI attacks are discussed in this work to introduce cyber-attacks on TL fault detectors and classifiers.

For any data-driven models, any FDI attacks corrupt historical data at different data points, which are used as inputs for the models. These maliciously crafted data can misguide the decision-making process for any model, as indicated in [56]. Especially for the trained models, as they are used to predict or classify utilizing clean data, the falsified data causes the models to perform incorrectly. For any deployed TL fault detection and classification systems, FDI attacks often focus on manipulating real-time sensor data during transmission to compromise the integrity of the information, resulting in notable impacts on fault detectors and classifiers for degrading the systems' performance and reducing stability and reliability, as presented later in Section 4.2. In this work, two different attack templates of FDI attacks have been utilized to investigate the susceptibility of ML-based algorithms to cyber-attacks during fault detections and classifications. Among the five different FDI attack templates discussed in [57], scaling and pulse attacks have been considered in this work.

3.2.1 | Scaling attack

Scaling attacks model data contamination through changing variables over a certain period of time by multiplying them using a scaling attack parameter λ_S, as mentioned in [57]. This attack template can be mathematically expressed as follows.

$$X_t^{*F} = (1 + \lambda_S) \times X_t^F, \quad \text{for } t_s < t < t_e \tag{1}$$

where t_s and t_e represent the beginning and end of the cyber-attack, respectively. X_t^F is the original value, which is not corrupted by cyber-attacks. X_t^{*F} is the value tampered with cyber-attacks.

3.2.2 | Pulse attack

Pulse attacks model data contamination through raising or lowering variables at a certain point during the attack period while setting the attack parameter to λ_p [57]. This attack template can be mathematically expressed as follows.

$$X_t^{*F} = (1 + \lambda_p) \times X_t^F, \quad \text{for } t = t_p \tag{2}$$

where t_p, X_t^F, and X_t^{*F} indicate the attack period of one pulse attack, uncorrupted original value, and corrupted value due to cyber-attacks, respectively.

3.3 | Performance evaluator metrics

Model performance is evaluated using two key metrics, which are the confusion matrix and the accuracy of the testing set. The confusion matrix provides a more detailed analysis of various models' classification capabilities, while the testing set accuracy gives a more comprehensive picture of the models' overall prediction ability. Additionally, accuracy deviation is used to point out the impacts of cyber-attacks on each model.

3.3.1 | Confusion matrix

Confusion matrix is a visual tool that provides a clear illustration of classification accuracy by comparing classification results with actual measured values. The confusion matrix, also known as the probability matrix or error matrix, grows into an X-by-X matrix with rows and columns denoting the many classes in issues involving X-classification [49]. In this work, a confusion matrix with a structure made up of four elements has been used, which is shown in Figure 3. The four elements discussed below

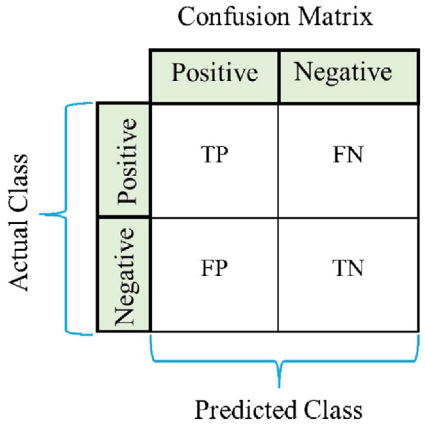

FIGURE 3 Confusion matrix structure.

are true positive (TP), true negative (TN), false positive (FP), and false negative (FN).

1. True positive (TP): A TP identifies when a model correctly classifies a sample as belonging to the positive class. TPs are desirable outcomes, demonstrating the model's ability to accurately detect the condition of interest (e.g. a specific fault type). A higher number of TPs indicates good sensitivity [58].
2. True negative (TN): A TN occurs when a model correctly classifies a sample as belonging to the negative class. TNs are also desirable, indicating the model's ability to correctly identify instances that don't belong to the target category. A higher number of TNs indicates good specificity [58].
3. False positive (FP): An FP, also known as a Type I error, occurs when a model incorrectly classifies a sample as positive when it actually belongs to the negative class. FPs diminish a model's performance, potentially leading to unnecessary actions or false alarms. A high FP rate can erode trust in a model's output [58].
4. False negative (FN): An FN, also known as a Type II error, occurs when a model incorrectly classifies a sample as negative when it actually belongs to the positive class. FNs can have severe consequences, especially in fault detection systems. Missing true faults (high FN rate) could lead to safety issues or equipment damage [58].

3.3.2 | Accuracy

The ratio of correctly predicted labels to the total number of samples in the testing set is used to measure the model's accuracy [49]. In fault detection and classification tasks, accuracy plays a multifaceted role. Firstly, it indicates how reliably a model can distinguish between normal operating conditions and the presence of a fault, which is crucial for timely intervention and preventing equipment damage. Secondly, accuracy reveals the model's ability to correctly identify the specific type of fault, facilitating targeted troubleshooting and effective repair strategies. This metric is used to perform a comparative analysis of different models using the testing set. Also, the impacts of cyber-attacks are indicated by evaluating the deviations in accuracies between normal conditions and cyber-attacks, which can be expressed as in Equation (3).

$$\text{Deviation} = \frac{A - A^*}{A} * 100\% \tag{3}$$

where A and A^* indicate the accuracy in normal conditions and during cyber-attacks, respectively. A high accuracy score suggests a model that effectively detects faults and accurately classifies their types, even under the threat of cyber-attacks.

4 | EXPERIMENTAL SETUP AND ANALYSIS

This section presents a summarized description of the experimental setup and a detailed analysis of experimental observations. The experimental setup includes details about system configuration, data collection with processing, and attack implementation process to support the investigations. The experimental analysis presents observations on detectors and classifiers separately. Also, to evaluate the impacts of cyber-attacks on detectors and classifiers, they have been investigated independently, although they work sequentially.

