Jestr

Journal of Engineering Science and Technology Review 17 (2) (2024) 35 - 44

Research Article

JOURNAL OF
Engineering Science and Technology Review

www.jestr.org

Bayesian Optimized CNN Model for Fault Classification in a Distribution System

Garima Tiwari* **and Sanju Saini**

Department of Electrical Engineering, Deenbandhu Chhotu Ram University of Science and Technology, Murthal, Sonipat, Haryana, 131027, India

Received 20 July 2023; Accepted 25 March 2024

Abstract

A fault in a power system is an anomalous state that must be recognized as soon as feasible. To minimize the repercussions of the fault, such as damages occurred to the device, loss of tangible assets and loss of human resources, it is critical to notice the problem promptly. In a power distribution system, there are several approaches for detecting different types of faults. In this paper, a neoteric approach using Bayesian optimized Convolutional Neural Network is used to detect and classify different symmetrical as well as unsymmetrical faults in power distribution systems. The effectiveness of the proposed CNN model is validated for an IEEE 13 bus radial distribution system grid modeled (and simulated) in PSCAD. Time series of the measured 3-phase fault currents (for eleven different categories of faults) are used to create training & testing data. This data has been imported in MATLAB software to develop a CNN classifier (whose hyper-parameters are optimized by using a Bayesian optimizer) for faults in power distribution system, under distinct fault situations by varying fault resistance, faulty node and fault inception angles. Findings of simulation clearly indicate that proposed model has very high categorizing accuracy and is superior and competitive to other techniques available in literature.

Keywords: Bayesian optimization, Convolutional Neural Network (CNN), Fault classification, Hyper-parameters, Radial distribution systems.

1. Introduction

Distribution systems are consistently exposed to danger of short circuit faults which lead to power loss. The relaying system's crucial stages of fault detection and classification must be completed successfully and quickly in order to clear faults so that speedy restoration may be accomplished. Techniques used for fault diagnosis in transmission grids cannot be used immediately for distribution grids on account of their structural complexities like non-homogeneity, presence of laterals etc. [1]. Moreover, the standard relaying techniques are becoming ineffective due to shifting fault current levels as distributed generators (DGs) are used more frequently in distribution networks. The chance of relay operations on non-faulted phases can be decreased with accurate real-time fault classification, which can allow superior grid operation. The practical objective of online fault classification is made possible by dispersed measurement equipment in smart distribution grids.

To take into account the complexities and uncertainties of distribution system, many researchers have been using machine learning based fault diagnosis systems for distribution networks using knowledge from the data corresponding to different conditions. Among them, W.H.Chen et.al. [2] have explored and demonstrated the fuzzy logic-based (FL) techniques for precise classifications of faults type in distributed power system. In [3], figures gathered by alarms & protective relays in power network have been evaluated by J. C. S. desouza et. al. utilizing Neuro fuzzy methods. D. Thukaram et. al. first utilized SVM (Support Vector Machines) to categorize distinct kinds of fault, and then an ANN (Artificial Neural Network) is used to locate the problem site [4]. L. C. Acacio et. al. [5] have used and compared the accuracy of different neural network structures for detection of single L-G (line to ground) faults. N. Wang et. al. [6] have utilized a SVM (support-vector-machine) and PCA (principle-component-analysis) for fault classification taking into account the impact of measurement noise as well as loading conditions. Hosseini et.al. [7] have used the data evaluated by smart meter as input of a multi-label SVM, for detecting defective lines in power distribution system.

Also, a mix of signal processing as well as machine learning is used by many researchers for distribution system fault detection, e.g., by using spectral properties of observations, feature extracted data is delivered to an artificial neural network (ANN) for distinct faults' classification [8]. In [9], a fault diagnosis scheme using ART (Adaptive Resonance Theory) neural network combined with time-time (T-T) transform is developed. S. Jana et. al. [10] have combined the concepts of wavelet entropy and artificial neural network for diagnosis of distribution grid faults. Fuzzy logic technique has also been employed in tandem with DWT (Discrete Wavelet Transform) for precise identification of distinct kind of faults [11 - 13]. J. Zhang et. al. [14] have used an adaptive neuro-fuzzy inference system combined with wavelet transform technique for diagnosing faults in a distribution grid. For the same problem, a new classifier called Robust Semi-Supervised Prototypical Network (RSSPN) based on Prototypical Network architecture and semi-supervised learning is proposed by T. Zheng et. al. [15]. For classification of single phase to ground faults, a multilabel classification model using 8-D feature space and a 14-label fault-type space is proposed by Y. Liang et. al. in [16]. For performing highly efficient fault data analysis, a kNN (k Nearest Neighbor) based fault identification model for single-phase-to-ground faults is proposed by J. Zhu et. al. [17]. The eigenvectors in this model are the wavelet energy ratio, the

*E-mail address: 18001902001garima@dcrustm.org
ISSN: 1791-2377
doi:10.25103/jestr.172.05

variance of the wavelet coefficients, and the wavelet power produced by decomposing transient components.

Latest innovations in field of machine/deep learning, have cached imaginations of academia and business equally. A major advancement in this field is effective deployment of CNN (Convolutional Neural Networks) as well as transfer learning algorithms in range of image processing as well as image recognition-related tasks [18]. CNN, with its deep architecture eliminates the requirement of feature extraction as it extracts features from set of training data (which is in the form of raw images) [19] It learns automatically the mid and high-level abstractions using raw data [20]-[22]. It consists of a series of three types of layers, i.e., convolve, pooling and ReLu (Rectified linear unit) layers. It is widely used in the computer vision and gives high image recognition accuracy [23]. A novel transfer learning framework based on a pre-trained LeNet-5 convolutional neural network is proposed by Shakiba et. al. [24] for fault diagnosis in transmission lines.

Md. Omaer et. al. [25] have proposed an extreme learning machine-based model for making a fast and accurate system for automatic fault classification on transmission lines. CatBoost classifier algorithm is used by V N. Ogar et. al. [26] as a machine language tool to train datasets for fault classification. A. Moradzadeh et. al. [27] have demonstrated the capabilities of SVM (support vector machine), DT (decision tree), *k*-NN *(k*-Nearest Neighbors), CNN (convolutional neural network), LSTM (long short-term memory) and Convolutional LSTM in categorizing the type and location of transmission line faults. A deep learning algorithm is used by P. Rai et. al. [28] for fault classification in the distributed networks integrated with DGs. Using Principal Component Analysis (PCA) and softmax regression, automated fault detection and classification is achieved by P. Chopra et. al. [29] using vibration and acoustic signals generated from the IC engines. The proposed technique does not require any hand-engineered feature extraction, as usually done and no pre-filtering is required on noisy industrial data [46]. In the recent years, deep learning methods have been adopted widely for fault diagnosis in Shipyard Power Systems [30]-[34].

