International Journal of Research and Technology in Electrical Industry

journal homepage: ijrtei.sbu.ac.ir

Shahid Beheshti University

A Transfer Learning-based Convolutional Neural Network for Event Classification with Small Databases

Mohammad Aryanfar [1], Mostafa Jazaeri [1,*]

[1] Department of Electrical and Computer Engineering, Semnan University, Semnan, Iran

ARTICLE INFO

Article history:
Received: 09 December 2024
Revised: 08 March 2025
Accepted: 09 April 2025

Keywords:
Transfer learning
Convolutional neural networks
Deep learning
Fault classification
Small database

ABSTRACT

Power system reliability hinges on accurate and timely fault classification, yet many real-world scenarios face data scarcity due to logistical and economic constraints. Traditional methods often struggle to maintain performance with limited training samples, creating a critical gap in practical applications. Fault classification in power systems often requires robust models that can be generalized from limited data. Traditional deep learning approaches, while highly effective, usually need large datasets to achieve acceptable performance. In this paper, we propose a novel convolutional neural networks (CNN) framework for fault classification tasks using small-scale databases. This is novel because it leverages transfer learning to adapt a pre-trained model in deep learning to the target domain of fault classification. Compared with other methods, our approach minimizes the dependency on large datasets besides achieving high accuracy and generalizability. Extensive experiments demonstrate that the proposed approach achieves state-of-the-art performance, validating its efficacy for scenarios with limited data availability. This research provides an essential step in applying deep learning to the fault classification problem of limited data resources, further pushing toward practical and accessible solutions for the field.

Abbreviation

CNN	Convolutional Neural Networks	GT	Generator Trip
ANNs	Artificial Neural Networks	LO	Line Outage
GAs	Genetic Algorithms	LD	Load Disconnection
DNNs	Deep Neural Networks	ROCOF	The Rate of Change of Frequency
DWT	Discrete Wavelet Transform	PDC	Phasor Data Concentrator
CE	Cross-Entropy	FC	Fully Connected
ReLU	Rectified Linear Unit	SNRs	Signal-to-Noise Ratios
OCs	Operating Conditions		

* Corresponding author
E-mail address: mjazaeri@semnan.ac.ir
https://orcid.org/0000-0003-0041-7968
http://dx.doi.org/10.48308/ijrtei.2025.238734.1075

1. Introduction

Fault classification has been one of the prime necessities for the reliability and safety of systems under consideration, especially in power systems, manufacturing, and transportation [1]. Correct identification and classification of faults are critical to timely intervention, thereby minimizing downtime and preventing possible hazard. In the last decade, machine learning, especially deep learning, has catapulted fault classification to new heights with abilities to produce superior accuracy and better feature extraction capabilities than traditional methods [2]. Deep learning models, such as CNNs, excel in capturing complex patterns and dependencies in data [3]. However, their performance is heavily reliant on the availability of large, labeled datasets. This poses a significant challenge in many real-world scenarios where data collection is expensive, time-consuming, or infeasible due to operational constraints [4]. In such cases, the limited size of the dataset can hinder the model's ability to generalize effectively, leading to suboptimal performance. Transfer learning has also emerged as a promising solution to address the limitations of small datasets. By utilizing knowledge from a pretrained model, transfer learning allows for the adaptation of a deep learning model to a new domain with minimal additional training [5]. This approach has been successfully applied in various fields, including computer vision, natural language processing, and medical diagnostics, demonstrating its potential to enhance performance in data-scarce environments [6].

Given the high importance of the subject under study, as mentioned in the previous section, many studies have been conducted in this field so far, and in this section, we will examine new studies. Fault location in power systems can be categorized into three methodologies: traditional, observant, and intelligent [7]. Traditional methods rely on customer reports, such as noticing downed wires, while observant methods utilize intelligent meters or local detectors to provide feedback to operators. Intelligent methods employ advanced technologies like smart sensors and expert systems, including Artificial Neural Networks (ANNs) and Genetic Algorithms (GAs), for fault detection. Additionally, data-driven approaches leverage machine learning to analyze large datasets, while model-driven techniques rely on mathematical models of the power system. The choice of methodology depends on the specific application and data availability. Smart grids, despite their advancements in communication and information technologies, remain prone to faults. Therefore, accurate fault detection, classification, and localization are essential for efficient maintenance, rapid fault identification, and power restoration [7].

It is also useful to use data-driven approaches to detect faults in nonlinear systems since they are not dependent on system structure [8,9]. As deep learning and parallel computing hardware have advanced rapidly, data-driven methods have emerged as highly promising solutions for real-time fault diagnosis. Further, data-driven algorithms are highly noise-resistant, making them ideal for dealing with complex classification problems. It has been demonstrated that Deep Neural Networks (DNNs), such as ANNs and CNNs, are capable of detecting and classifying faults [10-12].

