



An effective data-driven machine learning hybrid approach for fault detection and classification in a standalone low-voltage DC microgrid

Anindita Deb¹ · Arvind Kumar Jain²

Received: 16 December 2023 / Accepted: 27 February 2024
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract

DC microgrids are gaining more importance in maritime, aerospace, telecom, and isolated power plants for heightened reliability, efficiency, and control. Yet, designing a protective system for DC microgrids is challenging due to novelty and limited literature. Recent interest emphasizes standalone fault detection and classification, especially through data-driven machine-learning approaches. However, the emphasis remains on progressing state-of-the-art tools for fault diagnosis in DC microgrids. Therefore, this work emphasizes fault detection and classification in a low-voltage standalone DC microgrid using a data-driven machine learning hybrid approach: bagged ensemble learner and cosine k-nearest neighbour (C-kNN) algorithms. The proposed fault detection and classification scheme makes the use of local voltage and current measurements which enhances the admissibility of the proposed scheme. The bagged ensemble learner accurately identifies the faults in the line, whereas the cosine k-nearest neighbor classifies the fault as pole to ground or pole to pole for further corrective actions. A diverse set of test scenarios encompassing faulty and normal conditions has been analyzed and validated by randomizing data inputs. The test model comprising PV, battery source, and loads have been constructed in MATLAB/Simulink environment. The proposed scheme promises accurate fault identification and classification in normal and noisy environments. To establish the robustness of the proposed approach, the outcomes of the fault detection and classification scheme have been compared with the methods reported in the literature. The results indicate that the proposed method outperformed in comparison to existing methods.

Keywords Fault detection and classification · Bagged Ensemble learner · Cosine k-nearest neighbor · Low voltage DC MG · Confusion matrix

Abbreviations

DC	Direct current
EL	Ensemble learner
LV	Low voltage
MG	Microgrid
PP	Pole-to-pole
DG	Distributed generation
ML	Machine learning

PV	Photovoltaic
BEL	Bagged ensemble learner
C-kNN	Cosine k-nearest neighbor
PG	Pole-to-ground
2PG	Two pole-to-ground
AI	Artificial intelligence
HIF	High-impedance faults

✉ Anindita Deb
aninditae.sch@nita.ac.in

Arvind Kumar Jain
arvindj@nita.ac.in

¹ Department of Electrical Engineering, NIT Agartala, Jirania, Tripura 799046, India

² Department of Electrical Engineering, Faculty of Engineering, NIT Agartala, Jirania, Tripura 799046, India

1 Introduction

The penetration of renewable energy sources, DC electrical loads, and energy storage systems has sparked a growing interest in DC MGs, placing them as a competitive alternative to conventional AC MGs [1]. The integration of modern power electronics and sophisticated control algorithms has empowered the efficient deployment and reliable operation of

DC MGs [2, 3]. When compared to AC systems, DC systems have several benefits, such as higher dependability, better power quality, increased transmission capacity, and simplified control methods [4]. Despite this, protecting DC systems is more difficult than usual because of the peculiarities of DC. Quick and fast protective measures are required because of the swift transients and the lack of zero-crossings [5].

Predominantly, the most frequently encountered faults in DC MGs can be categorized as PP, PG, and 2PG faults. PP faults are less frequent, but they have the potential to produce extremely high currents that could harm equipment. HIFs, on the other hand, are difficult to identify since they tend to cause very little current variation, which makes them immune to standard preventative measures [6]. HIFs, if left unaddressed, can lead to circumstances that push the system to more severe faults like PP, PG, or 2PG faults. Classical fault detection techniques are not appropriate to be implemented in an MG that supports DG simply because they rely on predefined thresholds to identify faults. The prime reasons are but are not limited to: (i) the variability of DG output, which can cause voltage and current fluctuations, posing challenges in setting fixed thresholds for fault detection, (ii) the dynamic behavior of islanded Microgrids may cause a delay or imprecision in fault detection; (iii) Predefined thresholds may not be able to accurately represent the nonlinear behaviour of the system, which reduces the efficacy of threshold-based fault detection techniques [7]. Differential protection is another conventional distribution system protection method that has been proven applicable in an MG [8]; however, it is not suitable due to the difficulty of determining a multi-terminal protection zone with several inputs [9, 10]. Overcurrent protection, the fundamental form of protection, can fail to recognize a fault in an MG environment, whether it is island-based or grid-connected, due to tripping falsely or failing to trip due to a fault current level change. Moreover, the need for a communication link to provide selectivity, the dependency of the current derivative on line loading, cable length, and noise generated due to high sampling rates are some limitations of an overcurrent-based protection strategy [11]. Since the current derivative magnitude during faults is significant in DC, it simplifies the detection of faults in a DC MG [12]. Nonetheless, the protective measure may not work if the analytical model is incorrect.

DC systems are categorized as high, medium, and low voltage (HVDC, MVDC, and LVDC) depending on the power levels at which they operate. Since signal system-based protection techniques could potentially be used to safeguard DC MG systems in smart grid environments, they are regarded as cutting-edge techniques and are becoming more and more popular. These techniques include FFT, WT, S-transforms (ST), and Hilbert-Huang transforms (HHT) which offer more detailed observation than STFT when it comes to extracting the characteristics of faulty segments in

terms of signal features [14]. The primary limitation of the FFT is the insufficient time domain data. The frequency resolution of STFT is significantly influenced by the size of the window. Given that the STFT window size is constant, low frequencies will be poorly depicted by a small window and short disturbances will be difficult for a long window to detect. By utilizing WT with variable frequency resolution and ST with the time and frequency-dependent window function, this issue is resolved. WT can identify transient signals and abrupt changes in current wave patterns in the event of a short circuit [6, 15]. The main drawback of this strategy is that high sampling rate data acquisition components are required. HHT is appropriate for analyzing both non-stationary and nonlinear signals since it employs an adaptive basis function. However, when intermittent waves occur at a lower-frequency signal, mode mixing in the Empirical Mode Decomposition (EMD) section of HHT fosters a challenge [16]. In summary, the following are the main drawbacks of a fault detection and classification method based on signal processing: Wide-area protection necessitates the use of a global positioning system (GPS) and a high sampling rate data acquisition system [17].

In recent years, intelligent classifiers have emerged as an effective tool for fault detection. As classifiers, decision trees, support vector machines, naive Bayes, and artificial neural networks (ANN) are commonly used. AI is extensively used in the design of various protection schemes and power system controls in an MG environment because of its ability to reduce errors, learn from experience, and make decisions more quickly [18]. Instances of the direct utilization of intelligent classifiers for fault detection have been documented in the literature. The inputs of an artificial neural network (ANN) consist of voltage and current time signals, and the outputs are binary variables that indicate the direction and detection status of the fault. Typically, the time signal is first transformed through field transform-based techniques, after which features are extracted and supplied to the intelligent classifier. An improved approach based on wavelet transform (WT) and multiresolution analysis (MRA) for better characterization of the fault and input signal to the ANN has been suggested [19, 20]. AI and ML have generally improved the performance of fault diagnosis for all types of networks. A few ML- and AI-related pieces of literature on fault detection have been discussed as follows: For fault identification and classification, a support vector machine (SVM)-based machine learning technique has been presented in [21]. The technique trains the SVM classifier for both islanded and grid-connected types of MG operation using WT of the three-phase current as well as voltage signals. A similar work has been presented in [22], where discrete wavelet transforms (DWT) have been used in conjunction with support vector machines (SVM) for MG fault detection. To detect MG system faults (DWT) furnish features to Bayes classifiers

and KNN, decision tree (DT)/random forest (RF) [23, 24]. In contrast to other transients, the classifiers demonstrated the ability to identify HIFs. A discrete WT and Taguchi-ANN-based protection strategy for the fault analysis has been reported in [25], which has also been verified on the Opal RT platform. An alternative approach for locating and detecting faults in multi-terminal DC MGs that makes use of local voltage attributes has been explained in [26]. Short-circuit fault diagnosis using ANN in an LVDC ring-type test system has been reported in [27], which has been simulated in a PSCAD/EMTDC environment. A similar study on the LVDC network has been reported in [28, 29] using ANN and wavelet transform techniques. The main obstacles limiting the ANN-based method to the online application in a DC network are longer training times and the requirement to redesign the entire procedure by altering every system. Another study explains how a particular ML tool based on DT and WKNN for fault identification and classification operates. New transient scenarios have subsequently been used to test trained classification models to confirm their accuracy in fault detection [13]. In summary, machine learning approaches are well suited for fault diagnosis in microgrids due to their capacity to handle the complexity, variability, and dynamicity found in microgrid data as well as their adaptability, feature extraction skills, and scalability.