Since the execution of deep/machine learning models has strong dependence on the choice of their hyper parameter's values [35]-[36], so, there is a need to adopt a strategy for hyper parameters' optimization of the CNN used in this work. The majority of past hyper parameter tuning research has focused solely on Grid and/or Random Search, [37]-[41]. Bayesian Optimization algorithm (using informed search) has been utilized in literature [42]-[43] very effectively and efficiently for optimizing hyper parameters of various machine learning models. Its advantage is that there is not sampling of each and every combination (with in search space) as done in Grid Search and it is more systematic as compared to Random Search. Hence, to enhance the classification exactness and efficiency, hyper parameters of the CNN used are optimized in this work.

Most of the previous works on distribution system's fault have used one or more signal processing techniques like S-Transform [1], Wavelet transform [12,13,14], Principal component analysis [14], etc. for feature extraction [46] of data to be used for training. Some of these methods show poor fault classification accuracy and some studies have not performed fault classification in detail, e.g., they perform for single phase-to-ground faults classification only. Taking these research gaps in account, in this work, we propose a holistic deep learning-based framework for fault classification. The novelty of this work is use of images of time series of fault currents by Convolutional Neural Networks to identify visual patterns to monitor the state of the distribution system and thus to categorize eleven different types of faults including symmetrical as well as unsymmetrical faults. The main contributions of this work are as follows:

1. To develop a PSCAD (4.2 version). model for IEEE 13 node radial feeder power distribution system to simulate and collect time series for eleven different types of faults i.e..(lines (R or Y or B) to Ground (G), i.e., R-G, Y-G, B-G; between two lines to Ground, i.e., (R-Y-G, Y-B-G, B-R-G); between two lines, i.e., (R-Y, Y-B, B-R); between three lines i.e., (R-Y-B); between three lines to Ground, i.e., (R-Y-B-G). under varying fault resistance, node and inception angle values.
2. Use of a Convolutional Neural Network in MATLAB to extract the features of the faults from the time series images.
3. Use of a fully connected neural network structure using Softmax activation function for fault classification.
4. Optimization of the values of CNN's hyper parameters, i.e., Section depth, initial learning rate, momentum, and L2Regularization using a Bayesian optimizer.
5. A comparison of the proposed technique with state-of-the-art methods.

Five sections make up this paper. In Section II, background theory is given. Section III explains overall methodology employed in this work. In Section IV, simulation results and comparison are given followed by conclusion and list of references.

2. Theoretical Background

2.1 Convolution Neural Network

Deep learning approaches use neural networks with numerous hidden layers and sophisticated designs. With their ability to recognize visual details, discriminate noises, and make sophisticated judgments, they have revolutionized several domains of science and technology. The three most often utilized deep learning approaches are Convolutional Neural Networks (CNN), Auto encoders, and Generative Adversarial Networks (GANs). CNN is a supervised machine learning technique for recognizing picture characteristics created by spatial correlation. These networks are mostly used to investigate data local correlations, and the model is insensitive to tiny shifts because the CNN learns the features on its own, it can produce accurate classification even without comprehensive domain knowledge. [44].

The basic architecture of CNN as shown in Fig. 1, includes a number of layers and for increasingly sophisticated models, additional layers might be used.

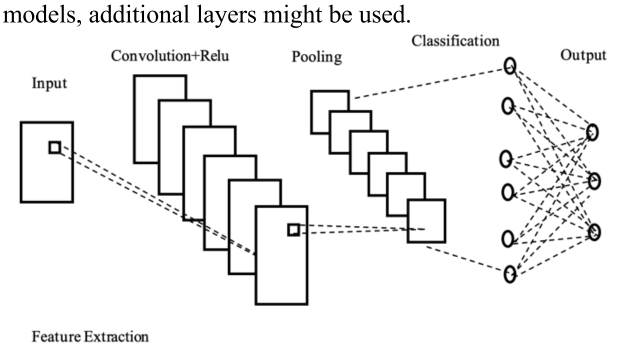

Fig. 1. The basic architecture of CNN

Each layer assists in the extraction of characteristics/features through multiple iterations. The convolution process helps to save generic features while tailoring the deeper network to the task, resulting in features which are more relevant and sophisticated. For sample reduction, a pooling layer is utilized, which can minimize the number of model parameters and, to some extent, prevent over fitting [45].

2.2 Hyper parameter Tuning [30]

Hyper parameters is a set of parameters utilized in the learning and testing process and must be pre-configured prior to the start of learning process in contrast to the internal (model) parameters (e.g. network's weights) which are automatically adjusted by the machine learning algorithms during learning [42]. The learning rate, hidden layers, count of iterations, batch size, activation functions, regularization and momentum constitute examples of hyper parameters in general. The field size of used convolutional layers, pooling layers, as well as the step size (which is dealt with help of stride parameter), are considered whfen using convolutional neural networks for image classification. The parameters may be integer, continuous or categorical variables with values ranging from lower to the upper bounds, or a combination of the two [44].

The hyperparameters used in each model differ depending on the problem. There is no such thing as a universal set of optimum hyper parameters that applies to all models. We have taken into account the learning rate, momentum, regularization, and network depth as hyper parameters in this study. The learning rate aids in the identification of generic patterns in the image, Ideal learning rate might depend on the data and/or network to be trained [16]. It's common to begin with a slow learning rate [44]. Momentum aids in the scanning of the full search region without missing important points. The gradient descent with momentum updates prior gradients at a faster rate than the traditional gradient descent. Addition of momentum with gradient descent reduces noise in the calculations and produces better results in less time by smoothening out the gradient descent steps [44].

Regularization allows the model to generalize more effectively without over fitting, allowing us to generate more accurate classification results. Network section depth decides the depth of the network and therefore helps in identifying the unique features of the images. These hyper parameters must be set (tuned/optimized) with extreme caution as they have direct control over CNN model training and its classification accuracy. For this purpose, manual or automatic search may be used [43]

Manual search is based on the core intuition and experience of an expert user with a greater professional background and practical experience. As humans are not adept at processing data which is high dimensional data and quickly misinterpret or fail to catch trends and relationship in hyper parameters, it becomes increasingly difficult to manage as the number and/or range of hyper parameters (to be tuned) grows [43].

To overcome these difficulties, automatic search algorithms, such as grid search and random search have been proposed [38]. In a high-dimensional search space, random search is superior to grid search, but it becomes unreliable when training some sophisticated models. To overcome these difficulties and make the tuning process more efficient and effective, Bayesian optimization based on Gaussian process is used for tuning of hyper parameters of CNN as discussed in the next section. It optimizes a reward-driven acquisition function that balances exploration and exploitation to choose hyper parameters intelligently. It also makes use of past data for optimizing the search domain's scope and updates the posterior distribution as the number of iterations increases [41].

2.3 Bayesian optimization model

We have used the Bayesian Optimization paradigm for optimizing the hyper parameters of the CNN classification model for its computational advantage over strategies like random and grid search-based cross-validation. It scales effectively with maximum resource utilization, handles noisy data well, and achieves global minima by utilizing non-continuous regions.