The complexity of ANN structures has also increased over the past few years, with different architectures designed to address different application scenarios. An advantage of ANN-based methods is that they do not require a pre-existing knowledge base for fault detection. In this way, they can detect, locate, and classify faults in the power system rapidly and precisely [12,13]. Fault classification has been done with CNNs, and they can be complemented by other techniques, such as the Discrete Wavelet Transform (DWT), to develop fault classification approaches [10]. In the smart grid, deep learning algorithms can be trained on labeled fault data to detect, classify, and locate faults in real time [14-16]. CNNs are commonly used for classification and computer vision tasks [17]. While ANNs are suitable for handling a variety of data types, CNNs are best suited for image-based data [14,18]. Power distribution grids, transmission lines, and photovoltaic modules have all been used in the detection and diagnosis of faults with these algorithms. These approaches are aimed at creating a system based on ANNs that can identify and classify transmission line faults as soon as they occur in a timely manner. ANNs and CNNs can diagnose faults in the smart grid in real time, providing the fault type and location, enabling operators to take appropriate actions. By utilizing these models, power systems can be enhanced in terms of reliability and efficiency, resulting in fewer downtimes and improved customer satisfaction. For the detection and isolation of faults in microgrids without shutting down the entire system, a study introduced a DNN-based approach (ANN and CNN). With these algorithms, smart grids can become more reliable and efficient [19-21]. Detecting faults in the network was done using measurements of current and voltage, which were pre-processed to identify characteristic changes in current and voltage signals.

The proposed model's DNN algorithm in [22] can detect faults in medium or low-voltage transmission systems as well as distribute systems. A data-driven model is proposed for identifying fault line identifiers, fault class types, and fault location estimators in smart grids using DNNs, including but not limited to ANNs and CNNs. The proposed data-driven model can detect a variety of fault types, including single-line to ground, double-line to-ground, and three-phase faults. As part of the proposed scheme, fault classes, faulty lines, and fault locations in the grid can be detected simultaneously. [23] eliminates preprocessing steps as in [24], feature engineering, and signal conversion, resulting in a more efficient method. In previous studies, voltage and current signals were converted into grayscale images or other transformations to extract meaningful features from signals.

While fault location methods based on machine learning have shown to be effective in some simulation scenarios, their application is always hindered by the small fault data sets from practical transmission lines. With large data sets that have similar distributions, transfer learning can reach a fast convergence with a small set of data. For VSC-HVDC transmission lines, [25] proposed a transfer learning-based fault location method and discussed its performance in different scenarios. As part of the method, stacked denoising auto-encoders are used to model the relationship between traveling waveforms and fault locations, and small data sets from the target transmission line are used to fine-tune the model. On a real-time digital simulation platform, [25] tests the proposed method with a simulated VSC-HVDC transmission line.

It is often necessary to consider many aspects of the specific problem to be solved in order to obtain satisfactory results with data-driven methods, regardless of whether they are based on classical machine learning or deep learning [26]. For example, the number and accuracy of training samples, number of classes to be classified, and the degree of separation between classes all play a role. The classical machine learning classifiers and DNNs will become overfitted if there are too few training samples [27]. When a model is overfitted, it cannot generalize well to new data and is thus unable to perform its intended classification task effectively. In nuclear power plant equipment fault diagnosis and system accident identification studies, the lack of labeled training samples, i.e., experimental data with known fault or accident patterns, is a common problem because gathering sufficient training data is expensive, time-consuming, or even impractical, especially for accidents. When data is lacking, models are usually unsatisfactory.

For training and testing machine learning models in the nuclear industry, most published research uses simulation data instead of real fault or accident data. The accident data in [28] is collected from a system analysis software called PCTRAN (Micro-Simulation Technology - Nuclear Power Plant Simulation, n.d.) when authors use five machine learning models to identify accident classes in a pressurized water reactor. Similarly, [29] uses Boolean networks to identify faults in a modular high-temperature gas-cooled reactor, and the accident data come from a full-scale simulation. [27] proposes a DNN-based transfer learning approach to reconcile the strong dependence of machine learning models on data with the scarcity of real fault or accident data in nuclear power plants. By using limited data, it is possible to train machine learning models that can handle nuclear power plant classification problems. To minimize the number of labelled samples or computing time required for training in the target domain, transfer learning uses the knowledge of the related domain (called the source domain). To avoid manual feature selection, the proposed method uses a CNN as a carrier.

As is evident, the existence of a small database may create limitations that can be referred to as weaknesses and study gaps in recent articles. In this paper, we present a transfer learning-based CNN framework designed explicitly for fault classification with small databases. Our methodology involves training a CNN model on a large-scale dataset from a related domain and fine-tuning it for fault classification using a smaller, domain-specific dataset. By combining the powerful feature extraction capabilities of deep learning with the efficiency of transfer learning, our approach addresses the challenges of limited data availability while maintaining high classification accuracy. The proposed method offers a practical and scalable solution for fault classification, contributing to the broader adoption of deep learning techniques in resource-constrained scenarios.