4.1 | Experimental setup

4.1.1 | System configuration

The experiments have been conducted using Google Colaboratory and a computer with an Intel 5 processor and 8 GB of RAM. Using this system, the outcomes from 30 experiment runs have been averaged to obtain the findings presented in this study.

4.1.2 | Data collection and processing

Open datasets have been used to facilitate TLs fault diagnosis and evaluate their performance under cyber-attacks. The details about the data collection process are discussed in [49], where both the dataset for detection and the dataset for classification have been used for independent investigations on the detectors and classifiers. The detect dataset was utilized for fault detection, containing two distinct labels such as normal and fault. On the other hand, six different labels, including normal state, two-phase short circuit fault (LL), three-phase short circuit fault (LLL), single-phase ground fault (LG), two-phase ground short circuit fault (LLG), and three-phase ground short circuit fault (LLLG), have been used in the class dataset for fault classification purposes [49].

The fault detection and classification models have been trained and tested using electrical measured data as inputs, which are the 3-phase voltages (V_a, V_b, V_c) and currents (I_a, I_b, I_c). To facilitate accurate and effective model training, data normalization is used as a pre-processing step to ensure each feature contributes equally to the model's performance [59]. For this pre-processing step, the package scikit-learn [60] has been used to import the 'MinMaxScaler' class from the 'sklearn.preprocessing' module. The 'MinMaxScaler' can scale and transform the data, bringing it within a standardized range between 0 and 1. For this investigation, the detect dataset and class dataset have been split into 80% for training and 20% for testing for the traditional ML models. On the other hand, for the modified ANN model used for detection, the detect dataset has been divided into 70% for training, 10% for validation, and 20% for testing. As the validation dataset can guide the optimization process [61], the training and validation datasets have been kept separate for the modified ANN model to prevent overfitting. By keeping 20% of the datasets for testing for all models, their performances have been evaluated on equal ground during normal conditions and cyber-attacks.

4.1.3 | Cyber-attack implementations

Attacks have been implemented on the test datasets to assess the resilience of the normally trained models against FDI attacks for both fault detection and classification tasks. This method has been used in order to evaluate how well the models performed in actual attack scenarios. Based on the attack modelling in [57], the values of λ_S in Equation (1) and λ_p in Equation (2) have been 0.2 and 0.1, respectively. As most works related to FDI attacks consider a single attack template, a mixed attack template has been considered in this work by applying both scaling and pulse attacks on 50% of the test data. Compared to a single template-based FDI attack, a mixed attack template can create more diversity in attack patterns for increasing the impacts of attacks while causing complexities in developing defensive measures. For the mixed attack template-based FDI attacks, Equations (1) and (2) have been used for simulating half of the data contamination by scaling attack and another half of the data contamination by pulse attack, respectively. During attack implementation, overlaps of the attacks have been avoided by selecting one input parameter at random in each case to attack. The detailed steps involved in the experimental analysis are outlined in Algorithm 1.

ALGORITHM 1 TLs fault diagnosis under cyber-attacks

Input: Electrical Measurements (Va, Vb, Vc, Ia, Ib, Ic) for Detect and Class Dataset
Output: Fault detection and classification results (normal and cyber-attack condition)
`/* Data Preprocessing */`
Normalize inputs using MinMaxScaler;
`/* Data Splitting */`
For Traditional ML Models: Divide datasets into 80% training, 20% testing;
For ANN Detector Model: Divide Detect dataset into 70% training, 10% validation, 20% testing;
`/* Model Training */`
Train traditional ML fault detector model;
Train traditional ML fault classifier model;
Train ANN detector model (optimize with validation set);
`/* Cyber-attack Implementation */`
Set $\lambda_S = 0.2, \lambda_p = 0.1$;
for *25% of test data instances* **do**
  Randomly select input parameter;
  Apply scaling attack (Eq. 1);
end
for *25% of test data instances* **do**
  Randomly select input parameter;
  Apply pulse attack (Eq. 2);
end
`/* Evaluation */`
Normal Conditions: Evaluate models on normal scenarios;
Cyberattack Conditions: Evaluate models on attacked scenarios;

4.2 | Experimental analysis

During the investigation, the data contamination has been kept moderate using the mixed attack template, as illustrated in Figures 4 and 5. As presented in Figure 4, there are changes in currents and voltages due to data contamination compared to the currents and voltages at normal conditions, including fault cases, but the changes may be barely visible to human eyes. The same phenomena can be observed for the classification datasets, as shown in Figure 5. These voltages and currents without and with data contaminations due to the mixed attack template of FDI attacks have been used for experimental analysis of different fault detectors and classifiers in this section. The experimental analysis presented in this section points to the susceptibility of the considered models to cyber-attacks while indicating the urgency of developing robust models, as discussed below.

4.2.1 | Analysis of fault detectors

As mentioned in Section 3.1, ANN, AdaBoost, k-NN, DTC, RF, ETsC, BC, XGBoost, and LGBM have been used for fault detection. For these models, the training time and the accuracy during normal conditions and cyber-attacks with the deviations in accuracy due to the attacks have been summarized in Table 1. As presented in Table 1, all the models have high accuracy at normal conditions, where the modified ANN, RF, and XGBoost have outperformed others. Although these models have accuracy above 99% at normal conditions, their performances have degraded significantly during cyber-attacks, as shown in Table 1. AdaBoost has been affected the most, and BC is the least affected model based on the percentage of deviations, while the impacts of cyber-attacks on other models are the same. In the case of the training time at the normal scenario, all the models require comparatively less time, except for the modified ANN.