To prove the robustness of the methods proposed in literature against noisy signals few case studies have been included in this paper. A PG fault is created at 0.5 s considering a fault resistance of 2Ω in a PV & battery sourced 30 kW, 350 V DC MG [30]. The proposed centralized protection scheme to differentiate between internal and external faults has been proven to be immune to noisy signals. Another study [13] has been reported here. In the presence of the white Gaussian noise, a fault current distortion spectrum at the natural frequency of a 380 V DC MG network has been observed. It has been shown that the current measured at the fault occurrence instances did not have any impact of noise, i.e., immune to noise in the system. Furthermore, to evaluate the robustness of the analysed model, a fault diagnostic model immune to noise has been built in [31, 32] by adding a specific amount of specified noise to the data samples. The outcomes demonstrate how effectively the suggested fault detection and classification method can identify MG vulnerabilities. Moreover, [33, 34] also pinpoints the efficient use of ML techniques in identifying and classifying faults. From the literature survey, it is observed that various techniques for fault detection and classification have been discussed by the researchers. However, the following limitations and drawbacks have been observed:

1. The accuracy and reliability of the existing methods are significantly impacted by variations in fault conditions, topologies, and diverse DG units.

2. The sporadic and unpredictable characteristics of DGs impede the effectiveness of relaying schemes, potentially leading to delayed responses and unwanted tripping [45].
3. The existing overcurrent schemes encounter challenges when dealing with low fault currents in scenarios of high-resistance faults and during the islanding mode of operation. In such instances, the relay may either fail to operate or exhibit an increased response time.
4. The fault magnitude in microgrids experiences considerable variation depending on the types of faults and the operating conditions of the microgrid. Therefore, the protection system needs to rapidly detect and pinpoint faults under diverse scenarios.
5. Additionally, many of these approaches come with considerable initial costs and operational intricacies. Furthermore, some of these methods exhibit low accuracy, diminishing the overall efficiency of the scheme.

Given the above facts, the prime contributions of this research work are.

1. To develop a voltage & current-assisted ML approach comprised of the BEL technique and the cosine KNN algorithm for the detection and classification of both PG and PP faults under a wide range of operating conditions, in both normal and noisy environments. Due to adaptability, resilience, and capacity to lessen overfitting, BEL has been used to enhance the accuracy of fault detection and classification [37–39]. One of the most basic, straightforward, and efficient ML techniques for data classification is the KNN algorithm [40]. Recent studies have demonstrated, the effectiveness of the KNN algorithm in identifying & classifying DC microgrid faults [13, 40–42]. Further, BEL has been used as a primary layer and cosine-KNN, a non-parametric algorithm that adapts to changes in data over time, has been employed as the second layer for fault detection and classification. The novel aspect of the suggested bi-layered method for fault identification and classification is that it makes use of BEL and Cosine kNN models, which have been tailored for fault identification and classification in DC microgrids. The proposed strategy is appropriate for DC microgrid systems where fault patterns may alter or new fault categories may appear.
2. Fault detection and classification have been addressed using only local voltage and current measurements, eliminating the requirement for additional communication channels.
3. A comparative study with existing ML approaches for fault detection & classification has also been incorporated in this work.

The research paper is structured as follows: in Sect. 2, the methodology of the proposed fault detection and classification strategy has been explained. Section 3 explains the

architecture of the standalone DC microgrid system encompassing the collection and discussion of simulated results. The conclusions and implications made from the proposed study are finally presented in Sect. 4.

2 Methodology

This section outlines the procedure for the proposed detection and classification of faults in standalone DC microgrids. The MATLAB/SIMULINK platform is utilized to simulate the test system, and local voltage and current samples are gathered from the workspace in datasets. Subsequently, each of these datasets is trained independently using the primary layer, BEL classifier for fault detection, and the secondary layer, Cosine KNN classifier for fault classification. These datasets are then individually trained using the BEL classifier for fault detection (primary layer) and the Cosine KNN classifier for fault classification (secondary layer). In the event of a fault being detected by the primary layer, the secondary layer of the proposed technique categorizes it as either a PP or PG fault. Random voltage and current data samples have been employed for testing to assess the efficacy of the suggested bi-layered system. Noisy data samples have also been considered for testing to ensure the robustness of the proposed approach. The sampling rate of 1 kHz has been considered here.

BEL [46] & cosine-KNN [50, 51] have been used as two layers of the proposed data-driven ML hybrid approach for fault identification & categorization scheme for the standalone 380 V LV DC MG testbed.

The flowchart for the proposed data-driven ML fault detection and classification has been shown in Fig. 1 and is explained below:

Step 1 Simulate the proposed 380 V LV DC MG testbed and generate the local voltage and current data samples from the network with and without noise inclusion.

Step 2 Obtain the provided data samples from the workspace and utilize them to train the BEL and cosine-KNN separately. Note that the classifiers have been trained using 0 dB, 30 dB, 60 dB, and 90 dB, respectively [52, 53]. The selected classifier models have been illustrated as follows:

2.1 Ensemble classifier

The primary fault detection layer in the presented data-driven ML hybrid approach is a BEL. One of the three primary ensemble learning techniques, bagging, is particularly useful for high-dimensional data because it can effectively handle problems like missing values and prevent generalization to new datasets [46]. Bootstrap aggregation, or bagging, is an ensemble learning method introduced by Leo Breiman in

1996 [47]. It creates diverse ensemble members by resampling the training data and generating multiple subsets with replacements. Every subset trains a different model, and the average or voting mechanism among the individual forecasts is frequently used to arrive at the final prediction. Bagging's main advantages are the elimination of all variation and the decrease in model overfitting since it produces a large number of classifiers with fixed bias and averages their outputs. When working with sparse data, the objective of employing an ensemble model in fault classification techniques is to improve accuracy. The basic steps of bagging are:

Bootstrapping This method allows for the repetition of instances, enhancing diversity within subsets and improving the robustness of the ensemble learning model.

Parallel training Weak or basic learners are used to separately and concurrently train these bootstrapped data.

Aggregation Depending on the task (regression or classification), predictions are averaged for regression (soft voting), or the majority class is accepted for classification (hard voting/majority voting). This process yields a more accurate estimate by combining the individual classifiers' outputs.

Consider a dataset A with N samples, and we want to create an ensemble of bagged decision trees [48, 49]. The process of Bagging is summarised below:

Bootstrap sampling Generate B bootstrap samples from the original dataset: A_1, A_2, \dots, A_B

Train decision trees Train a decision tree on each bootstrap sample independently: T_1, T_2, \dots, T_B .

Consider,

A_b is the b -th bootstrap sample, $b = 1, 2, 3, \dots, B$.

T_b is the decision tree trained on A_b .

Aggregation Now, combining predictions of individual trees to make the final ensemble prediction, we get-

$$\hat{y} = \frac{1}{B} \sum_{i=1}^B \hat{y}_{T_i}$$

where, \hat{y} indicates the ensemble prediction, & \hat{y}_{T_i} indicates the prediction of the i -th tree.