The remainder of this paper is organized as follows: Section 2 represents the formulation of the problem. Section 3 outlines the proposed methodology, including the CNN architecture and transfer learning strategy. Section 4 presents the simulation, results, and analysis. Finally, Section 5 concludes the paper and discusses potential directions for future research.

2. Transfer learning formulation

Transfer learning is a machine learning technique that leverages pre-trained models on large datasets to enhance learning efficiency on new, often smaller, datasets. Fine-tuning is a common approach within transfer learning where a pre-trained model is slightly adjusted to perform well on a new task. Instead of training a model from scratch, which requires significant data and computational resources, fine-tuning modifies the weights of a pre-trained model (typically from a related task) by continuing its training on the new dataset for a few more epochs. This approach helps to retain the knowledge already learned while adapting it to the new task. Fine-tuning is particularly beneficial in cases where labeled data is scarce. The basic formulation of fine-tuning in transfer learning can be expressed as follows:

1. Pre-trained Model $M_{pre-trained}$: A model M trained on a large dataset D_{large}, where M has learned generalizable features.
2. Fine-Tuning: Let $M_{fine-tuned}$ be the model that undergoes fine-tuning on a new dataset D_{New}, where only a few layers or the entire model might be updated. $$M_{fine-tuned} = FineTune(M_{pre-trained}, D_{New}) \quad (1)$$
3. Loss Function: The objective is to minimize the loss function L on the new dataset: $$L(M_{fine-tuned}, D_{New}) = \frac{1}{N}\sum_{i=1}^{N} L(y_i, \hat{y}_i) \quad (2)$$ where y_i is the true label and $\hat{y}_i$ is the predicted output for each sample in D_{New}.
4. Learning Rate: Fine-tuning typically involves using a lower learning rate η to prevent overfitting and retain the model's learned features:

$$\theta_{new} = \theta_{old} - \eta \nabla L(\theta_{old}, D_{New}) \quad (3)$$

where θ_{old} stand for the weights of the pre-trained model and θ_{new} are the adjusted weights after fine-tuning.

3. Proposed Methodology

The proposed model consists of three parts: database generation, model training and online application. Figure 1 shows the general flowchart of the multi-structure model.

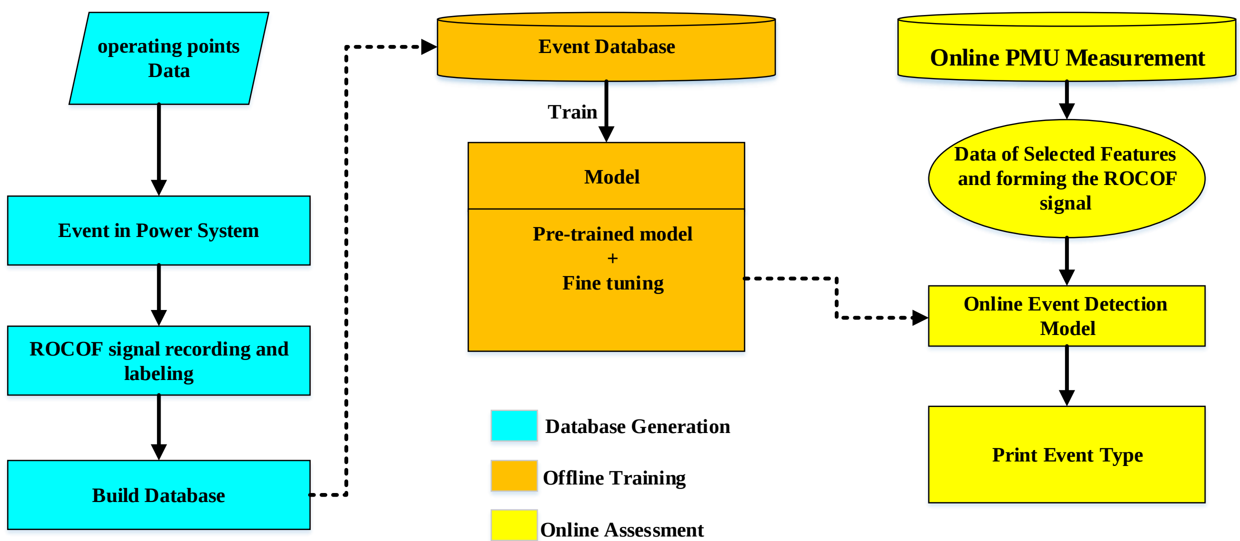

Fig 1: Event detection model flowchart

3.1. Data acquisition

Creating a suitable database is the first step in training a model. In this section, a small database containing Generator Trip (GT), Line Outage (LO), Load Disconnection (LD), and fault events is constructed. To achieve this, power system events are applied, and time domain simulation is carried out. The rate of change of frequency (ROCOF) signal collected by phasor data concentrator (PDC) illustrates these variations more clearly than the frequency signal. Each of these events causes a sudden change in the system's frequency. The ROCOF equation is as follows:

$$ROCOF_{t(k)} = \frac{F_{t(k-\tau+1)} - F_{t(k)}}{t(k-\tau+1) - t(k)} \quad (4)$$

where the value of ROCOF at timestamp k is indicated by $ROCOF_{t(k)}$; $F_{t(k)}$ and τ are the frequency at timestamp k and time interval, respectively. At the end, the ROCOF signal is stored for each of events at different operating points and a corresponding label is assigned for classification.