To provide a better presentation of the impacts of cyber-attacks on fault detectors, confusion matrices for normal and attacked scenarios have been used for each model, as illustrated in Figure 6. In the confusion matrix of fault detector models, label 0 and label 1 correspond to the normal condition or no-fault scenario and the fault condition, respectively. As shown in Figure 6, ANN, RF, and XGBoost have the same performance at normal conditions, while others are very close to them. On the other hand, the impacts of cyber-attacks have caused notable performance for them, as mentioned earlier. The AdaBoost has the worst performance with notable false cases, as shown in Figure 6(b). Among other models, k-NN, RF, ETsC, XGBoost, and LGBM have performed similarly during cyber-attacks by predicting all cases as faults irrespective of the actual scenarios. Also, the modified ANN has performed closely to these five models. Although DTC and BC have performed closely and comparatively better than others based on the confusion matrices, BC has been found superior to DTC due to

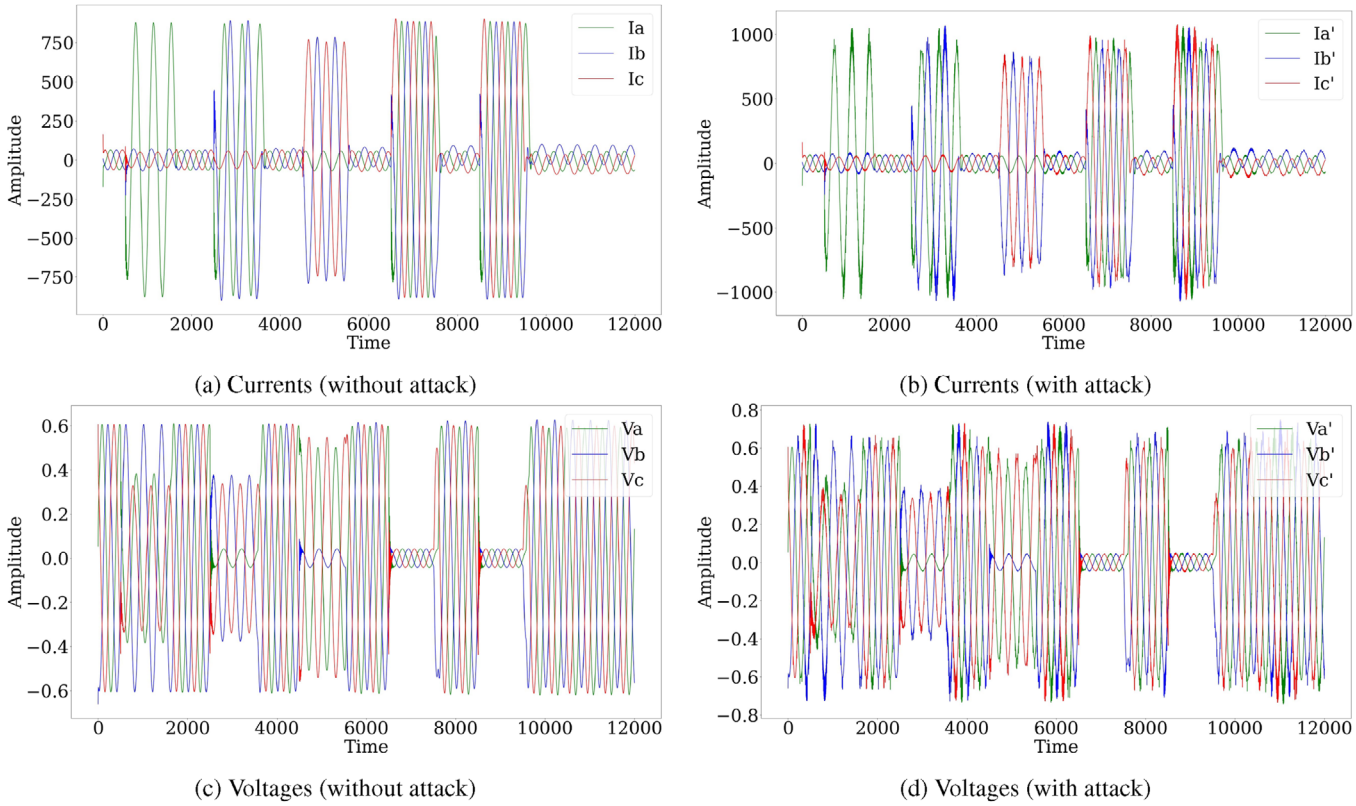

FIGURE 4 Three phase voltages and current for fault detection without and with FDI attacks.