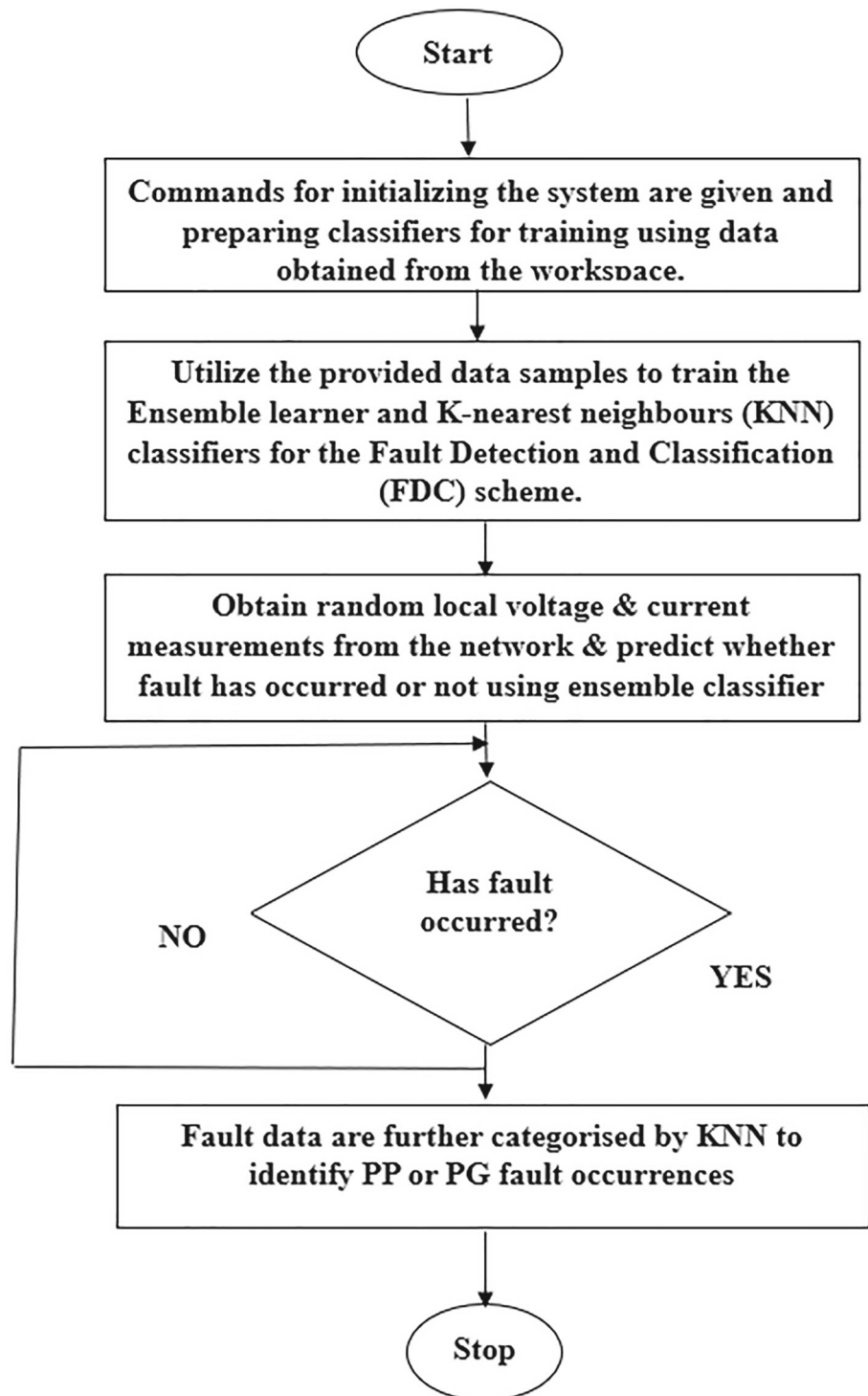
Final Prediction The ensemble prediction, \hat{y} , is the output of the bagged tree ensemble model.

2.2 Cosine KNN classifier

In the initial scenario of DC MG fault identification, the cosine KNN technique has been employed to classify faulty conditions in the MG. The approach involves training the classifier using a diverse range of fault instances including PP & PG faults to enable

accurate fault classification. Before delving into the fault classifying process the basic principles are explained. Among the classifiers discussed in the existing literature, KNN is

Fig. 1 Flowchart of the proposed fault detection and classification strategy



a commonly used and straightforward tool [50, 51]. KNN can be applied to various classification tasks where prior knowledge of the data distribution is unavailable. Cosine-KNN is a similarity-based algorithm used for classification or recommendation tasks. It calculates the cosine of the angle across feature vectors to determine the degree of resemblance between data points. It uses the cosine distance metric for

evaluation. The process of cosine KNN for classifying a new observation can be succinctly summarized in these standard steps [54]:

- (i) *Cosine Similarity* The ratio of the dot product of

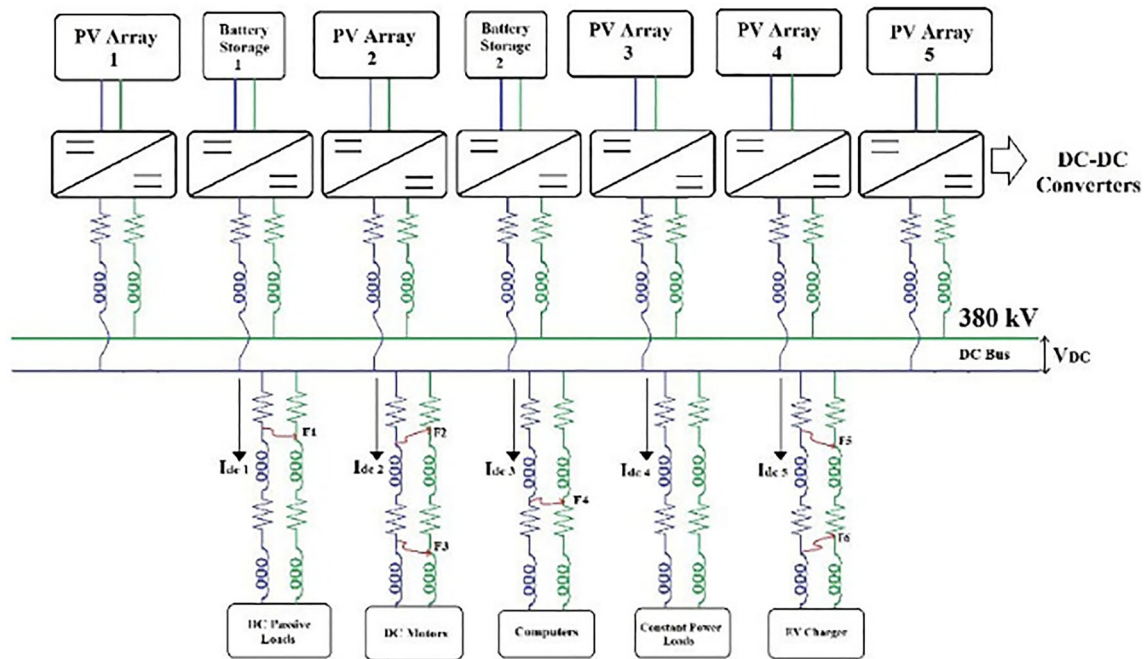


Fig. 2 Constructed standalone 380 V LV DC microgrid testbed

the vectors and the product of their magnitudes yields the cosine similarity between vectors, A and B :
 Cosine Similarity (A, B) = $\frac{A \cdot B}{\|A \cdot B\|}$.

- (ii) *Cosine Distance*: The cosine distance is complementary to the cosine similarity and is defined as $1 - \text{Cosine Similarity}$.

$$\text{Cosine Distance} = 1 - \text{Cosine Similarity}(A, B)$$

When the cosine similarity coefficient between two vectors is higher, it indicates a strong similarity between them.

Step 3 Obtain random current and voltage data samples from the network considering 30 dB, 60 dB, 90 dB SNR and test the accuracy of the bi-layered trained classifiers.

The approach proposed in this paper can be utilized in different DC MG configurations as well owing to its dependability on only local voltage and current data samples.

The steps demonstrating the application of the suggested method in real grids are as follows:

1. *Data acquisition* Throughout the grid infrastructure, multiple sensors, meters, and monitoring devices can be used for this purpose. This will gather the local voltage and current information from the respective areas.
2. *Preprocessing of data* To handle missing data, eliminate data distortions, and normalize the features, data pre-processing will be carried out.
3. *Classifier model training* Using historical information, the BEL and Cosine-KNN models will be trained to identify

patterns linked to different circumstances, faulty/ non-faulty, PP/PG faults.

4. *Model validation* The validation strategies will be applied to assess how well the trained models work. Models' performance in detecting and classifying problems will be evaluated using metrics including accuracy, receiver operating characteristic (ROC) curves, and parallel coordination plots (PCP).

5. *Model implementation* The developed model will be implemented through Supervisory Control and Data Acquisition (SCADA), Advanced Metering Infrastructure (AMI), and Distributed energy resources management system (DERMS) for real-time grid monitoring, enabling automatic fault identification, classification, and response.

6. *Continuous monitoring and upgrading* The performance of the bi-layered scheme needs to be tracked continuously, gather feedback from the grid, and utilize it to regularly revise and retrain the models to increase the accuracy and reliability of fault detection and classification strategy.