3.2 Model training

In this step, the CNN model is first trained with the CIFAR-10 large database. Then, with the small database generated for the related events, the classification model is built. The architecture of the proposed CNN model is depicted in Fig. 2.

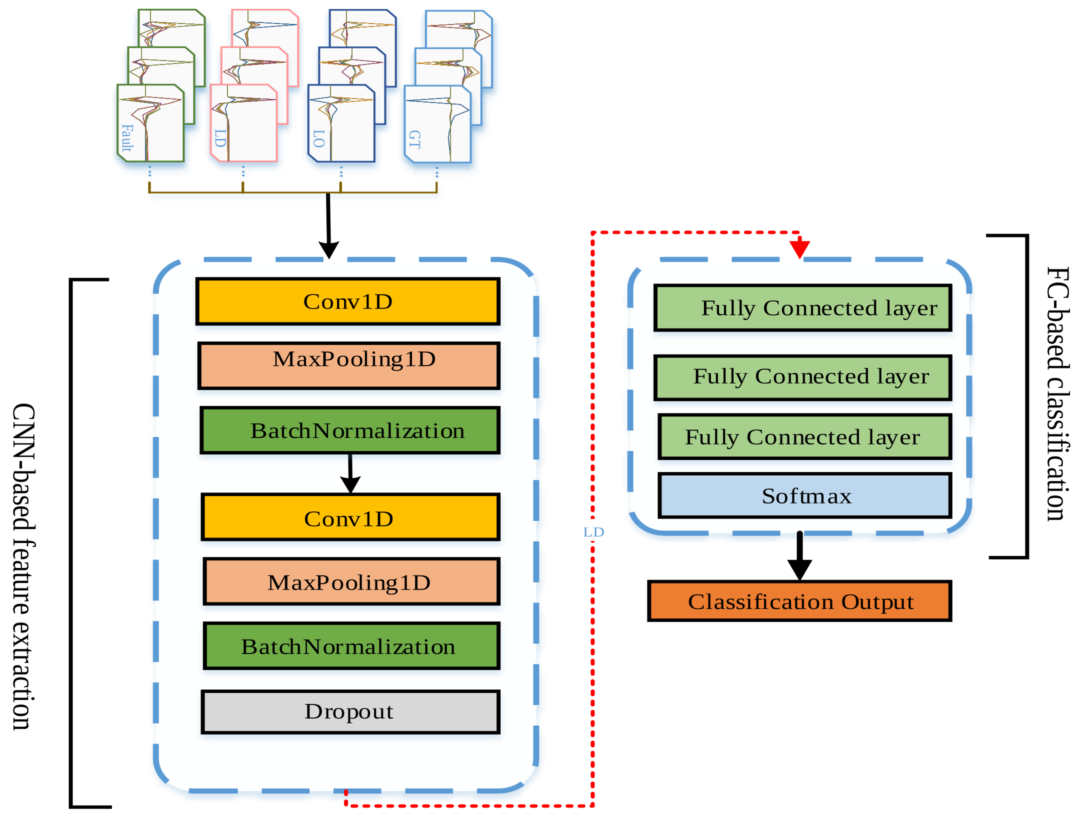

Fig. 2: The architecture of the proposed CNN model

There are two convolution layers in this study, two maxpooling layers, three fully connected layers for interpreting features, and one fully connected layer for predicting classes. The convolutional layers were constructed using kernels of size 3, and each activation block was based on the Rectified Linear Unit (ReLU). In the structure of CNN, the maximum value function is used to design the pooling layer, which has a pooling size and strides of 2 and 1, respectively. The batch normalization process has also been used to improve domain adaptation and learning convergence. After the dropout layer, the output is reshaped into one-dimensional vectors using the flatten block before the densely connected classifier. Dropout was used to reduce overfitting before the flatten layer. The dense layers were activated using ReLU, while the output layer was classified using softmax.

$$y = soft\,max(w_d \times s + b_d) \tag{5}$$

In eq. (5), s represents the input of the softmax layer. Also, w_d and b_d are the weight and bias matrices that the assessment model must learn during training.

An offline training involves training the CNN-based assessment model to minimize the difference between the predictions and the actual states, and determining the learning parameters. In order to accomplish this goal, a loss function and a learning parameter optimization algorithm are necessary. Model predictions and actual states are compared using the loss function, and the optimization algorithm attempts to reduce the loss function by iteratively updating the learning parameters. A number of studies have utilized the cross-entropy (CE) function for classification tasks, so it has been extensively utilized [30]. Adam's algorithm has been used to optimize CE in this work. It is an important algorithm in deep learning.