The workflow involves the usage of a probabilistic surrogate model driven by Gaussian process priors for modeling the decision search boundary and a smart querying strategy that tradeoffs exploration and exploitation. For a given function f(x), a Gaussian process model is used as a surrogate as given in eqn. 1.

$$f(x) \sim N\big(m(x), k(x, x')\big) \quad (1)$$

Where the mean function,$m(x)$ s taken to be zero and the Kernel function $k(x, x')$ incorporates our prior belief about the model in the search space. In this case, we have used an ARD (Automatic Relevance Determination) Matérn 5/2 kernel as covariance kernel function as given in equation 2.

$$k(x, x') = \left(1 + \sqrt{5a} + \tfrac{5}{3}a^2\right) exp\left(-\sqrt{5a}\right) \quad (2)$$

where,

$$a = \sqrt{\sum_{i=1}^{n} \frac{(x_n - x_n')^2}{\sigma_i^2}} \text{ for all n hyper parameters.}$$

For selecting query points, we used the Expected Improvement Acquisition function which gives expectation of the up gradation over the current optimal value over the posterior distribution of the Gaussian Process model as given in eqn. 3.

$$EI(x) = E_{f'}\,[\max(0, \mu_{f'}(x^*) - f(x))] \quad (3)$$

Where the posterior distribution of the surrogate is model and is the optimal mean value of the posterior distribution at the point. To find the optimal value of (hyper-parameters of CNN), gradient descent-based optimization techniques have been employed over the entire search space sequentially. The flow chart for Bayesian optimization of CNN hyperparameters values is given in Fig. 2.

3. Methodology

In this paper, a Convolution Neural network model is used for classification of different types of symmetrical and unsymmetrical faults in an IEEE 13 node radial feeder, power distribution system. Hyper-parameters of the CNN are optimized by Bayesian optimization technique based on Gaussian process. Overall methodology given in Fig. 3 is explained in details in the following subsections:

3.1 Modelling of IEEE 13 node Radial Distribution System

An IEEE 13 node radial test feeder (Single line diagram as shown in Fig. 4) is modelled in PSCAD software (4.2 version).

Its complete specifications are given in Table 1. Simulink model can generate 11 types of faults i.e., lines (R or Y or B) to Ground (G), i.e., R-G, Y-G, B-G; between two lines to Ground, i.e., (R-Y-G, Y-B-G, B-R-G); between two lines, i.e., (R-Y, Y-B, B-R); between three lines i.e., (R-Y-B); between three lines to Ground, i.e., (R-Y-B-G). IEEE 13 bus system provides a good test for most common features of distribution system analysis as it contains various components including highly loaded 4.16kV feeder, one three-phase voltage regulator along with three single-phase units connected in star, different types of overhead and underground lines/cables, in-line transformer and shunt capacitor bank.

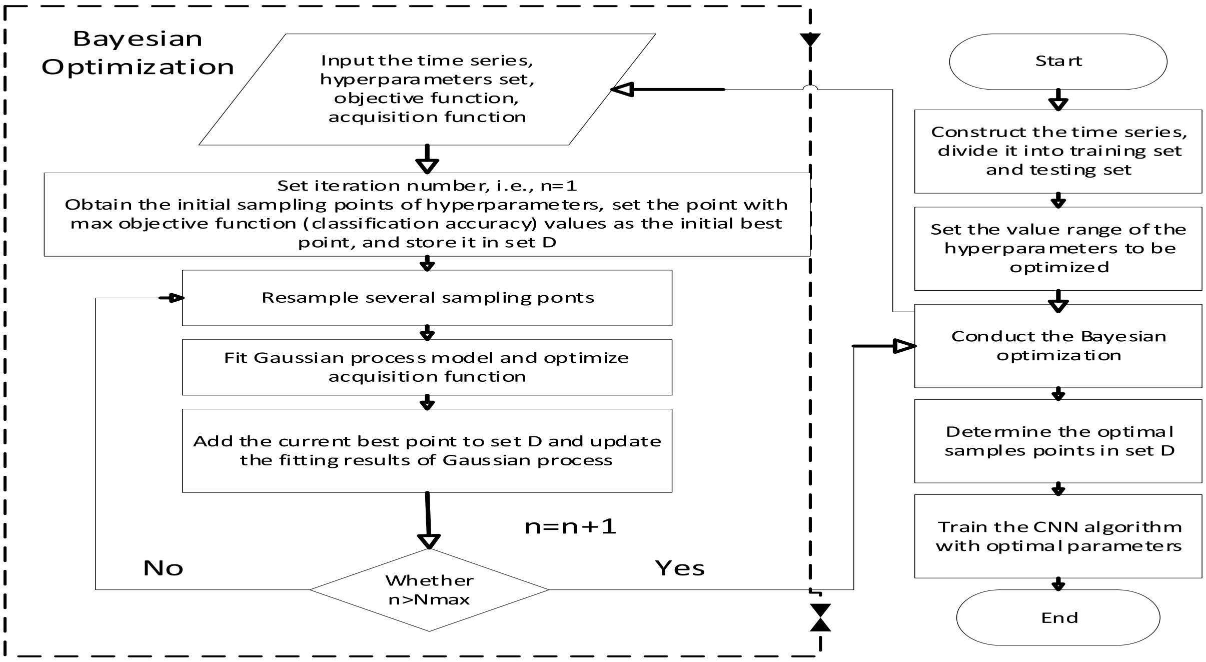

Fig. 2. Flow chart of Bayesian optimization for CNN's Hyperparameters tuning

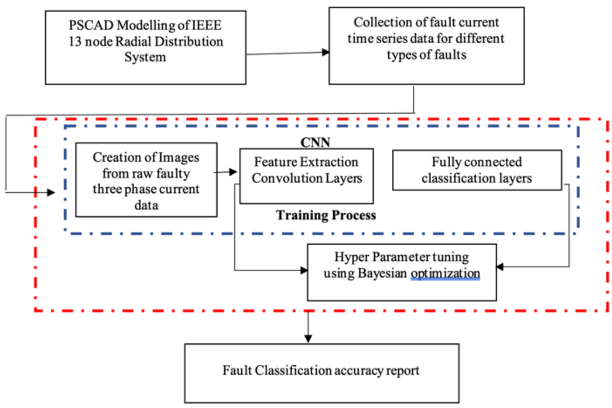

Fig. 3. Proposed Fault Classification approach

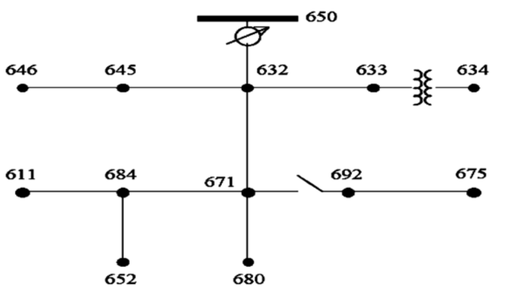

Fig. 4. Single Line Diagram of IEEE 13 Node feeder Distribution System

Table 1. Specifications of IEEE 13 node radial test feeder

Parameters	Specification
Voltage level	4.16 kV
System Frequency	60 Hz
Load types	Spot and distributed
Load Variations	±10%

3.2 Collection of fault current data

Different types of faults are applied at 8 different nodes of IEEE 13 node test feeder (as mentioned in TABLE 2) in the PSCAD Simulink model and time series of three-phase fault currents are observed at the corresponding nodes for eight different fault resistances and eight different fault inception angles (values given in TABLE 2).