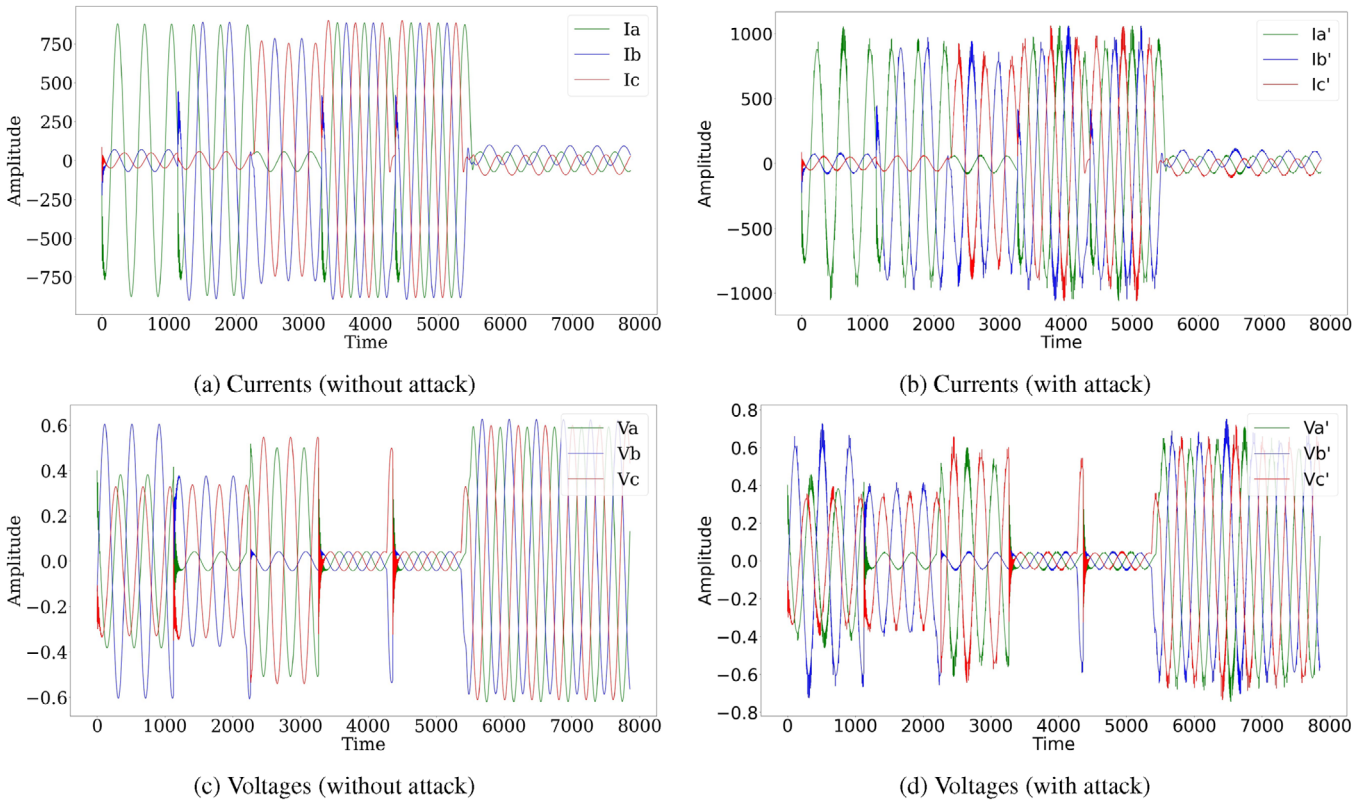

FIGURE 5 Three phase voltages and current for fault classification without and with FDI attacks.

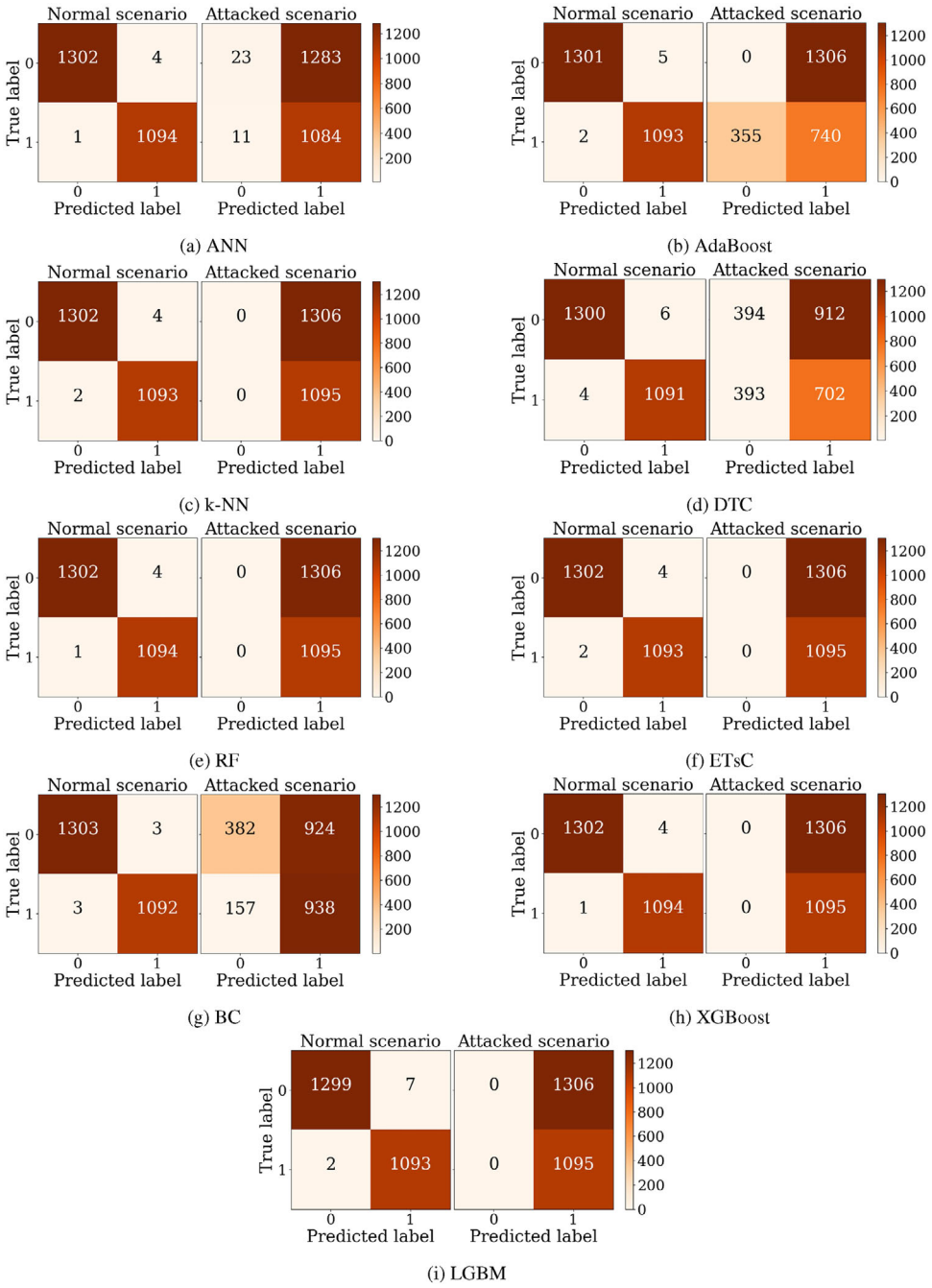

FIGURE 6 Comparison of different fault detectors at normal and attacked scenarios based on confusion matrices.

having more true cases. Based on the observations in Table 1 and Figure 6, BC has comparatively better performances than others considering all metrics.