In the proposed transfer approach for model training, the structure and parameters of the pre-trained model are transferred to the new classification model, and all layers are fine-tuned with a small database. In this case, the optimal parameters of the pre-trained model are selected as the initial parameters of the new model.

$$w' = initialize(w_{out}, w_l, w_{l-1}, \dots, w_1) \tag{6}$$

$$b' = initialize(b_{out}, b_l, b_{l-1}, \dots, b_1) \tag{7}$$

3.3 Online Application Stage

In the online stage, the data measured by the PMUs are used as input to the model. Once the information about the PMUs is collected, they are entered into the trained model, whose optimal parameters have been determined during the offline training phase. Finally, the results of a power system are determined instantly. To assess the performance of the models, this paper examines their accuracy, F1-score, recall, and precision [31, 32].

Phase 1: Pretraining on CIFAR-10

The base CNN is trained on the CIFAR-10 dataset to develop strong feature extraction capabilities. This phase employs the following steps:

1. **Training Objective:** The model minimizes the categorical cross-entropy loss.
2. **Optimization:** The model parameters are updated

Phase 2: Fine-tuning with Transfer Learning

1. **Model Modification:** The pretrained CNN model is modified by adding two fully connected (FC) layers at the end to accommodate the specific fault classification task. The final architecture is structured as follows:
    - The convolutional layers from the pretrained model are frozen to retain the learned feature extraction capabilities.
    - The two new FC layers are initialized randomly.
2. **Fine-tuning Objective:** The added layers are trained using the smaller fault classification dataset.
3. **Optimization:** The fine-tuning process updates only the parameters of the new FC layers, while keeping the convolutional layers frozen.

The training process can be summarized as follows:

1. Train the base CNN on CIFAR-10 and save the pretrained model.
2. Add two fully connected layers to the pretrained model for the fault classification task.
3. Freeze the convolutional layers and fine-tune the new layers using the small fault classification dataset.
4. Evaluate the model's performance using metrics such as accuracy and F1-score.

This two-phase approach ensures that the model leverages the large-scale CIFAR-10 dataset for robust feature extraction, while efficiently adapting to the target domain with minimal additional data.

4. Simulation results

In order to analyze the efficiency of the proposed method, it has been implemented on the IEEE 39-bus system. The diagram of the studied system, which includes 10 generators, 46 transmission lines and 39 buses, can be seen in Fig. 3. The simulation results for the proposed method have also been compared with LSTM and GRU. Table 1 shows the main parameters of CNN. Hyperparameters and transfer learning configurations were selected through trial and error. Digsilent Power Factory and Python have been used to develop the proposed method.

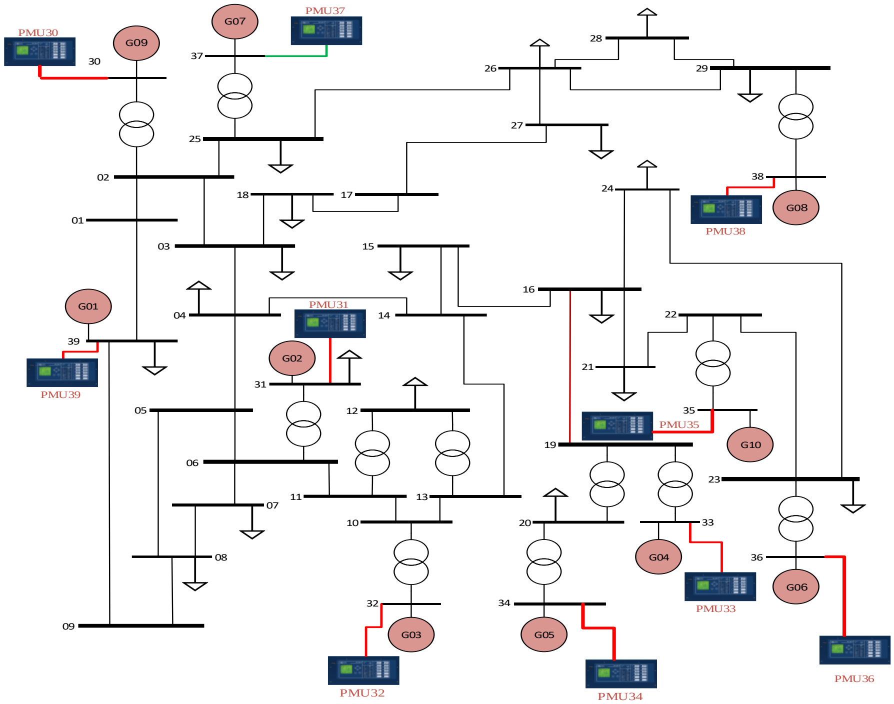

Fig. 3: The IEEE 39-bus system.

Table 1. CNN structure and hyper-parameters.