Table 2. Different fault Conditions for training

Parameters	Details
Indices of Faulty Nodes	632,633,634,650 671,675,680,692
Fault inception angle (degrees)	10, 35, 60, 85, 110, 135, 160 & 185
Fault resistance (ohms)	0, 0.5, 5, 50, 100, 500, 1000 & 1500
Total no. of cases	5568

3.3 Pre-Processing of fault current data

CNN receives a time series of fault currents and uses the Gramian Angular Fields (GAF) method to automatically generate images [47]. Instead of using the more common Cartesian coordinates, Gramian Angular Field (GAF) depicts time series in a polar coordinate system. The temporal correlation between each time point is thus represented by each GAF. As training and test input data for CNN, these RGB images are employed & CNN performs feature extraction automatically. Batches of training, validation, test, and prediction data are created along with pre-processing tasks; such as scaling, rotation, and reflection. Images are resized to make them consistent with the input scale of the deep learning network. Image data is augmented with randomized pre-processing procedures to prevent the network from over fitting and remembering the precise characteristics of the training images.

3.4 Fault type classification using Bayesian optimized CNN

Using the pre-processed fault current data, The CNN model is used to predict the correct fault class for an input data. Firstly, all the hyper parameters of CNN (to be optimized using Bayesian optimization) are chosen and their ranges are specified. Also stated is whether selected variable is integer and whether the interval should be searched in logarithmic space. The training and validation data are used as inputs to generate an objective function (e.g., error rate on the validation set) for the Bayesian optimizer. In this work, Bayesian optimization is performed by minimizing of validation set's classification error (CE), defined by eqn 4.

$$CE = \frac{Overall\ number\ of\ the\ cases\ correctly\ classified}{Total\ no.of\ cases} X\ 100 \quad (4)$$

Final chosen CNN model using Bayesian optimized hyper-parameter values is trained to minimize loss function and then tested on the independent test set. In this work, cross entropy loss function [37], as defined by eqn. 5 is used.

$$loss = -\frac{1}{N}\sum_{n=1}^{N}\sum_{i=1}^{K} w_i t_{ni} ln y_{ni} \quad (5)$$

where,
N= Total samples
K=Total classes
W_i=weight for i^{th} class,
t_{ni}=indicator showing how nth sample is related to i^{th} class and y_{ni}= Probability that the network is associating n^{th} input to i^{th} class.

3.5 Implications of the proposed scheme

The adoption of Bayesian optimized Convolutional Neural Networks (CNNs) for fault detection in the field of power distribution and smart grid management can have several broader implications that positively impact the efficiency, reliability, and sustainability of power systems. Here are some key implications:

Improved Fault Detection Accuracy:
Bayesian optimization helps fine-tune hyperparameters, enhancing the performance of CNNs. This optimization can lead to improved fault detection accuracy, reducing false positives and negatives in the identification of faults within the power distribution network.
Enhanced Predictive Maintenance:
The Bayesian optimized CNN can contribute to more accurate predictions of equipment failures. By detecting subtle patterns indicative of potential faults, the system aids in implementing proactive and predictive maintenance strategies, minimizing downtime and reducing overall maintenance costs.
Increased Grid Reliability:
Accurate fault detection contributes to overall grid reliability. By identifying and addressing faults promptly, the system helps prevent cascading failures, ensuring a more resilient power distribution network.
Optimized Operation and Energy Efficiency:
A reliable fault detection system enables operators to make informed decisions, optimizing the operation of the power distribution network. Minimizing downtime and reducing energy losses associated with faults lead to improved energy efficiency.
Quick Response to Grid Events:
Bayesian optimization aids in tuning the CNN model for faster and more effective responses to grid events. This quick response is crucial in minimizing the impact of faults and ensuring a more resilient and adaptive smart grid.
Adaptability to Evolving Grid Architectures:
The adaptability of Bayesian optimized CNNs allows the fault detection system to evolve alongside changes in the power distribution infrastructure. This is particularly relevant in the context of smart grid technologies, distributed energy resources, and evolving grid architectures.
Facilitation of Grid Modernization:
By incorporating advanced machine learning techniques, the fault detection system contributes to the modernization of power grids. This aligns with the broader goal of creating intelligent, responsive, and self-healing grids capable of meeting the demands of modern energy systems.
Cost Reduction and Resource Optimization:
Proactive fault detection and maintenance optimization can lead to cost reductions by minimizing equipment failures and associated repair costs. Efficient resource allocation, guided by the fault detection system, helps optimize manpower and financial resources.
Data-Driven Decision-Making:
Bayesian optimization enhances the data-driven decision-making process. By leveraging historical and real-time data, the CNN model can adapt to changing conditions, facilitating more informed decisions for grid operators and managers.
Cybersecurity Considerations:
The integration of advanced fault detection systems should also include considerations for cybersecurity. Protecting the integrity and confidentiality of the data and model is critical to ensure the overall security of the smart grid.
Technology Leadership and Innovation:
Embracing state-of-the-art technologies like Bayesian optimized CNNs positions power distribution utilities and smart grid operators as technology leaders. This fosters a culture of innovation and readiness to harness the benefits of advanced analytics in the energy sector.

In summary, the broader implications of deploying Bayesian optimized CNNs for fault detection in power distribution and smart grid management extend beyond localized improvements. They contribute to the transformation of power systems into more intelligent, adaptive, and resilient infrastructures, aligning with the overarching goals of modernizing energy grids for the future

4. Simulation Results and Discussion

This section presents all the simulation procedures and results in details. All simulations are performed on a computer with Core i5-2.67 GHz CPU with 4 GB RAM. The suggested fault classification technique is tested on a PSCAD-modeled IEEE 13-Node test distribution grid as stated earlier. Eleven (11) different types of faults including all types of short circuit faults have been simulated. Faulty cases are also simulated by varying the locations (at 8 different nodes), fault resistances (7 different values) and inception angles (8 different values) as given in TABLE 2. There are total of 5568 cases for 11 types of faults consisting of 5120 [10*8(resistances)*8(nodes (locations))*8(inception angles)] cases for 10 different types of faults except ABC fault for which 448 cases [7(resistances)*8(nodes) *8(inception angles)] are considered (avoiding zero fault resistance in this case). The simulation run time of the distribution system model is 1.3 sec in which each fault is made to occur for a time interval from 0.2 sec to 1 seconds.

4.1 CNN model and Hyper-parameters

Time series of three phase fault currents are sampled (at a sampling frequency of 2000 samples per second) to get 2600 samples (out of which 400 samples are used for pre-fault, 1600 samples during the fault, and 600 for post-fault) for fault current of each phase.