4.2.2 | Analysis of fault classifiers

For fault classifications, QDAC, DTC, RF, ETsC, BC, and XGBoost have been utilized, as mentioned in Section 3.1. The comparative analysis of these classifiers has been summarized in Table 2 in terms of the training time and the accuracy during normal conditions and cyber-attacks with the deviations in accuracy due to the attacks. Among the considered models, only QDAC has an accuracy above 90% at normal conditions. Although QDAC has outperformed others based on accuracy and training time in normal scenarios, it has been affected the most during cyber-attacks. On the other hand, the other models with comparatively low accuracy at normal conditions have been less affected by the cyber-attacks. The XGBoost model with the lowest accuracy is the least affected model during

TABLE 1 Comparison of different models for fault detections.

Model Name	Accuracy		Deviation (%)	Training time (s)
	Normal	Attack		
ANN	**0.9979**	0.4611	54	127.0366
AdaBoost	0.9971	0.3082	69	0.6957
k-NN	0.9975	0.4561	54	**0.0216**
DTC	0.9958	0.4565	54	0.4416
RF	**0.9979**	0.4561	54	1.5834
ETsC	0.9975	0.4561	54	0.0970
BC	0.9975	**0.5498**	**45**	0.4652
XGBoost	**0.9979**	0.4561	54	0.9443
LGBM	0.9963	0.4561	54	0.1791

Bold font - Best performance

TABLE 2 Comparison of different models for fault classifications.

Model Name	Accuracy		Deviation (%)	Training time (s)
	Normal	Attack		
QDAC	**0.9822**	0.7088	28	**0.0244**
DTC	0.8919	**0.8519**	04	0.1095
RF	0.8875	**0.8519**	04	1.4811
ETsC	0.8907	**0.8519**	04	0.6277
BC	0.8786	**0.8519**	03	0.3800
XGBoost	0.8557	0.8493	**01**	4.2441

Bold font - Best performance

cyber-attacks, as indicated by the percentage of deviation. Also, among these models, only RF and XGBoost have required more than 1 s for training, while the rest have completed training in less than 1 s. These observations in Table 2 indicate the trade-offs between accuracy and robustness. Also, DTC has been found comparatively better than others based on training time, deviation, and accuracy in both scenarios, while ETsC has been found to be competitive.

For a more comprehensive understanding of the performance of the fault classifiers, the confusion matrix has been used for each model to provide insights into the performances during normal conditions and cyber-attacks, as shown in Figure 7. The labels assigned in the confusion matrix of fault classifier models are as follows: 0 for the normal condition or no-fault scenario, 1 for LG fault, 2 for LL fault, 3 for LLG fault, 4 for LLL fault, and 5 for LLLG fault. The high accuracy of QDAC at normal conditions compared to others becomes more visible using a confusion matrix, while highlighting the notable performance degradations during cyber-attacks. On the other hand, DTC, ETsC, RF, XGBoost, and BC have shown almost similar in classifying each type of fault under normal conditions based on their confusion matrices. Also, the changes in performances due to cyber-attacks become barely noticeable for these models. Among these five models, DTC and ETsC have shown comparatively better performances than others in classifying faults during normal conditions, while ETsC and RF have been found to be better in fault classification during cyber-attacks. Based on these observations in Table 2, ETsC has been found better than others in attack classification in all scenarios.

4.2.3 | Summary

For fault detectors, although BC has been found to be comparatively better than others, considering all scenarios none of the fault detection models have performed reasonably well during cyber-attacks. This highlights the vulnerabilities of the fault detector models to cyber-attacks. Hence, the comparative analysis of ML-based fault detectors presented in this work signifies the necessity of developing attack-resilient fault detectors to ensure reliable, effective, and robust performances. Similarly, ETsC has been found to outperform others considering all scenarios, but its low accuracy during normal conditions is concerning. Although the classifier models have performed better than the detector models during cyber-attacks, the trade-offs between accuracy and robustness point to the necessity of developing an attack-resilient fault classification system.

As the highly performed models under normal conditions suffer severely during cyber-attacks, pre-processing based defensive mechanisms such as attack detection can be one possible solution for developing robust fault detection and classification systems for utilizing the existing models. On the other hand, there is some room for improvement from the model perspective. As the models are usually trained with clean data, they are inherently vulnerable to any form of data abnormalities. A comparative analysis of models belonging to the same group can help to understand the changes in inherent robustness with changes in architectures, which can lead to the development of a robust architecture. While defensive security measures help to prevent the impacts of cyber-attacks, a robust architecture development can help in mitigating the impacts of attacks in the absence or failure of defensive measures. Nonetheless, considering the severity of cyber-attacks addressed in this work, improving the resiliency of fault detectors and classifiers by any means is of high importance for smart grids.