Layers	Hyper-parameters
Convolution Layer	Number of kernels:32 Size of kernels: 3 Strides: 1
Pooling Layer (MaxPooling)	Size of pooling:2 Strides: 1
Batch Normalization Layer	-
Convolution Layer	Number of kernels:64 Size of kernels: 3 Strides: 1
Pooling Layer (MaxPooling)	Size of pooling:2 Strides: 1
Batch Normalization	-
Dropout	dropout rate: 0.1
Flatten	-
Dense	128 units
Dense	64 units
Dense	32 units
Output Layer (Dense)	4 classifications

4.1 Generation of Original Small Database

In this part of the study, a small original database for different operating conditions (OCs) is generated to train and evaluate the proposed model, as described in the following. The OCs are obtained by random sampling of the load within the range of its practical changes. Specifically, according to the base load level for each bus, random sampling of load changes is performed. The range of load changes is considered between 0.7 and 1.25 of the base load. Then, the generators' output power is calculated using the optimal power flow for each load level. Different OCs will be obtained as a result of this work. Then, labelling is performed according to the GT, LO, LD, and three-phase fault events. In the next step, 80% of the OCs are randomly selected for training and 20% for testing.

4.2 Performance testing of the proposed model

In this section, the performance of the proposed model with and without transfer learning is examined and compared with the GRU and LSTM algorithms. Table 2 shows the performance accuracy results of different models. As the results show, using transfer learning has increased the accuracy of the CNN-based model by 3.75%. Also, the proposed model has superior performance and achieved higher accuracy compared to the GRU and LSTM models. The performance accuracy of the proposed model is 3.67 % and 3.94 % higher than GRU and LSTM, respectively. The proposed model achieves high performance accuracy without the need for a large database, which is a prerequisite for excellent performance of DL-based models. However, data augmentation techniques that are commonly used to address the challenge of small databases cannot achieve high fidelity in data generation. Therefore, the proposed model can be used as an effective approach in practical applications. The proposed model's online computation time for each sample is 0.15 milliseconds, which is suitable for online applications.

Table 2: Comparison of classifiers from accuracy point of view

Classifier	Accuracy (%)
Proposed	99.44
CNN	95.69
LSTM	95.50
GRU	95.77

In order to better demonstrate the proposed model's superiority, its performance is further analysed based on other evaluation indicators, the results of which are shown in Table 3. Recall, Precision and F1-score for the proposed model are 99.32%, 99.65% and 99.53% respectively. The results show that the performance pf the proposed model based on transfer learning is much better than other models.

Recall and precision represent false negatives and false positives. In this work, recall is less important than precision. This indicates that the model has more false negatives than false positives. Also, this work, the F1 score is used as a suitable trade-off between recall and precision. It has a high value and indicates the proper performance of the proposed model.

Table 3: Comparison of different classifiers from viewpoints of statistical indicators

Classifier	Recall (%)	Precision (%)	F1-score (%)
Proposed	99.32	99.65	99.53
CNN	94.23	96.41	98.97
LSTM	95.25	94.84	98.73
GRU	96.91	95.42	96.16

Finally, this section examines the effect of transfer learning on reducing database size. As shown in Fig. 4, the transfer learning-based CNN requires 750 samples for effective training. In this case, the accuracy of the model is 99.44%. While the CNN model without transfer learning reaches 99.44% accuracy with 1650 samples. This shows that transfer learning allows for training the model with a small database. In addition, it also reduces the update time if updates are needed.

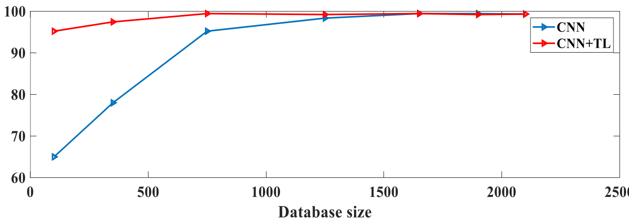

Fig. 4 :Impact of transfer learning on database size

4.3 Performance test with renewable energy integration

As renewable energy units continue to merge into the power grid, the dynamic characteristics of the system following disturbances grow increasingly complex. In order to assess the implications of renewable energy unit integration on the proposed methodology, this work examine its performance under varying rates of renewable energy penetration. The outcomes of these assessments are depicted in Table 4.

Table 4: Performance of the proposed approach under different renewable energy penetration rates

Penetrations rates (%)	Accuracy (%)
0	99.44
10	99.31
20	99.15
30	99.15

The results of Table 4 show that the proposed model with the penetration of renewable energy sources also has a suitable and promising performance. However, with the influence of these sources, the accuracy of the model's performance has decreased, and the authors plan to examine this issue with more focus in future work. However, up to a penetration factor of 30%, the proposed model has achieved satisfactory accuracy and is robust.