Table 3. Description of optimized hyper parameters

Parameters	Ranges	Optimized Values
Section Depth	[1,3]	2
Initial Learning Rate	[0.001,1]	0.0113
Momentum	[0.8,0.98]	0.9225
L2Regularization	[1.0e-10, .01]	0.0022818

For each case samples of fault currents of each phase are placed in a row (one after the other) to get an input vector. Taking all the cases together gives input matrix for training, which is imported to MATLAB, where images having size of 40*65*3 are created for each row of this input matrix. The resolution of images created by CNN is 4.76 Megapixels. There is total 5568 images (one for each case and using 7800 samples) to be used as input for learning of CNN. Out of these, 4176 images (75%) are used as training input dataset, 749 images (≈ 13%) as validation and 643 images (≈ 12%) as testing input dataset for CNN. Output data for CNN model is class labels or names of 11 different types of faults [AG, BG, CG, ABG, BCG, CAG, AB, BC, CA ABCG, and ABC].

It uses nine convolutional layers, with batch normalization and ReLU operations occurring after each one. Sizes of each convolutional layer in first, second and third batch of consecutive three convolutional layers are 9*40*65(9 filters resulting in a 40*65 feature map), 18*20*33(18 filters resulting in a 20*33 feature map) and 36*10*17 (36 filters resulting in a 10*17 feature map) respectively as shown in Fig. 5. After the first two batches of consecutive three convolutional layers, a max-pooling procedure was performed resulting in output dimensions of 9*20*33 & 18*10*17 (halving both height and width) respectively. In contrast, an average pooling operation was performed after the last batch of three convolutional resulting in an output tensor with dimensions 36x3x10. In average pooling, for each local region in the input feature map, the average value is computed, and that value becomes the corresponding element in the output feature map. The number of channels remains the same, but the height and width dimensions may change based on the pooling operation parameters.

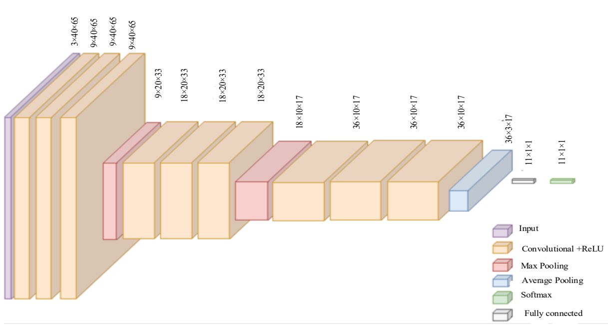

Fig. 5. CNN Model's architectural layout.

To get the 11 classification outputs, the final average pooling process was followed by a fully connected layer consisting of 11 neurons, then a softmax layer of size 11*1*1(here, each element represents the probability of the input belonging to a specific class in a classification task with 11 classes). The input image's dimensions are 3*40*65, with 3 denoting the image's depth. Structure of the CNN is obtained by extensive hit and trial simulations with a target to maximize the classification accuracy. Four hyperparameters of CNN, i.e., learning rate, momentum, regularization and network depth have been optimized by using Bayesian optimization approach. Ranges of these parameters are decided via extensive hit & trial simulations & the final optimized values are on the list in Table 3.

4.2. Fault Classification

Using optimized hyper parameters, the proposed CNN model has been trained. While training, piecewise learn rate schedule is used in which learning rate is updated every 40 epochs (i.e., learn rate drop period) by multiplying with 0.1(learn rate drop factor). Classification accuracy for test data is 99.87% for all types of faults as shown in confusion matrix of Fig. 6 and training progress of the model is shown in Fig. 7

True Class													
AG	51											100%	
BG		54										100%	
CG			53									100%	
ABG				63								100%	
BCG					64							100%	
CAG						53						100%	
AB							66					100%	
BC								59				100%	
CA									63			100%	
ABC										57		100%	
ABCG										1	59	98.3%	1.7%
	100%	100%	100%	100%	100%	100%	100%	100%	100%	98.3%	100%		
										1.7%			
Predicted Class	AG	BG	CG	ABG	BCG	CAG	AB	BC	CA	ABC	ABCG		

Fig. 6. Confusion Matrix for test data

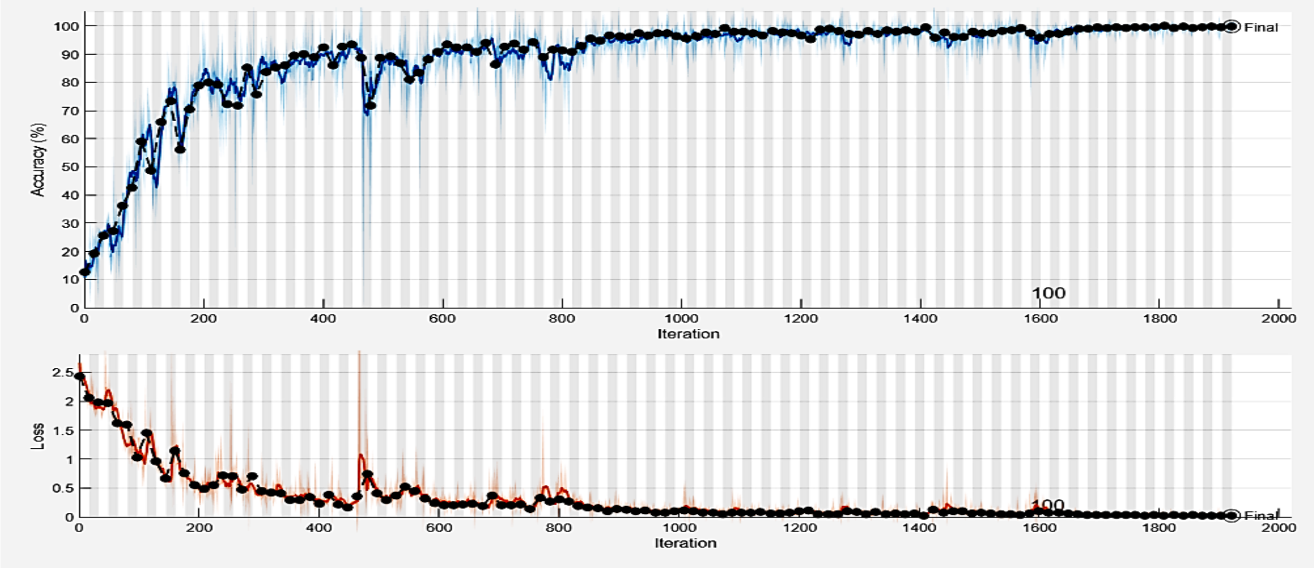

Fig. 7. Training Progress Graph of model CNN

As shown in Table 4. It is self-evident that the proposed strategy is capable of diagnosing the faults in a distribution grid with maximum accuracy and is better than the other techniques presented in the literature. Additionally, this method has shown positive results regardless of the fault resistance magnitude or fault inception angle measure and location of faulty node.