5 | CONCLUSIONS

The vulnerability of the protective system for TLs has been explored through a comparative analysis of ML-based fault detection and classification systems. Nine fault detectors and six fault classifiers have been utilized under normal conditions and cyber-attacks, which has indicated the severity of cyber-attacks on their performances. A mixed attack template for an FDI attack has been used by combining scaling and pulse attack models for corrupting 50% of data. The impacts of cyber-attacks on detectors and classifiers have been investigated separately by using different datasets, which has highlighted the susceptibility of detector models to cyber-attacks comparatively more than classifier models. As no model has been found to be reasonably accurate and robust for fault detection and

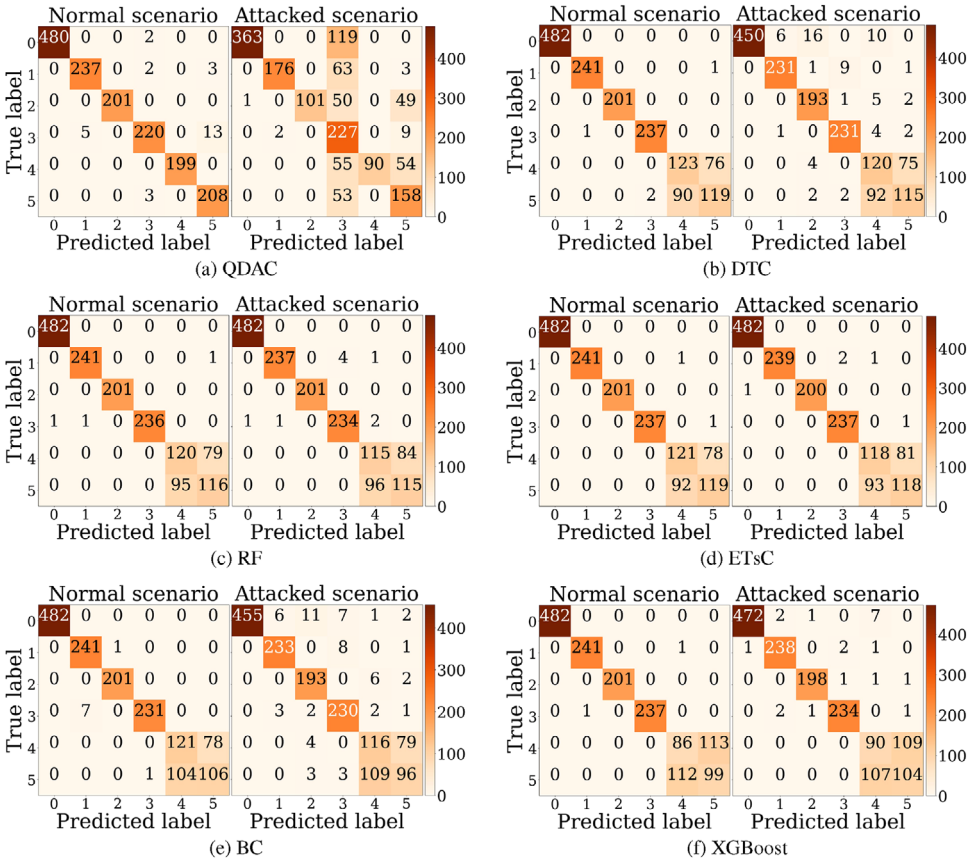

FIGURE 7 Comparison of different fault classifiers at normal and attacked scenarios based on confusion matrices.

classification, the study addresses the theoretical and industrial significance of conducting more research in this area to improve the fault detection and classification system of TLs. In the future, different types of cyber-attacks will be investigated to identify the vulnerability of fault monitoring systems of TLs from a broad perspective, including adversarial examples and DoS attacks. Simultaneously, a comparative analysis of ML-based algorithms can be conducted to identify their limitations under normal and attack scenarios, while finding their compatibility with different defence mechanisms to develop robust fault identification systems. Furthermore, as DL-based models are data-dependent like ML-based models, this work can be extended to DL-based fault detectors and classifiers to find their vulnerabilities and improve their resilience.

AUTHOR CONTRIBUTIONS

Animesh Sarkar Tusher: Conceptualization; data curation; investigation; methodology; resources; writing—original draft. **M. A. Rahman**: Conceptualization; investigation; resources; writing—original draft. **Md. Rashidul Islam**: Conceptualization; supervision; writing—review and editing. **M. J. Hossain**: Supervision; writing—review and editing.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data used in this work are confidential.

ORCID

Animesh Sarkar Tusher https://orcid.org/0009-0006-9318-361X
M. A. Rahman https://orcid.org/0000-0003-0864-4882
Md. Rashidul Islam https://orcid.org/0000-0001-8415-0206
M. J. Hossain https://orcid.org/0000-0001-7602-3581

REFERENCES

1. Musleh, A.S., Chen, G., Dong, Z.Y.: A survey on the detection algorithms for false data injection attacks in smart grids. IEEE Trans. Smart Grid 11(3), 2218–2234 (2020)
2. Shakiba, F.M., Shojaee, M., Azizi, S.M., Zhou, M.: Generalized fault diagnosis method of transmission lines using transfer learning technique. Neurocomputing 500, 556–566 (2022)
3. Stefanidou-Voziki, P., Sapountzoglou, N., Raison, B., Dominguez-Garcia, J.: A review of fault location and classification methods in distribution grids. Electr. Power Syst. Res. 209, 108031 (2022)
4. Islam, M.R., Hasan, M.M., Anower, M.S., Sheikh, M.: Detection and localization of fault in an electrical power transmission line using sub-transient current measurement and wavelength calculation method. Int. J. Electr. Eng. 18(2), 472–479 (2018)
5. Prasad, A., Edward, J.B., Ravi, K.: A review on fault classification methodologies in power transmission systems: part-II. J. Electr. Syst. Inf. Technol. 5(1), 61–67 (2018)