7. *Integration with grid management* The proposed fault detection and classification technique will be integrated with grid operations and decision-making processes to provide alerts, notifications, and recommendations to enable prompt reaction and corrective measures if faults are found.

Table 1 Specifications and ratings for the analyzed LV DC microgrid configuration

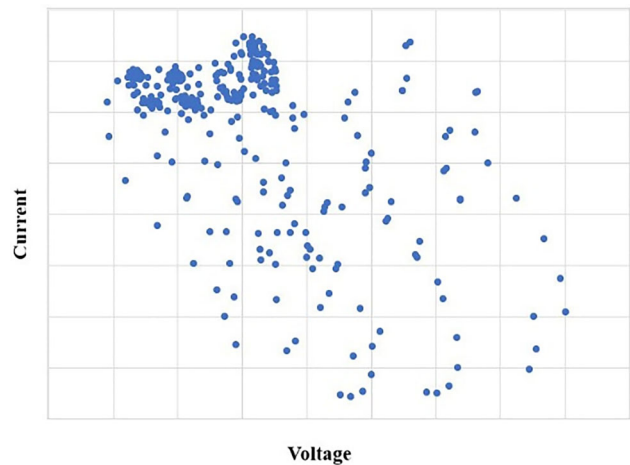
Equipment	Specifications
PV Panel (1 Soltech 1 STH-215-P)	$V_{mp} = 29$ V, $I_{SC} = 7.84$ A, $W_{max} = 213.5$ W
Battery (Li-ion)	Nominal voltage = 96 V, Rated Capacity (Ah) = 92.2, SOC _{initial} = 60%
DC-DC converter	380 V
Passive loads	14 kilowatts, 380 V
DC motor	6 kilowatts, 240 V
Constant power loads	6 kilowatts, 240 V
Computer loads	4 kilowatts, 240 V
EV Charger	48 V, Rated capacity (Ah) = 26.7, SOC _{initial} = 60%
Line impedance(pi-section)	$33 + j310$ m Ω

3 Test system, simulations, results, and discussions

3.1 The constructed standalone LV DC microgrid framework

The LV DC MG test bed shown in Fig. 2 is used to evaluate the proposed machine learning-driven fault detection and classification approach. A 380 V standalone LV DC MG is comprised of two auxiliary batteries, five solar PV systems & various loads, such as motor, passive loads, EV charger, constant power load, and constant current load. Every component of the test bed is linked to a single DC bus through DC-DC converters. The MPPT controller is used for monitoring and extracting the maximum power from solar PVs. During the power deficit, the state of charge-controlled auxiliary batteries supplies the power to the loads. The rating of each component of the test bed is given in Table 1. MATLAB/Simulink software has been utilized for developing the model depicted in Fig. 2. To train the BEL, the standalone LV DC MG has been simulated to produce a variety of faulty and normal conditions, and the resulting voltage and current values have been recorded. It is crucial for ML-based algorithms that the developed scheme undergoes verification across diverse environments and scenarios to ensure high reliability and robustness.

Therefore, in this work data with and without noise has been used to train and test the proposed algorithm. 2102 data samples of fault & non-fault situations have been utilized for training of the BEL and hold-out validation (25% held-out) is carried out to examine the predictive accuracy of the model. 1402 data samples of PG and PP faults have been utilised to train the cosine KNN where cross-validation has been used to get a good estimate of the predictive accuracy of the

**Fig. 3** Scattered data plot of current vs voltage with noise inclusion in the system

final model trained with all the data. Moreover, 901 random data samples of local voltage & current measurements have been used to test the individual classifiers' performances. To ensure the robustness of the suggested method, a white Gaussian noise of 30 dB, 60 dB, and 90 dB has been allowed to infiltrate the network response of the constructed testbed. The suggested approach has demonstrated resilience even in the presence of signal noise. This resilience is attributed to the immunity of the magnitudes of the measured signals, utilized for training the data-driven ML algorithms, to the introduced noise. Figure 3 shows the data scatter obtained from the proposed model including noise in the test system which is then trained using data-driven ML algorithms for fault identification and categorisation scheme. The plot is useful in visualizing the relationship between current and voltage variables in the dataset. In the later section, it has been proven that the current & voltage measurement at the fault occurring time instants have a negligible impact on the noise, and the measured signals with the noisy signals can also be considered as the inputs of the proposed data-driven ML hybrid technique.

3.2 Assessment of classifiers' performance during training

For the proposed work only two common input parameters i.e., voltage and current are required [13]. The primary and secondary layer of the proposed fault identification and categorisation scheme gives 100% training accuracy with BEL and Cosine KNN. The ratio of accurately predicted occurrences to actual occurrences in the training set determines the accuracy of the training. In this work, both the classifiers' performance has been verified with confusion matrix (CM), receiver operation characteristics (ROC) curve, and parallel coordinates plot (PCP).

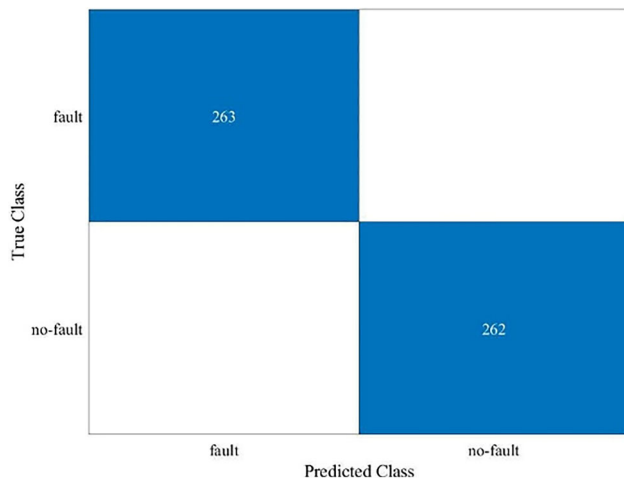


Fig. 4 CM for trained BEL scheme

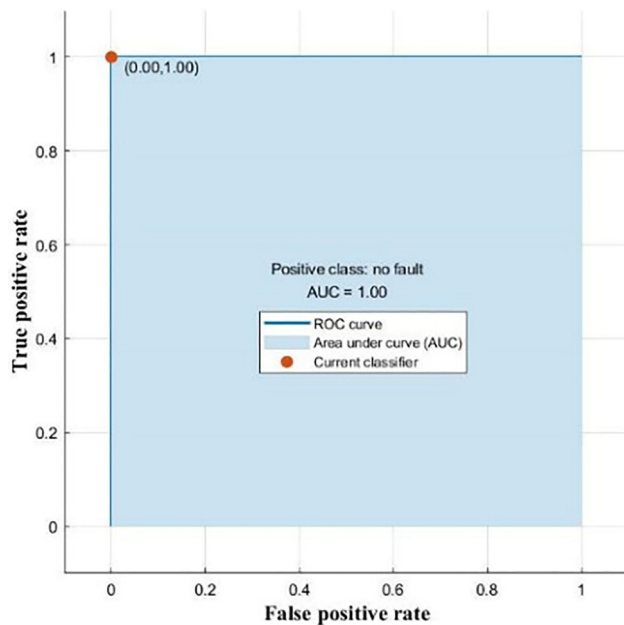


Fig. 5 ROC for trained BEL fault detection strategy

3.2.1 Training results of bagged ensemble learner without noise

The decision tree learner type has been implemented with 30 base/weak learners for the Bagged tree ensemble learner. Figures 4, 5 and 6 show the CM, ROC, and PCP characteristics, respectively, for the training of the Bagged tree ensemble learner. From the training results, it has been shown that the proposed ensemble model offers 100% detection accuracy and takes 0.256 s for training. From Fig. 4 it can be seen that the mislabelled boxes (off-diagonal) in CM are empty which depicts the full accuracy. The numerical expression

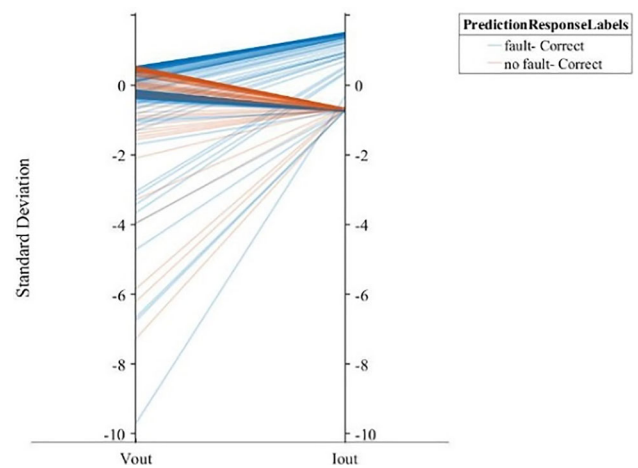


Fig. 6 PCP for trained BEL fault detection strategy

for calculating training accuracy is:

$$\frac{\text{Total number of instances in the training dataset}}{\text{Number of correctly classified instances}} \times 100\%$$

The ROC curve shown in Fig. 5 confirms the highest accuracy of the classifier model as the Area Under the Curve (AUC) shows 1. A two-dimensional representation of a multifaceted dataset can be shown with the help of Fig. 6's Parallel Coordinates Plot (PCP). The two variables in this instance are represented by the vertical lines: voltage and current. There are no lines indicating incorrect predictions.