Table 4. Comparison of fault classification results

Reference no.	Year	Technique	Overall Accuracy (%)
Proposed Technique	**2023**	**Bayesian Optimized CNN**	**99.87%**
[25]	2023	EML (Extreme Learning Machine)	99.09%
[15]	2022	Robust Semi- Supervised Prototypical Network (RSSPN)	91.10%
[24]	2022	LeeNet-5	99.48%
[26]	2022	CatBoost Classifier	99.54%
[27]	2022	CNN-LSTM	97.5%
[28]	2021	CNN	99.66%
[16]	2021	KNN- Bayesian Method	97%
[17]	2021	Grid Search, Random Search- Bayesian Optimization	98.06%
[46]	2021	wavelet packet transform & SVM	98.8%
[29]	2018	PCA	99.54%
[1]	2018	Hybrid S-transform	99.47%
[12]	2016	Wavelet-based fuzzy logic algorithm	89.5%
[14]	2013	Wavelet-based ANFIS system	99.84%
[9]	2013	Time-time transform based ART Network	99.18%
[10]	2012	Wavelet-based Neural Network	98.4%

5. Conclusion & Discussion

A deep learning approach involving the use of a convolution neural network is presented in this paper. The hyper parameters of the CNN are tuned using Bayesian optimization and it is trained with training data to enable it to take decision for distinct faults. From the simulation results, it can be seen that the suggested approach has the ability to categorize eleven different types of faults for an 11 kV distribution grid with 99.87% accuracy (shunt capacitor bank for a variety of fault location, inception angle, and fault resistance conditions). The proposed technique's classification results are shown to be superior to many other strategies published in the literature. The proposed technique aligns with several industry trends, regulatory frameworks, and the evolving landscape in the following ways:

Industry Trends:

a. Advanced Analytics and Machine Learning: The use of Bayesian optimization and CNNs reflects the industry trend towards leveraging advanced analytics and machine learning for fault detection. These techniques enable automated, data-driven decision-making in real-time.
b. IoT and Sensor Integration:
The incorporation of Bayesian optimization with CNNs likely involves the utilization of data from sensors and the Internet of Things (IoT). This aligns with the trend of integrating sensor data for enhanced monitoring and fault detection capabilities.
c. Automation and Predictive Maintenance:
Proposed scheme aligns with the trends towards automation in optimizing model configurations and the broader move towards predictive maintenance in power distribution systems.

Regulatory Frameworks:

a. Reliability and Resilience Standards:
Proposed scheme aligns with regulatory standards that emphasize the reliability and resilience of power distribution networks. These frameworks often mandate the use of advanced technologies to enhance fault detection and response.
b. Data Security and Privacy Regulations:
As the fault detection system likely relies on data from various sources, adherence to data security and privacy regulations is crucial. Ensuring compliance with these regulations is aligned with the increasing emphasis on data protection in the regulatory landscape.

Evolving Landscape of Power Distribution:

a. Integration with Smart Grid Technologies: Proposed scheme can be integrated into smart grid architectures, aligning with the evolving landscape of power distribution. Smart grids leverage advanced technologies for efficient management, and fault detection systems play a vital role in ensuring grid stability.

b. Adaptability to Renewable Energy Integration: The evolving landscape involves increased integration of renewable energy sources. A fault detection system using Bayesian optimization and CNNs should be adaptable to handle challenges associated with the intermittent nature of renewable energy.

In summary, a Bayesian optimized CNN-based fault detection system aligns with industry trends by embracing advanced analytics, aligns with regulatory frameworks by emphasizing reliability and data security, and aligns with the evolving landscape by integrating with smart grid technologies and adapting to changes in the power distribution paradigm. Demonstrating the system's alignment with these aspects enhances its acceptance and effectiveness in the context of real-world power distribution applications

In future research, following limitations, potential challenges, and considerations associated with the real-world deployment, adaptability to diverse environments, and scalability of the proposed scheme need to be taken care of:

Limitations:

a. Data Quality and Quantity: The effectiveness of CNNs relies heavily on the availability and quality of labeled training data. In real-world scenarios, obtaining diverse and comprehensive fault data for training can be challenging.

b. Generalization Across Environments:
The model's generalizability across diverse power distribution environments may be limited. Factors such as variations in equipment types, configurations, and network structures could impact the model's performance.

c. Hyperparameter Tuning Overhead: Bayesian optimization involves tuning hyperparameters, which can be computationally expensive. This could pose challenges in real-time applications where quick responses are crucial.

d. Sensitivity to Hyperparameter: The performance of Bayesian optimization is sensitive to the choice of optimization parameters. Suboptimal choices may lead to inefficient model training.

e. Interpretability: CNNs are known for their complexity, and Bayesian optimization may further increase the model's opacity. Understanding and interpreting the decision-making process of such a model might be challenging.

Challenges in Real-World Deployment

a. Integration with Existing Systems: Integrating a new fault detection system with existing power distribution infrastructure can be complex. Compatibility issues, data format differences, and communication protocols need to be addressed.

b. Operational Impact: Implementing a fault detection system might disrupt ongoing operations. Minimizing downtime during deployment and ensuring a smooth transition are critical considerations.

c. Maintenance and Updating: Continuous maintenance and updates are necessary to keep the model effective over time. This includes adapting to changes in the power distribution network and ensuring the model stays relevant.

Adaptability to Diverse Environments:

a. Variability in Fault Types:
Power distribution systems exhibit a wide range of fault types. Ensuring the model's adaptability to diverse fault scenarios, including rare and unconventional faults, is essential.

b. Dynamic Network Topologies: Power distribution networks may change topology due to maintenance, repairs, or new installations. The model should be adaptable to dynamically changing network structure.

Scalability Considerations:

a. Computational Resources: Bayesian optimization and CNNs can be resource-intensive, especially for large-scale power distribution networks. Scalability requires careful consideration of computational requirements and potential hardware limitations.

b. Communication Overhead:
In a scalable system, the communication overhead between distributed components (sensors, computing units) must be managed efficiently to avoid delays in fault detection responses.

c. Real-Time Processing: As the size of the power distribution network increases, maintaining real-time processing capabilities becomes crucial. Ensuring low-latency responses for large-scale networks may be challenging.

Overall Considerations:

a. Cost Implications: The cost of implementing and maintaining the system, including hardware, software, and personnel training, should be carefully considered in the context of the overall budget and resource constraints.

b. Regulatory Compliance: Compliance with regulatory standards and cybersecurity measures is critical. Adhering to industry-specific regulations ensures the legal and secure deployment of the fault detection system.

c. Continuous Monitoring and Evaluation:
Continuous monitoring and evaluation of the model's performance in real-world conditions are necessary. Regular updates and improvements should be part of the deployment strategy.

It's essential to conduct thorough testing and validation in real-world conditions to ensure the model's robustness and effectiveness in diverse environments. In the future, sensitivity of the proposed technique can be investigated in the presence of distributed generators and/or noise and/or by varying the grid configuration. Calculation of accuracy of this algorithm on experimental data can also be done in the future. In future, the proposed scheme will be thoroughly tested and validated in real-world conditions to ensure its robustness and effectiveness in diverse environments.