6. Ning, J., Ren, Y., Lin, J., Jiang, C., Zhang, Y., Zhang, Z.: Power system fault diagnosis method based on artificial intelligence and information fusion. Power Grid Technol. 45(08), 2925–2936 (2021)
7. Gunduz, M.Z., Das, R.: Cyber-security on smart grid: threats and potential solutions. Comput. Networks 169, 107094 (2020)
8. Rahman, M.A., Rana, M.S., Pota, H.R.: Mitigation of frequency and voltage disruptions in smart grid during cyber-attack. J. Control, Autom. Electr. Syst. 31, 412–421 (2020)
9. Hussain, S., Fernandez, J.H., Al-Ali, A.K., Shikfa, A.: Vulnerabilities and countermeasures in electrical substations. Int. J. Crit. Infrastruct. Prot. 33, 100406 (2021)
10. Hasan, M.K., Habib, A.A., Shukur, Z., Ibrahim, F., Islam, S., Razzaque, M.A.: Review on cyber-physical and cyber-security system in smart grid: standards, protocols, constraints, and recommendations. J. Netw. Comput. Appl. 209, 103540 (2023)
11. Tharzeen, A., Natarajan, B., Srinivasan, B.: Phasor data correction and transmission system state estimation under spoofing attacks. Electr. Power Syst. Res. 221, 109435 (2023)
12. Zhang, Z., Gong, S., Dimitrovski, A.D., Li, H.: Time synchronization attack in smart grid: impact and analysis. IEEE Trans. Smart Grid 4(1), 87–98 (2013)
13. Chen, Y., Tan, Y., Deka, D.: Is machine learning in power systems vulnerable? In: 2018 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm), pp. 1–6. IEEE, Piscataway, NJ (2018)
14. Biggio, B., Roli, F.: Wild patterns: ten years after the rise of adversarial machine learning. In: Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security. CCS '18, pp. 2154–2156. ACM, New York (2018)
15. Liang, G., Zhao, J., Luo, F., Weller, S.R., Dong, Z.Y.: A review of false data injection attacks against modern power systems. IEEE Trans. Smart Grid 8(4), 1630–1638 (2017)
16. Swetapadma, A., Yadav, A.: Protection of parallel transmission lines including inter-circuit faults using naïve Bayes classifier. Alexandria Eng. J. 55(2), 1411–1419 (2016)
17. Aker, E., Othman, M.L., Veerasamy, V., Aris, I.b., Wahab, N.I.A., Hizam, H.: Fault detection and classification of shunt compensated transmission line using discrete wavelet transform and naïve Bayes classifier. Energies 13(1), 243 (2020)
18. Mishra, P.K., Yadav, A., Pazoki, M.: A novel fault classification scheme for series capacitor compensated transmission line based on bagged tree ensemble classifier. IEEE Access 6, 27373–27382 (2018)
19. Patel, B.: A new fdost entropy based intelligent digital relaying for detection, classification and localization of faults on the hybrid transmission line. Electr. Power Syst. Res. 157, 39–47 (2018)
20. Moravej, Z., Pazoki, M., Khederzadeh, M.: New pattern-recognition method for fault analysis in transmission line with UPFC. IEEE Trans. Power Delivery 30(3), 1231–1242 (2015)
21. Chaitanya, B., Yadav, A.: Decision tree aided travelling wave based fault section identification and location scheme for multi-terminal transmission lines. Measurement 135, 312–322 (2019)
22. Wilches-Bernal, F., Jiménez-Aparicio, M., Reno, M.J.: An algorithm for fast fault location and classification based on mathematical morphology and machine learning. In: 2022 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), pp. 1–5. IEEE, Piscataway, NJ (2022)
23. Fonseca, G.A., Ferreira, D.D., Costa, F.B., Almeida, A.R.: Fault classification in transmission lines using random forest and notch filter. J. Control, Autom. Electr. Syst. 33(2), 598–609 (2022)
24. Ahmed, S., Islam, M.R.: Simplified artificial neural network based fault classification and location for transmission line. In: 2019 5th International Conference on Advances in Electrical Engineering (ICAEE), pp. 485–489. IEEE, Piscataway, NJ (2019)
25. Jamil, M., Sharma, S.K., Singh, R.: Fault detection and classification in electrical power transmission system using artificial neural network. SpringerPlus 4(1), 1–13 (2015)
26. Kumari, S., Mishra, A., Singhal, A., Dahiya, V., Gupta, M., Gawre, S.K.: Fault detection in transmission line using ann. In: 2023 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), pp. 1–5. IEEE, Piscataway, NJ (2023)
27. Rai, P., Londhe, N.D., Raj, R.: Fault classification in power system distribution network integrated with distributed generators using CNN. Electr. Power Syst. Res. 192, 106914 (2021)
28. Ahmed, S.I., Rahman, M.F., Kundu, S., Chowdhury, R.M., Hussain, A.O., Ferdoushi, M.: Deep neural network based fault classification and location detection in power transmission line. In: 2022 12th International Conference on Electrical and Computer Engineering (ICECE), pp. 252–255. IEEE, Piscataway, NJ (2022)
29. Koley, E., Kumar, R., Ghosh, S.: 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 (2016)
30. Chavan, J.N., Kale, A.A., Deore, S.R.: Transmission line fault detection using wavelet transform & ANN approach. In: 2022 IEEE Integrated STEM Education Conference (ISEC), pp. 426–432. IEEE, Piscataway, NJ (2022)
31. Yadav, A., Swetapadma, A.: A single ended directional fault section identifier and fault locator for double circuit transmission lines using combined wavelet and ANN approach. Int. J. Electr. Power Energy Syst. 69, 27–33 (2015)
32. Shukla, S.K., Koley, E., Ghosh, S.: Grey wolf optimization-tuned convolutional neural network for transmission line protection with immunity against symmetrical and asymmetrical power swing. Neural Comput. Appl. 32, 17059–17076 (2020)
33. Moradzadeh, A., Teimourzadeh, H., Mohammadi-Ivatloo, B., Pourhossein, K.: Hybrid cnn-lstm approaches for identification of type and locations of transmission line faults. Int. J. Electr. Power Energy Syst. 135, 107563 (2022)
34. Li, M., Cai, C., Sun, Y.: A method for identifying transmission line faults based on deep learning. In: 2022 China Automation Congress (CAC), pp. 3295–3300. IEEE, Piscataway, NJ (2022)
35. Zhang, Z., Zuo, K., Deng, R., Teng, F., Sun, M.: Cybersecurity analysis of data-driven power system stability assessment. IEEE Internet Things J. 10(17), 15723–15735 (2023)
36. Zhu, Y., Yan, J., Tang, Y., Sun, Y.L., He, H.: Joint substation-transmission line vulnerability assessment against the smart grid. IEEE Trans. Inf. Forensics Secur. 10(5), 1010–1024 (2015)
37. Wei, F., Wan, Z., He, H., Lin, X.: Ultrafast active response strategy against malfunction attack on fault current limiter. IEEE Trans. Smart Grid 11(3), 2722–2733 (2020)
38. Ameli, A., Hooshyar, A., El-Saadany, E.F.: Development of a cyber-resilient line current differential relay. IEEE Trans. Ind. Inf. 15(1), 305–318 (2019)
39. Chawla, A., Agrawal, P., Panigrahi, B.K., Paul, K.: Deep-learning-based data-manipulation attack resilient supervisory backup protection of transmission lines. Neural Comput. Appl. 35(7), 4835–4854 (2023)
40. Potdar, V., Sharif, A., Chang, E.: Wireless sensor networks: a survey. In: 2009 International Conference on Advanced Information Networking and Applications Workshops, pp. 636–641. IEEE, Piscataway, NJ (2009)
41. Wischkaemper, J., Brahma, S.: Machine learning and power system protection [viewpoint]. IEEE Electrif. Mag. 9(1), 108–112 (2021)
42. Freund, Y., Schapire, R.E.: A decision-theoretic generalization of on-line learning and an application to boosting. J. Comput. Syst. Sci. 55(1), 119–139 (1997)
43. Shakiba, F.M., Azizi, S.M., Zhou, M., Abusorrah, A.: Application of machine learning methods in fault detection and classification of power transmission lines: a survey. Artif. Intell. Rev. 56(7), 5799–5836 (2023)
44. Geurts, P., Ernst, D., Wehenkel, L.: Extremely randomized trees. Mach. Learn. 63, 3–42 (2006)
45. Richman, R., Wüthrich, M.V.: Nagging predictors. Risks 8(3), 83 (2020)
46. Chen, T., Guestrin, C.: XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794 (2016)
47. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., et al.: LightGBM: a highly efficient gradient boosting decision tree. In: NIPS'17: Proceedings