3.2.2 Training results of bagged ensemble learner with noise

To ensure the robustness of the data-driven ML identification and categorization scheme, the models have been trained with the noisy data, and the white Gaussian noise of different signal-to-noise ratios (SNR) has been added to train the classifiers. SNRs of 30 dB 60 dB 90 dB have been included and simulated for the constructed LV DC microgrid. Figures 7, 8 and 9 show the performances of the ensemble classifier with SNRs of 30 dB 60 dB and 90 dB, respectively. The figures have shown that the performance of the BEL-trained classifier model does not get much affected even with noisy data. Furthermore, the results given in Table 2 show that the deviation in the training accuracy after introducing noise is also minimal.

3.2.3 Training results of cosine KNN with and without noise

For the PP or PG fault classification, the cosine KNN classifier has been trained which provides 100% accuracy in fault classification within a training time of 1.507 s. Cosine distance metric function with equal distance weights has been

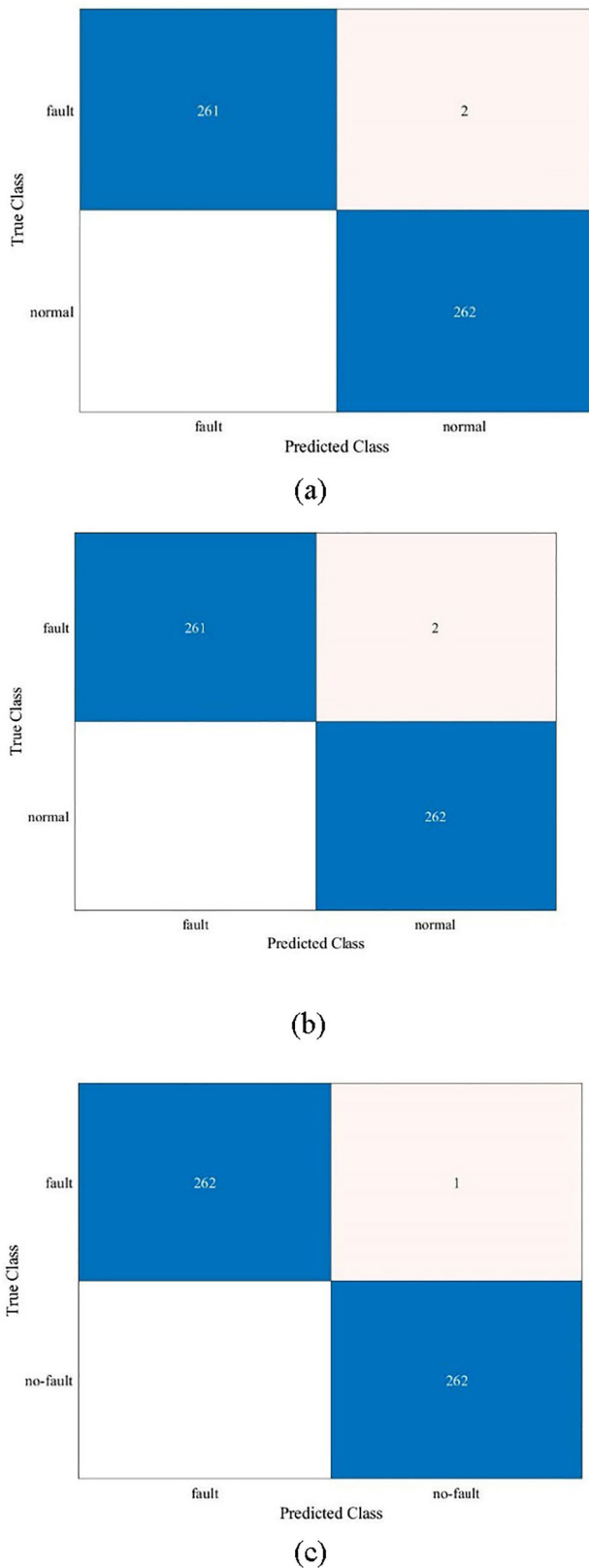


Fig. 7 CM of the trained BEL classification learner using a 30 dB b 60 dB c 90 dB

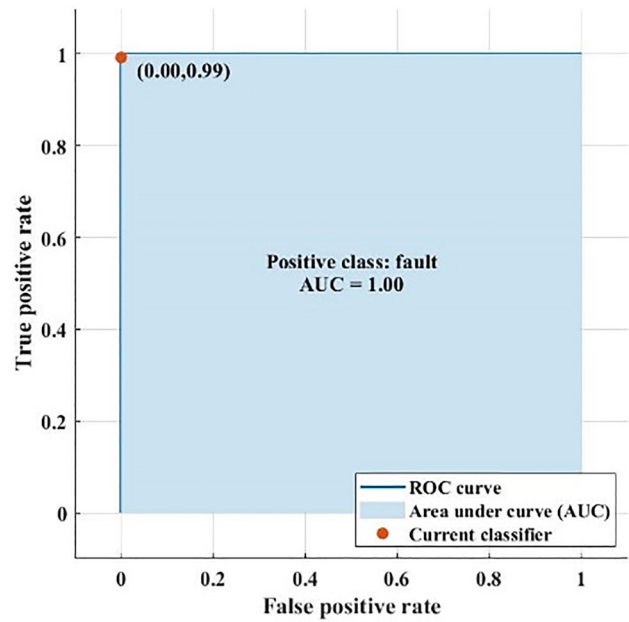


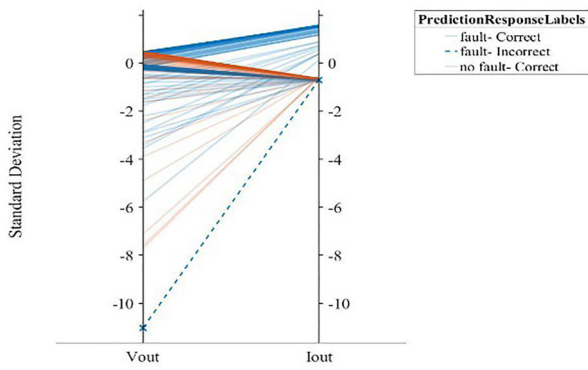
Fig. 8 ROC for the fault detection strategy with trained BEL considering noise

considered. In this scenario number of nearest neighbors equals 10. The average training accuracy of the second layer of the fault detection and classification scheme, i.e., classifying the fault as PP or PG fault using a Cosine KNN classifier without noise inclusion is found to be 100%. Further, Figs. 10, 11 and 12 show the CM, ROC & PCP of the trained-cross validated classifier. Five folds of cross-validation have been implemented here to prevent data overfitting. Figure 13 presents the performance of the trained classifier considering SNR of 30 dB 60 dB 90 dB, respectively.

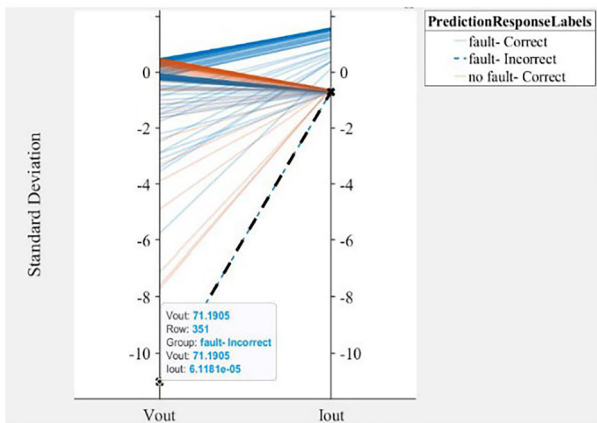
From Fig. 13, it is clear that miss-classification is minimal in all the cases. Table 2 summarises the training accuracies of the proposed data-driven ML fault detection and classification classifiers with 0 dB, 30 dB, 60 dB, and 90 dB noise inclusion. From the training results, it has been seen that the training accuracies of both the proposed models lie between 99.6 and 100%.