References

[1] M. Shafiullah and M. A. Abido, "S-Transform Based FFNN Approach for Distribution Grids Fault Detection and Classification," *IEEE Access*, vol. 6, pp. 8080–8088, Oct. 2018, doi: 10.1109/ACCESS.2018.2809045.

[2] Wen-Hui Chen, Chih-Wen Liu, and Men-Shen Tsai, "On-line fault diagnosis of distribution substations using hybrid cause-effect network and fuzzy rule-based method," *IEEE Trans. Power Deliv.*, vol. 15, no. 2, pp. 710–717, Apr. 2000, doi: 10.1109/61.853009.

[3] J. C. S. deSouza, E. M. Meza, M. Th. Schilling, and M. B. D. C. Filho, "Alarm Processing in Electrical Power Systems Through a Neuro-Fuzzy Approach," *IEEE Trans. Power Deliv.*, vol. 19, no. 2, pp. 537–544, Apr. 2004, doi: 10.1109/TPWRD.2003.823205.

[4] D. Thukaram, H. P. Khincha, and H. P. Vijaynarasimha, "Artificial Neural Network and Support Vector Machine Approach for Locating Faults in Radial Distribution Systems," *IEEE Trans. Power Deliv.*, vol. 20, no. 2, pp. 710–721, Apr. 2005, doi: 10.1109/TPWRD.2005.844307.

[5] L. C. Acacio, P. A. Guaracy, T. O. Diniz, D. R. R. P. Araujo, and L. R. Araujo, "Evaluation of the impact of different neural network structure and data input on fault detection," in *2017 IEEE PES Innov. Smart Grid Technol. Conf.- Latin America (ISGT Latin America)*, Quito: IEEE, Sep. 2017, pp. 1–5. doi: 10.1109/ISGT-LA.2017.8126699.

[6] N. Wang, V. Aravinthan, and Y. Ding, "Feeder-level fault detection and classification with multiple sensors: A smart grid scenario," in *2014 IEEE Worksh. Stat. Sign. Proces. (SSP)*, Gold Coast, Australia: IEEE, Jun. 2014, pp. 37–40. doi: 10.1109/SSP.2014.6884569.

[7] Z. S. Hosseini, M. Mahoor, and A. Khodaei, "AMI-Enabled Distribution Network Line Outage Identification via Multi-Label SVM," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 5470–5472, Sep. 2018, doi: 10.1109/TSG.2018.2849845.

[8] Y. Aslan and Y. E. Yağan, "Artificial neural-network-based fault location for power distribution lines using the frequency spectra of fault data," *Electr. Eng.*, vol. 99, no. 1, pp. 301–311, Mar. 2017, doi: 10.1007/s00202-016-0428-8.

[9] I. Nikoofekr, M. Sarlak, and S. M. Shahrtash, "Detection and classification of high impedance faults in power distribution networks using ART neural networks," in *2013 21st Iranian Conf. Electr. Engin. (ICEE)*, Mashhad, Iran: IEEE, May 2013, pp. 1–6. doi: 10.1109/IranianCEE.2013.6599760.

[10] S. Jana, "Wavelet Entropy and Neural Network Based Fault Detection on A Non Radial Power System Network," *IOSR J. Electr. Electron. Eng.*, vol. 2, no. 3, pp. 26–31, 2012, doi: 10.9790/1676-0232631.

[11] M. Jamil, R. Singh, and S. K. Sharma, "Fault identification in electrical power distribution system using combined discrete wavelet transform and fuzzy logic," *J. Electr. Syst. Inf. Technol.*, vol. 2, no. 2, pp. 257–267, Sep. 2015, doi: 10.1016/j.jesit.2015.03.015.

[12] J. Klomjit and A. Ngaopitakkul, "Selection of proper input pattern in fuzzy logic algorithm for classifying the fault type in underground distribution system," in *2016 IEEE Region 10 Conf.(TENCON)*, Singapore: IEEE, Nov. 2016, pp. 2650–2655. doi: 10.1109/TENCON.2016.7848519.

[13] M. Dehghani, M. H. Khooban, and T. Niknam, "Fast fault detection and classification based on a combination of wavelet singular entropy theory and fuzzy logic in distribution lines in the presence of distributed generations," *Int. J. Electr. Power Energy Syst.*, vol. 78, pp. 455–462, Jun. 2016, doi: 10.1016/j.ijepes.2015.11.048.

[14] J. Zhang, Z. Y. He, S. Lin, Y. B. Zhang, and Q. Q. Qian, "An ANFIS-based fault classification approach in power distribution system," *Int. J. Electr. Power Energy Syst.*, vol. 49, pp. 243–252, Jul. 2013, doi: 10.1016/j.ijepes.2012.12.005.

[15] T. Zheng *et al.*, "RSSPN:Robust Semi-Supervised Prototypical Network for Fault Root Cause Classification in Power Distribution Systems," *IEEE Trans. Power Deliv.*, vol. 37, no. 4, pp. 3282–3290, Aug. 2022, doi: 10.1109/TPWRD.2021.3125704.

[16] Y. Liang, K.-J. Li, Z. Ma, and W.-J. Lee, "Multilabel Classification Model for Type Recognition of Single-Phase-to-Ground Fault Based on KNN-Bayesian Method," *IEEE Trans. Ind. Appl.*, vol. 57, no. 2, pp. 1294–1302, Mar. 2021, doi: 10.1109/TIA.2021.3049766.

[17] J. Zhu, L. Mu, D. Ma, and X. Zhang, "Faulty Line Identification Method Based on Bayesian Optimization for Distribution Network," *IEEE Access*, vol. 9, pp. 83175–83184, Apr. 2021, doi: 10.1109/ACCESS.2021.3087131.

[18] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.

[20] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," 2014, doi: 10.48550/ARXIV.1409.1556.

[21] C. Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, and Scott Reed, "Going deeper with convolutions," in *2015 IEEE Conf. Comp. Vis. Patt. Recogn. (CVPR)*, Boston, MA, USA: IEEE, Jun. 2015, pp. 1–9. doi: 10.1109/CVPR.2015.7298594.

[22] M. D. Zeiler and R. Fergus, "Visualizing and Understanding Convolutional Networks," 2013, doi: 10.48550/ARXIV.1311.2901.

[23] Y. LeCun, K. Kavukcuoglu, and C. Farabet, "Convolutional networks and applications in vision," in *Proceed. 2010 IEEE Intern. Symp. Circ. Sys.*, Paris, France: IEEE, May 2010, pp. 253–256. doi: 10.1109/ISCAS.2010.5537907.

[24] F. M. Shakiba, M. Shojaee, S. M. Azizi, and M. Zhou, "Transfer Learning for Fault Diagnosis of Transmission Lines," 2022, doi: 10.48550/ARXIV.2201.08018.

[25] Md. O. F. Goni., Md. Nahiduzzaman, Md. M. Rahman, Md. R. Islam, and Md. S. Anower, "Fast and Accurate Fault Detection and Classification in Transmission Lines using Extreme Learning Machine," *E-Prime - Adv. Electr. Eng. Electron. Energy*, vol. 3, p. 100107, Mar. 2023, doi: 10.1016/j.prime.2023.100107.