4.4 Performance with PMU missed data

PMU information may be incomplete for various reasons, such as cyber-attacks, communication line interruptions, etc. In these conditions, data-driven models usually suffer from reduced accuracy. In this section, the performance of the proposed model for PMUs with missing data is examined. Fig. 5 shows the average accuracy of the model performance for different cases where part of the 5 PMU information is missing. As the results show, the model achieves an accuracy of 97.47% for the case where part of the 5 PMU information is missing, which is remarkable. Also, the model accuracy is higher for times when less PMU information is missed. It is important to note that given the importance of event classification, it is necessary to recover the missed PMU information. This topic is beyond the focus of this paper

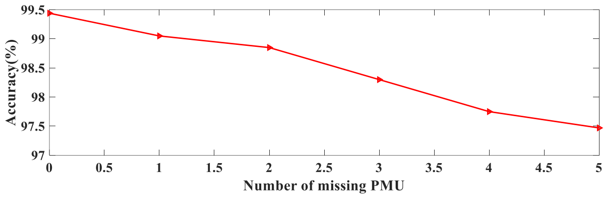

Fig. 5: Model performance with missing data

4.5 Robustness Test under Noisy Environments

Since noises are inevitable during the collection and transmission process of PMU data, the robustness of the proposed method is tested under noisy environments. Here, Gaussian white noises with different signal-to-noise ratios (SNRs) are added into the PMU data, where a smaller SNR indicates a higher noise level. Here, 3 scenarios with SNRs 50 dB, 40 dB, and 30 dB are respectively tested in this section. It can be seen from Table 5 that although the assessment accuracy will is decreased slightly with the increase of the background noise level, it is always above 99.05% in various noise environments.

Fig. 5: Accuracy of the proposed models for noisy data

SNR	Accuracy (%)
50	99.35
40	99.18
30	99.05

5. Conclusion

This paper introduces a CNN-based method for data-driven event classification of power systems, which solves the challenge of small database. For this purpose, transfer learning-based CNN has been used. In the proposed transfer learning approach, fine tuning of the entire model is applied. Using transfer learning has increased the accuracy of the model by 3.75%. The proposed model's accuracy was 99.44%. Recall, precision, and F1-score were recorded as 99.32%, 99.65%, and 99.54%, respectively. The proposed model shows a high accuracy of 99.05 against noisy data, indicating its robustness. Incomplete PMU data caused the model accuracy to drop by 1.97%. This drop in accuracy indicates that data recovery should be considered in this situation. PMU data recovery is a focus of the authors' future work. For this purpose, the authors use parallel models to achieve the least time delay in event classification.

The performance of machine learning models when working with small and unbalanced datasets is challenging. A small dataset makes it difficult for the model to train well and obtain optimal parameters, and the imbalance causes the model to be biased towards the majority class. These challenges negatively affect the accuracy and robustness of the results. Therefore, the authors plan to address this challenge in future work.

References

1. Shukla, P. K., & Deepa, K. (2024). Deep learning techniques for transmission line fault classification–A comparative study. *Ain Shams Engineering Journal*, *15*(2), 102427.
2. Zhou, M., Kazemi, N., & Musilek, P. (2024). Distribution Grid Fault Classification and Localization using Convolutional Neural Networks. *Smart Grids and Sustainable Energy*, *9*(1), 1-17.
3. Siddique, M. N. I., Shafiullah, M., Mekhilef, S., Pota, H., & Abido, M. A. (2024). Fault classification and location of a PMU-equipped active distribution network using deep convolution neural network (CNN). *Electric Power Systems Research*, *229*, 110178.
4. Punitha, K., Sivapriya, G., & Jayachitra, T. (2025). CNN Based Fault Classification and Predition of 33kw Solar PV System with IoT Based Smart Data Collection Setup: CNN PV Fault IoT. *EAI Endorsed Transactions on Energy Web*, *12*.
5. Yan, P., Abdulkadir, A., Luley, P. P., Rosenthal, M., Schatte, G. A., Grewe, B. F., & Stadelmann, T. (2024). A comprehensive survey of deep transfer learning for anomaly detection in industrial time series: Methods, applications, and directions. *IEEE Access.*
6. Zhang, Y., Wang, X., He, J., Xu, Y., Zhang, F., & Luo, Y. (2020). A transfer learning-based high impedance fault detection method under a cloud-edge collaboration framework. *IEEE Access*, *8*, 165099-165110.
7. De La Cruz, J., Gómez-Luna, E., Ali, M., Vasquez, J. C., & Guerrero, J. M. (2023). Fault location for distribution smart grids: Literature overview, challenges, solutions, and future trends. *Energies*, *16*(5), 2280.
8. Li, B.; Delpha, C.; Diallo, D.; Migan-Dubois, A. Application of Artificial Neural Networks to photovoltaic fault detection and diagnosis: A review. Renew. Sustain. Energy Rev. 2021, 138, 110512.
9. Samanta, A.; Chowdhuri, S.; Williamson, S.S. Machine learning-based data-driven fault detection/diagnosis of lithium-ion battery: A critical review. Electronics 2021, 10, 1309.