3.3 Assessment of classifiers' performance during testing

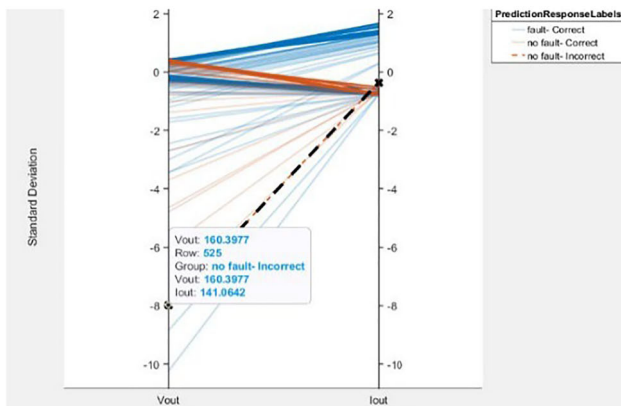
To test the performance of the trained classifier 901 random and shuffled data samples of current and voltage have been recorded as test data. Test results indicate that the proposed fault detection scheme has 100% accuracy in identifying “normal/no-fault” & “fault” conditions and further categorizing it as PP & PG fault. In addition, to complement the robustness and resilience of the proposed scheme further, the



(a)



(b)



(c)

Fig. 9 PCP of the trained BEL classification learner using a 30 dB b 60 dB c 90 dB

Table 2 Performance of the data-driven ML fault detection and classification technique with and without noise inclusion in training

SNR	Training accuracy (%)	
	Bagged tree ensemble learner	Cosine KNN
Noise-free	100	100
30 dB	99.6	99.9
60 dB	99.6	99.9
90 dB	99.8	99.8

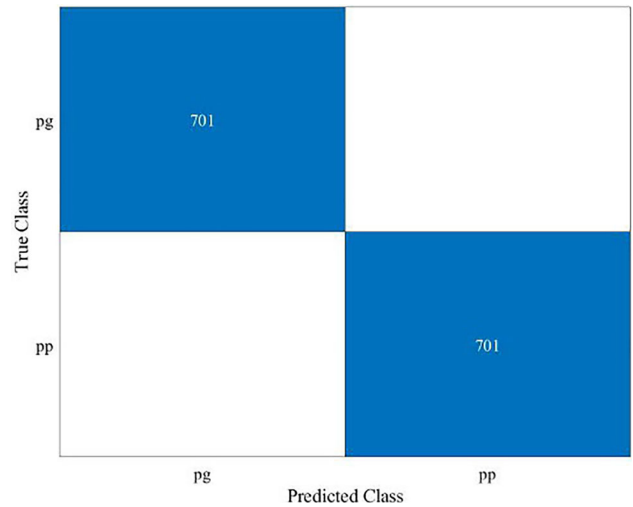


Fig. 10 CM for trained cosine KNN fault classification scheme

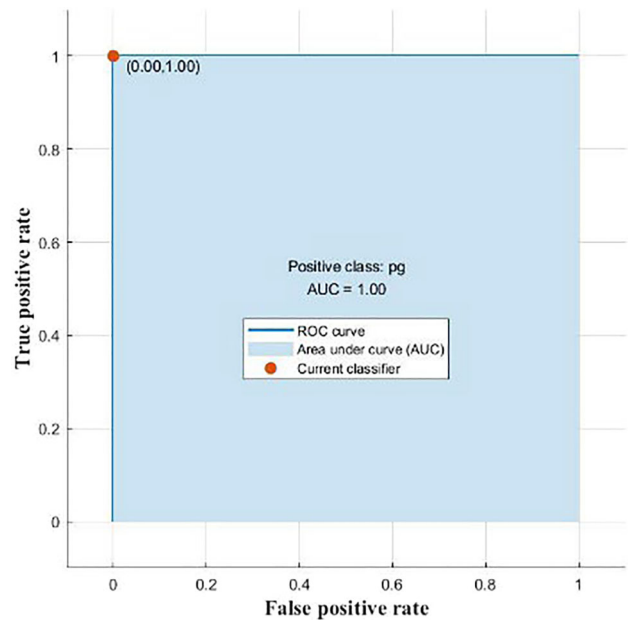


Fig. 11 ROC for trained cosine KNN fault classification scheme

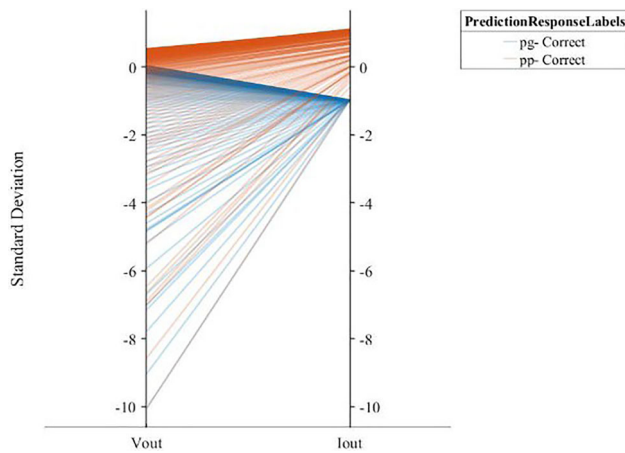


Fig. 12 PCP for trained cosine KNN fault classification scheme

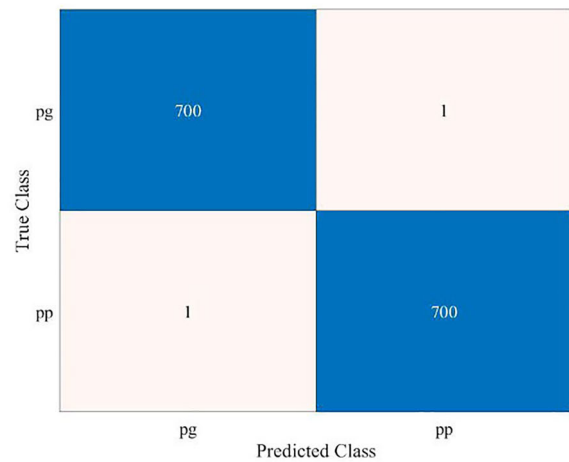
proposed classifiers have been tested with random contaminated data sets and have been shown in Table 3.

From Table 3, it is observed that the testing accuracy without noise for both classifiers is 100%. However, in a noisy environment, with different SNRs, the performance of the BEL reduces minimally whereas the testing accuracy of cosine-KNN remains 100%. It established that the proposed fault detection and classification scheme is robust and resilient.

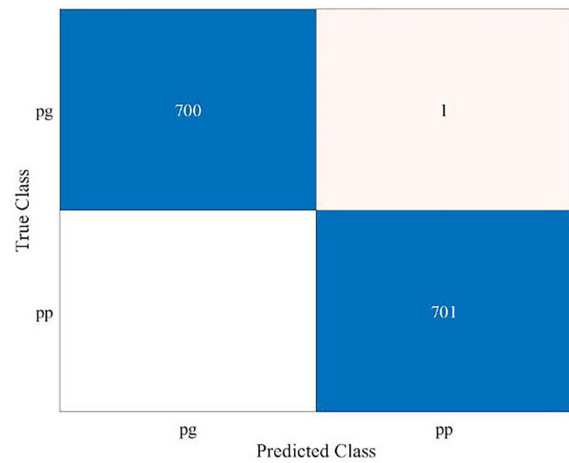
3.4 Comparative study of proposed data-driven ML fault detection and classification schemes with methods reported in the literature

To establish the superiority of the proposed classifiers, the results obtained from BEL and cosine-KNN have been compared with the other ML-based techniques considering objectives, communication channel requirements, complexity involved, and detection/classification capacity. Comparative results are shown in Table 4. It can be summarized that the proposed fault detection and classification approaches outperformed the existing fault identification in the following ways:

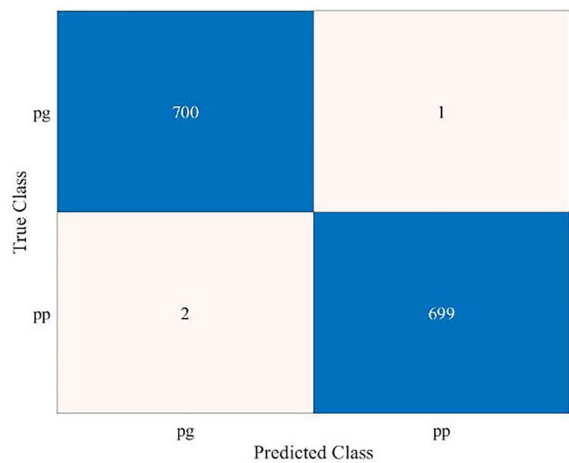
1. Existing methods considered small test cases and samples to validate the same. Further, few works did not confirm the testing of the suggested approaches. The proposed scheme has been tested using 901 random data samples of local voltage and current.
2. In reported works, lower training and detection accuracy have been achieved. The proposed work provides 99–100% accuracy for both training and testing.
3. The majority of the reported work did not consider a noisy environment. Hence, the robustness of the proposed techniques is in question. In the proposed work, the noise of



(a)



(b)



(c)

Fig. 13 CM of the trained cosine KNN classification learner using a 30 dB b 60 dB c 90 dB

Table 3 Performance of the bagged tree ensemble learner and cosine-KNN techniques for fault detection and classification with and without noise in testing

SNR	Testing accuracy (%)	
	Bagged tree ensemble learner	Cosine KNN
Noise-free	100	100
30 dB	99.74	100
60 dB	99.103	100
90 dB	99.134	100

different SNRs has been included in training and testing. The results established the robustness of the proposed method.

4. In comparison to the existing methods, the proposed method has lesser computational time and complexities which makes it superior to implement in complex networks.

4 Conclusion

LV DC microgrid has gained substantial significance for tapping renewable energy sources. However, the protection of LV DC microgrid is a major concern. Therefore, in this work, the prime focus is to introduce a strategic methodology for accurately detecting and classifying the faulty and non-faulty scenarios in a standalone DC MG to facilitate the isolation of the faulty section by the circuit breaker. The bagged ensemble learner effectively pinpointed faults, while the cosine-based KNN categorizes them as PP or PG faults, assisting operators in taking immediate corrective measures. To ensure the robustness of the proposed fault detection and classification scheme, the classifiers are trained with noisy data samples. The technique undergoes comprehensive training, validation, and testing, accounting for no noise and system noise. Results indicate that the proposed scheme seamlessly identifies and categorizes the faults. Thus, it is suitable to integrate with the relaying components of DCCBs or hybrid circuit breakers (HCBs) for quick fault detection and interruption. Comparative results with existing machine-learning techniques for microgrid fault detection established the superiority of the

Table 4 Comparative Analysis of the proposed Bagged tree ensemble learner and cosine-KNN approaches with other ML Schemes reported in the literature

Classifiers used	Objective	Accuracy (%)	Communication channel	Complexities
ANN [29], 2016	Detection and localization of faults in a DC MG system with a ring configuration	100	Required	The protection scheme has two ANNs, which adds complexity and training time
DWT based SVM [24], 2017	Fault detection & and classification of 34.5 kV DC microgrid	95–98	Not required	This technique requires a high sample frequency
ANN [35], 2019	Fault detection & classification	96.1	Not required	To execute practically it fails to fulfil the stringent time constraints for DC fault interruption
Wavelet Transform based ANN, [31] 2019	Detection and localization of faults in a DC MG system with a grid-type configuration	97–99	Not required	Massive data processing is an issue
Wavelet energy- fuzzy neural network-based technique [36], 2020	Fault detection	98–99	Not required	It is difficult to choose a threshold for a high number of fuzzy rules
Recurrent Neural Networks (RNN), [43], 2021	Fault detection	99.99	Not required	Selecting features for RNNs to avoid overfitting is complex
DT & KNN [44], 2022	k-nearest neighbors is utilized for fault categorization, while decision trees are used for fault detection	99	Not required	Calculation of the differential current increases the computational time
Bagged tree Ensemble learning with Cosine KNN (Proposed method)	Ensemble learning has been used for fault detection & Cosine k-nearest neighbors have been used for fault classification	100	Not required	There is no complexity as the proposed methodology is robust and efficient

proposed schemes. In future work, (i) a novel algorithm will be developed to estimate the location of faults (ii) multiple faults on the network will also be analysed in the future.

Author contributions AD wrote the main manuscript text and prepared all the Figures. AKJ prepared the Tables & reviewed the manuscript.

Funding Not Applicable.

Data availability Not Applicable.

Declarations

Conflict of interest There are no competing interests, according to the authors. The authors further state that none of their known conflicting financial interests or personal connections could exist to have an impact on the work presented in this study.

Ethical approval Not applicable.

References

- Chowdhury D, Hasan AK, Khan MZR (2020) Islanded DC microgrid architecture with dual active bridge converter-based power management units and time slot-based control interface. *IEEJ Trans Electr Electron Eng* 15:863–871
- Hasan ASMK, Chowdhury D, Khan MZR (2018) Scalable DC microgrid architecture with a one-way communication-based control Interface. In: Proceedings 2018 10th international conference on electrical and computer engineering (ICECE), Dhaka, Bangladesh, 20–22; pp 265–268
- Chowdhury D, Hasan ASMK, Khan MZR (2018) Scalable DC microgrid architecture with phase-shifted full bridge converter-based power management unit. In: Proceedings 2018 10th international conference on electrical and computer engineering (ICECE), Dhaka, Bangladesh, 20–22; pp 22–25
- Mirsaeidi S, Dong X, Said DM (2018) Towards hybrid AC/DC microgrids: critical analysis and classification of protection strategies. *Renew Sustain Energy Rev* 90:97–103
- Beheshtaein S, Cuzner RM, Forouzesh M, Savagheb M, Guerrero JM (2019) DC microgrid protection: a comprehensive review. *IEEE J Emerg Sel Top Power Electron* 149:111401
- Grcić I, Pandžić H, Novosel D (2021) Fault detection in DC microgrids using short-time fourier transform. *Energies* 14:277. <https://doi.org/10.3390/en14020277>
- Dagar A, Gupta P, Niranjana V (2021) Microgrid protection: a comprehensive review. *Renew Sustain Energy Rev* 149:111401. <https://doi.org/10.1016/j.rser.2021.111401>
- Li C, Rakhra P, Norman PJ, Burt GM, Clarkson P (2021) Multi-sample differential protection scheme in DC Microgrids. *IEEE J Emerg Sel Top Power Electron* 9(3):2560–2573. <https://doi.org/10.1109/JESTPE.2020.3005588>
- Prince SK, Affijulla S, Panda G (2023) Protection of DC microgrids based on complex power during faults in on/off-grid scenarios. *IEEE Trans Ind Appl* 59(1):244–254. <https://doi.org/10.1109/TIA.2022.3206171>
- Sistani A, Hosseini SA, Sadeghi VS, Taheri B (2023) Fault detection in a single-bus DC microgrid connected to EV/PV systems and hybrid energy storage using the DMD-IF method. *Sustainability* 15:16269. <https://doi.org/10.3390/su152316269>
- Haron AR, Mohamed A, Shareef H, Zayandehroodi H (2012) Analysis and solutions of overcurrent protection issues in a microgrid. In: Proceedings of the IEEE International Conference on Power and Energy (PECon), Kota Kinabalu, Malaysia; pp 644–649
- Farhadi M, Mohammed OA (2015) Event-based protection scheme for a multiterminal hybrid DC power system. *IEEE Trans Smart Grid* 6:1658–1669
- Yugeswar RO, Soumesh C, Chakraborty AK (2022) Bilayered fault detection and classification scheme for low-voltage DC microgrid with weighted KNN and decision tree. *Int J Green Energy* 19:11
- Marija Č, Hrvoje P, Juraj H (2023) Mathematical morphology-based fault detection in radial DC microgrids considering fault current from VSC. *IEEE Transactions on Smart Grid*. <https://doi.org/10.1109/TSG.2022.3229979>, 14,4,(2545-2557)
- Som S (2018) Efficient protection scheme for low-voltage DC micro-grid. *IET Gener Transm Distrib* 12:3322–3329
- Shen WC, Chen YH, Wu AY (2014) Low-complexity sinusoidal-assisted EMD (SAEMD) algorithms for solving mode-mixing problems in HHT. *Digit Signal Process* 24:170–186
- Sarangi S, Sahu BK, Rout PK (2020) A comprehensive review of distribution generation integrated DC microgrid protection: issues, strategies, and future direction. *Int J Energy Res* 45:5006–5031
- Verma A, Yadav A (2015) ANN-based fault detection and direction estimation scheme for series compensated transmission lines. In: Proceedings of the IEEE International Conference on electrical, computer and communication technologies (ICECCT), Coimbatore, India, 5–7, pp 1–6
- Li W, Monti A, Ponci F (2014) Fault detection and classification in medium voltage DC shipboard power systems with wavelets and artificial neural networks. *IEEE Trans Instru Meas* 63(11):2651–2665
- Yue M, Atif M, Damian O, Keith C (2022) Wavelet transform data-driven machine learning-based real-time fault detection for naval DC pulsating loads. *IEEE Trans Transp Electr* 8(2):1956–1965
- Manohar M., Koley E (2017) SVM based protection scheme for microgrid. In: international conference on intelligent computing, instrumentation and control technologies (ICICICT). Kerala, India: IEEE
- Delpha C, Diallo D, Samrout Al H, Moubayed N (2018) Multiple incipient fault diagnosis in three-phase electrical systems using multivariate statistical signal processing. *Engg App Artif Intell* 73:68–79
- Asadi M, Samet A, Ghanbari T (2017) k-NN based fault detection and classification methods for power transmission systems. *Prot Control Mod Power Syst* 2:32
- Manohar M, Pravat KR (2017) Detection and classification of micro-grid faults based on HHT and machine learning techniques. *IET Gener Trans Distrib* 12(2):388–397
- Somesh L, Chakravarty A, Maiti AD (2022) Fault diagnosis in power transmission line using decision tree and random forest classifier. Durgapur, India, In: IEEE 6th international conference on condition assess. Techniques in electrical systems (CATCON)
- Amal Krishna TS, Hari Kumar R (2023) Fault detection and classification for DC microgrid using binary classification models. Thiruvananthapuram, India, In: international conference on control, Communication and Computing (ICCC)
- Ying-Yi H, Mark MCTA (2019) Fault detection, classification, and location by static switch in microgrids using wavelet transform and taguchi-based artificial neural network. *IEEE Syst J* 14(2):2725–2735
- Abdali A, Mazlumi K, Noroozian R (2019) High-speed fault detection and location in DC microgrids systems using multi-criterion system and neural network. *Appl Soft Comput* 79:341–353
- Yang Q, Li J, Blond LS, Wang C (2016) Artificial neural network-based fault detection and fault location in the DC microgrid. *Energy Proc* 103:129–134