[26] V. N. Ogar, S. Hussain, and K. A. A. Gamage, "Transmission Line Fault Classification of Multi-Dataset Using CatBoost Classifier," *Signals*, vol. 3, no. 3, pp. 468–482, Jul. 2022, doi: 10.3390/signals3030027.

[27] A. Moradzadeh, H. Teimourzadeh, B. Mohammadi-Ivatloo, and K. Pourhossein, "Hybrid CNN-LSTM approaches for identification of type and locations of transmission line faults," *Int. J. Electr. Power Energy Syst.*, vol. 135, p. 107563, Feb. 2022, doi: 10.1016/j.ijepes.2021.107563.

[28] P. Rai, N. D. Londhe, and R. Raj, "Fault classification in power system distribution network integrated with distributed generators using CNN," *Electr. Power Syst. Res.*, vol. 192, p. 106914, Mar. 2021, doi: 10.1016/j.epsr.2020.106914.

[29] P. Chopra and S. K. Yadav, "PCA and feature correlation for fault detection and classification," in *2015 IEEE Rec. Adv. Intellig. Comput. Sys. (RAICS)*, Trivandrum, Kerala, India: IEEE, Dec. 2015, pp. 195–200. doi: 10.1109/RAICS.2015.7488413.

[30] S. Senemmar and J. Zhang, "Deep Learning-based Fault Detection, Classification, and Locating in Shipboard Power Systems," in *2021 IEEE Elect. Ship Technol. Sympos. (ESTS)*, Arlington, VA, USA: IEEE, Aug. 2021, pp. 1–6. doi: 10.1109/ESTS49166.2021.9512342.

[31] N. K. Chanda and Yong Fu, "ANN-based fault classification and location in MVDC shipboard power systems," in *2011 North American Pow. Sympos.*, Boston, MA, USA: IEEE, Aug. 2011, pp. 1–7. doi: 10.1109/NAPS.2011.6025172.

[32] W. Li, A. Monti, and F. Ponci, "Fault Detection and Classification in Medium Voltage DC Shipboard Power Systems with Wavelets and Artificial Neural Networks," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 11, pp. 2651–2665, Nov. 2014, doi: 10.1109/TIM.2014.2313035.

[33] Y. Ma, D. Oslebo, A. Maqsood, and K. Corzine, "DC Fault Detection and Pulsed Load Monitoring Using Wavelet Transform-Fed LSTM Autoencoders," *IEEE J. Emerg. Sel. Top. Power Electron.*, vol. 9, no. 6, pp. 7078–7087, Dec. 2021, doi: 10.1109/JESTPE.2020.3019382.

[34] Y. Ma, D. Oslebo, A. Maqsood, and K. Corzine, "Pulsed-Power Load Monitoring for an All-Electric Ship: Utilizing the Fourier Transform Data-Driven Deep Learning Approach," *IEEE Electrification Mag.*, vol. 9, no. 1, pp. 25–35, Mar. 2021, doi: 10.1109/MELE.2020.3047164.

[35] C. Thornton, F. Hutter, H. H. Hoos, and K. Leyton-Brown, "Auto-WEKA: Combined Selection and Hyperparameter Optimization of Classification Algorithms," 2012, doi: 10.48550/ARXIV.1208.3719.

[36] F. Hutter, H. H. Hoos, and K. Leyton-Brown, "Sequential Model-Based Optimization for General Algorithm Configuration," in *Learn, Intellig. Optimiz.*, vol. 6683, C. A. C. Coello, Ed., in Lecture Notes in Computer Science, vol. 6683. , Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 507–523. doi: 10.1007/978-3-642-25566-3_40.

[37] B. H. Shekar and G. Dagnew, "Grid Search-Based Hyperparameter Tuning and Classification of Microarray Cancer Data," in *2019 Sec. Inter. Conf. Adv. Comp. Commun. Parad. (ICACCP)*, Gangtok, India: IEEE, Feb. 2019, pp. 1–8. doi: 10.1109/ICACCP.2019.8882943.
[38] B. Bischl, M. Binder, M. Lang, T. Pielok, J. Richter, S. Coors, and J. Thomas, "Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges," *WIREs Data Min. Knowl. Discov.*, vol. 13, no. 2, pp. e1484, Mar. 2023, doi: 10.1002/widm.1484.
[39] P. Liashchynskyi and P. Liashchynskyi, "Grid Search, Random Search, Genetic Algorithms: A Big Comparison for NAS," 2019, doi: 10.48550/ARXIV.1912.06059.
[40] L. Villalobos-Arias, C. Quesada-López, J. Guevara-Coto, A. Martínez, and M. Jenkins, "Evaluating hyper-parameter tuning using random search in support vector machines for software effort estimation," in *Proc. 16th ACM Inter. Conf. Predict. Models Data Analytics Softw. Engin.*, Virtual USA: ACM, Nov. 2020, pp. 31–40. doi: 10.1145/3416508.3417121.
[41] R. Andonie and A.-C. Florea, "Weighted Random Search for CNN Hyperparameter Optimization," 2020, doi: 10.48550/ARXIV.2003.13300.
[42] E. Elgeldawi, A. Sayed, A. R. Galal, and A. M. Zaki, "Hyperparameter Tuning for Machine Learning Algorithms Used for Arabic Sentiment Analysis," *Informatics*, vol. 8, no. 4, p. 79, Nov. 2021, doi: 10.3390/informatics8040079.
[43] S. Shekhar, A. Bansode, and A. Salim, "A Comparative Study of Hyper-Parameter Optimization Tools," in *2021 IEEE Asia-Pacific Conf. Comp. Scie. Data Engin. (CSDE)*, Brisbane, Australia: IEEE, Dec. 2021, pp. 1–6. doi: 10.1109/CSDE53843.2021.9718485.
[44] T. T. Joy, S. Rana, S. Gupta, and S. Venkatesh, "Fast hyperparameter tuning using Bayesian optimization with directional derivatives," *Knowl.-Based Syst.*, vol. 205, p. 106247, Oct. 2020, doi: 10.1016/j.knosys.2020.106247.
[45] S. J. Pan and Q. Yang, "A Survey on Transfer Learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010, doi: 10.1109/TKDE.2009.191.
[46] K. Moloi, M. Ntombela, Thapelo. C. Mosetlhe, Temitope. R. Ayodele, and Adedayo. A. Yusuff, "Feature Extraction based Technique for Fault Classification in Power Distribution System," in *2021 IEEE PES/IAS Power Africa*, Nairobi, Kenya: IEEE, Aug. 2021, pp. 1–5. doi: 10.1109/PowerAfrica52236.2021.9543314.
[47] A. Брагин, A. Bragin, B. Спицын, and V. Spicyn, "Electroencephalogram Analysis Based on Gramian Angular Field Transformation," in *GraphiCon'2019 Proc. Volume 2*, Bryansk State Technical University, Nov. 2019, pp. 273–275. doi: 10.30987/graphicon-2019-2-273-275.