10. Pan, P.; Mandal, R.K.; Rahman Redoy Akanda, M.M. Fault classification with convolutional neural networks for microgrid systems. Int. Trans. Electr. Energy Syst. 2022, 2022, 8431450.
11. Karić, A.; Konjić, T.; Jahić, A. Power system fault detection, classification and location using artificial neural networks. In Advanced Technologies, Systems, and Applications II: Proceedings of the International Symposium on Innovative and Interdisciplinary Applications of Advanced Technologies (IAT), Banja Vrucica, Bosnia and Herzegovina, 25–28 May 2017; Springer: Cham, Switzerland, 2019; pp. 89–101.
12. Bishal, M.R.; Ahmed, S.; Molla, N.M.; Mamun, K.M.; Rahman, A.; Al Hysam, M.A. ANN Based Fault Detection & Classification in Power System Transmission line. In Proceedings of the 2021 International Conference on Science & Contemporary Technologies (ICSCT), Dhaka, Bangladesh, 5–7 August 2021; pp. 1–4.
13. Shakiba, F.M.; Shojaee, M.; Azizi, S.M.; Zhou, M. Real-time sensing and fault diagnosis for transmission lines. Int. J. Netw. Dyn. Intell. 2022, 1, 36–47.
14. Mohd Amiruddin, A.A.A.; Zabiri, H.; Taqvi, S.A.A.; Tufa, L.D. Neural network applications in fault diagnosis and detection: An overview of implementations in engineering-related systems. Neural Comput. Appl. 2020, 32, 447–472.
15. Mamuya, Y.D.; Lee, Y.D.; Shen, J.W.; Shafiullah, M.; Kuo, C.C. Application of machine learning for fault classification and location in a radial distribution grid. Appl. Sci. 2020, 10, 4965.
16. Alvarez, G.P. Real-time fault detection and diagnosis using intelligent monitoring and supervision systems. In Fault Detection, Diagnosis and Prognosis; IntechOpen: London, UK, 2020.
17. Wen, L.; Gao, L.; Li, X.; Zeng, B. Convolutional neural network with automatic learning rate scheduler for fault classification. IEEE Trans. Instrum. Meas. 2021, 70, 1–12.
18. Gao, H.X.; Kuenzel, S.; Zhang, X.Y. A hybrid ConvLSTM-based anomaly detection approach for combating energy theft. IEEE Trans. Instrum. Meas. 2022, 71, 1–10.
19. Alrifaey, M.; Lim, W.H.; Ang, C.K.; Natarajan, E.; Solihin, M.I.; Juhari, M.R.M.; Tiang, S.S. Hybrid deep learning model for fault detection and classification of grid-connected photovoltaic system. IEEE Access 2022, 10, 13852–13869.
20. Yoon, D.H.; Yoon, J. Deep learning-based method for the robust and efficient fault diagnosis in the electric power system. IEEE Access 2022, 10, 44660–44668.
21. Zhang, S.; Zhang, S.; Wang, B.; Habetler, T.G. Deep learning algorithms for bearing fault diagnostics—A comprehensive review. IEEE Access 2020, 8, 29857–29881.
22. Coban, M.; Tezcan, S.S. Feed-Forward Neural Networks Training with Hybrid Taguchi Vortex Search Algorithm for Transmission Line Fault Classification. Mathematics 2022, 10, 3263.
23. Alhanaf, A. S., Balik, H. H., & Farsadi, M. (2023). Intelligent fault detection and classification schemes for smart grids based on deep neural networks. *Energies*, *16*(22), 7680.
24. Azizi, S.; Jegarluei, M.R.; Dobakhshari, A.S.; Liu, G.; Terzija, V. Wide-area identification of the size and location of loss of generation events by sparse PMUs. IEEE Trans. Power Deliv. 2020, 36, 2397–2407.
25. Shang, B., Luo, G., Li, M., Liu, Y., & Hei, J. (2023). Transfer learning-based fault location with small datasets in VSC-HVDC. *International Journal of Electrical Power & Energy Systems*, *151*, 109131.
26. 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. *International Journal of Electrical Power & Energy Systems*, *134*, 107399.
27. Zhong, X., & Ban, H. (2022). Pre-trained network-based transfer learning: A small-sample machine learning approach to nuclear power plant classification problem. *Annals of Nuclear Energy*, *175*, 109201.
28. Li, J., & Lin, M. (2022). Research on robustness of five typical data-driven fault diagnosis models for nuclear power plants. *Annals of Nuclear Energy*, *165*, 108639.
29. Dong, Z. (2017). Boolean network-based sensor selection with application to the fault diagnosis of a nuclear plant. *Energies*, *10*(12), 2125.
30. Azad, S., & Ameli, M. T. (2024). A domain adaptation-based convolutional neural network incorporating data augmentation for power system dynamic security assessment. *The Journal of Engineering*, *2024*(7), e12400.
31. Azad, S., & Ameli, M. T. (2025). An Imbalanced Deep Learning Framework for Pre-Fault Flexible Multi-Zone Dynamic Security Assessment Via Transfer Learning Based Graph Convolutional Network. Results in Engineering, 104172.