30. Som S, Samantaray SR (2018) Efficient protection scheme for low-voltage DC micro-grid. *IET Gener Transm Distrib* 12(13):3322–3329
31. Jayamaha DKJS, Lidula NWA, Rajapakse AD (2019) Wavelet multi resolution analysis-based ANN architecture for fault detection and localization in DC microgrids. *IEEE Access* 7:145371–145384
32. Mohanty R, Sahoo S, Pradhan AK, Blaabjerg F (2021) A cosine similarity based centralized protection scheme for DC microgrids. *IEEE J Emerg Sel Topics Power Electr* 9(5):1–1
33. Fahim SR, Sarker S, Muyeen SMRI, Das SK (2020) Microgrid fault detection and classification: machine learning based approach, comparison, and reviews. *Energies* 13(13):3460
34. Chen K, Hu J, He J (2016) Detection and classification of transmission line faults based on unsupervised feature learning and convolutional sparse autoencoder. *IEEE Trans Smart Grid* 9:1748–1758
35. Jayamaha DKJS, Lidula NWA, Rajapakse AD (2019) Wavelet based artificial neural networks for detection and classification of DC microgrid faults. *IEEE power & energy society general meeting (PESGM)*
36. Chen CI, Lan CK, Chen YC, Chen CH, Chang CH (2020) Wavelet energy fuzzy neural network-based fault protection system for microgrid. *Energies MDPI* 13(4):1–13
37. Chatterjee S, Roy BKS (2020) Bagged tree based anti-islanding scheme for multidig microgrids. *J Ambient Intell Humaniz Comput*. <https://doi.org/10.1007/s12652-020-02324-0>
38. Islam M, Usman M, Newaz A, Faruque M (2022) Ensemble voting-based fault classification and location identification for a distribution system with microgrids using smart meter measurements. *IET Smart Grid* 6(3):219–232. <https://doi.org/10.1049/stg.2.12091>
39. Wang T, Tan Y, Wang Y, Jin B, Monti A, Sangiovanni-Vincentelli AL (2022) Synthetic data in DC microgrids: label creation for ensemble learning for fault isolation. *IEEE Trans Power Deliv* 37(3):2301–2313. <https://doi.org/10.1109/TPWRD.2021.3110182>
40. Okfalisa, Gazalba I, Mustakim, Reza NGI (2017) Comparative analysis of k-nearest neighbor and modified k-nearest neighbor algorithm for data classification. In: proceedings of the 2nd international conferences on information technology, information systems and electrical engineering (ICITISEE), Yogyakarta, Indonesia, 1–3, 294–298
41. Deb A, Jain AK (2023) Design of a robust and fast fault diagnosis strategy of a stand-alone LVDCMG using gaussian SVM & weighted KNN algorithms. In: 3rd international conference on energy, power and electrical engineering (EPEE), Wuhan, China, 2023, pp. 1247–1253, doi: <https://doi.org/10.1109/EPEE59859.2023.10352007>.
42. Gangwar AK, Shaik AG (2023) k-Nearest neighbour-based approach for the protection of distribution network with renewable energy integration. *Electr Power Syst Res* 220:109301. <https://doi.org/10.1016/j.epsr.2023.109301>
43. Grcić I, Pandžić H (2021) Fault detection in DC microgrids using recurrent neural networks. In: international conference on smart energy system and technologies (SEST), Vaasa, Finland, 1–6
44. Saxena A, Sharma NK, Samantaray SR (2022) An enhanced differential protection scheme for LVDC microgrid. *IEEE J Emerg Sel Top Power Electr* 10(2):2114–2125
45. Meskin M, Domijan A, Grinberg I (2020) Impact of distributed generation on the protection systems of distribution networks: analysis and remedies—review paper. *IET Gener Transmiss Distrib* 14(24):5944–5960
46. Zhou Z (2021) Ensemble learning. *Machine learning*. Springer, Singapore, pp 181–210
47. Breiman L (1996) Bagging predictors. *Mach Learn* 24:123–140
48. Deleplace A, Atamuradov V, Allali A, Pellé J, Plana R, Alleaume G (2020) Ensemble learning-based fault detection in nuclear power plant screen cleaners. *IFAC-Papers On-Line* 53(2):10354–10359
49. Gangavva C, Mangai J, Bansal M (2022) Advances in parallel computing algorithms, tools and paradigms: an investigation of ensemble learning algorithms for fault diagnosis of roller bearing, pp 117–125
50. Naimi A, Deng J, Doney P, Sheikh-Akbari A, Shimjith SR, Arul AJ (2022) Machine learning-based fault diagnosis for a PWR nuclear power plant. *IEEE Access* 10:126001–126010
51. Johnson JM, Yadav A (2016) Fault detection and classification technique for HVDC transmission lines using KNN. *International Conference on ICT for sustainable development ICT4SD, LNNS*, 10
52. Majumdar A, Jain AK, Debnath R (2023) Assessment of transient disturbance using discrete fourier transform and feed forward neural network based hybrid classifier. In: 2023 IEEE 3rd international conference on sustainable energy and future electric transportation (SEFET), Bhubaneswar, India, pp. 1–6, doi: <https://doi.org/10.1109/SeFeT57834.2023.10245639>.
53. Karan S, Yeh HG (2020) Fault classification in microgrids using deep learning. In: 2020 IEEE green energy and smart systems conference (IGESSC), Long Beach, CA, USA, pp 1–7, <https://doi.org/10.1109/IGESSC50231.2020.9285101>
54. Nayak S, Bhat M, Reddy NVS, Rao BA (2022) Study of distance metrics on k - nearest neighbor algorithm for star categorization. *J Phys Conf Ser* 2161:012004

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.