of the 31st International Conference on Neural Information Processing Systems, pp. 3149–3157. ACM, New York (2017)
48. Tharwat, A.: Linear vs. quadratic discriminant analysis classifier: a tutorial. Int. J. Appl. Pattern Recognit. 3(2), 145–180 (2016)
49. He, Q., Chen, L., Tang, W., He, G., Tan, S., Shu, Y., et al.: Application and comparative analysis of traditional machine learning and deep learning in transmission line fault classification. In: 2022 IEEE 5th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), Vol. 5, pp. 1715–1719. IEEE, Piscataway, NJ (2022)
50. Abhishek, L.: Optical character recognition using ensemble of SVM, MLP and extra trees classifier. In: 2020 International Conference for Emerging Technology (INCET), pp. 1–4. IEEE, Piscataway, NJ (2020)
51. Tharwat, A., Gaber, T., Awad, Y.M., Dey, N., Hassanien, A.E.: Plants identification using feature fusion technique and bagging classifier. In: The 1st International Conference on Advanced Intelligent System and Informatics (AISI2015), pp. 461–471. Springer, Cham (2016)
52. Nikam, A., Bhandari, S., Mhaske, A., Mantri, S.: Cardiovascular disease prediction using machine learning models. In: 2020 IEEE Pune Section International Conference (PuneCon), pp. 22–27. IEEE, Piscataway, NJ (2020)
53. Chaber, R., Arthur, C.J., Depciuch, J., Łach, K., Raciborska, A., Michalak, E., et al.: Distinguishing Ewing sarcoma and osteomyelitis using FTIR spectroscopy. Sci. Rep. 8(1), 15081 (2018)
54. Unsal, D.B., Ustun, T.S., Hussain, S.S., Onen, A.: Enhancing cybersecurity in smart grids: false data injection and its mitigation. Energies 14(9), 2657 (2021)
55. Tusher, A.S., Rahman, M.A., Islam, M.R., Hossain, M.J.: Adversarial training-based robust lifetime prediction system for power transformers. Electr. Power Syst. Res. 231, 110351 (2024)
56. Biggio, B., Nelson, B., Laskov, P.: Poisoning attacks against support vector machines. arXiv:12066389v3 (2013)
57. Cui, M., Wang, J., Yue, M.: Machine learning-based anomaly detection for load forecasting under cyberattacks. IEEE Trans. Smart Grid 10(5), 5724–5734 (2019)
58. Fawcett, T.: An introduction to ROC analysis. Pattern Recognit. Lett. 27(8), 861–874 (2006)
59. Singh, D., Singh, B.: Investigating the impact of data normalization on classification performance. Appl. Soft Comput. 97, 105524 (2020)
60. Barupal, D.K., Fiehn, O.: Generating the blood exposome database using a comprehensive text mining and database fusion approach. Environ. Health Perspect. 127(9), 097008 (2019)
61. Lago, J., Marcjasz, G., De Schutter, B., Weron, R.: Forecasting day-ahead electricity prices: a review of state-of-the-art algorithms, best practices and an open-access benchmark. Appl. Energy 293, 116983 (2021)

How to cite this article: Tusher, A.S., Rahman, M.A., Islam, M.R., Hossain, M.J.: A comparative analysis of different transmission line fault detectors and classifiers during normal conditions and cyber-attacks. J. Eng. 2024, e12412 (2024).
https://doi.org/10.1049/tje2.12412