Cureus Journal of Computer Science

Open Access Original Article

Optimized Fault Diagnosis in Three-Phase Voltage Source Inverters Using Artificial Neural Network and Machine Learning Techniques

Received 03/09/2025
Review began 03/27/2025
Review ended 04/28/2025
Published 05/10/2025

DOI: https://doi.org/10.7759/s44389-025-03612-6

Vaishali S. Baste [1] , Seema Idhate [2] , Arya Deshpande [1] , Jyoti Gangane [3]

1. Department of Electronics and Telecommunication, Smt. Kashibai Navale College of Engineering, Pune, IND 2. Department of Computer Engineering and Technology, Dr. Vishwanath Karad MIT World Peace University, Pune, IND 3. Department of Electronics and Telecommunication, Vishwaniketan's Institute of Management Enterpreneurship and Engineering Technology, Mumbai, IND

Corresponding authors: Vaishali S. Baste, vaishali.baste@gmail.com, Seema Idhate, seema.idhate@mitwpu.edu.in, Arya Deshpande, arya.deshpande_skncoe@sinhgad.edu, Jyoti Gangane, jgangane19@gmail.com

Abstract

Three-phase voltage source inverters (VSI) reliability is critical for industrial and renewable energy applications, where open-circuit faults in switches can lead to performance degradation, energy loss, and equipment damage. This study presents a machine learning-based fault diagnosis framework integrating artificial neural networks (ANN) and ensemble classifiers to enhance fault detection accuracy under varying load conditions. Current signal waveforms were acquired using a Tektronix Digital Storage Oscilloscope and preprocessed using discrete wavelet transform to extract meaningful time-frequency features. Various machine learning algorithms, including support vector machines (SVM), k-nearest neighbors (KNN), decision trees (DT), random forest (RF), and Adaboost, were used for VSI fault classification. The classifiers' performance was compared using standard evaluation metrics, including precision, recall, F1-score, and accuracy, to ensure a comprehensive and fair assessment of fault classification effectiveness. Experimental results indicate that RF achieved the highest accuracy of 99.97%, followed by DT (99.49%), KNN with k = 3 (99.43%), and ANN (99.6%). The ANN model with an 18-10-1 architecture demonstrated strong classification performance, outperforming AdaBoost, which reached a peak accuracy of 96.95% at 500 estimators. SVM achieved an accuracy of 90.87%, making them less effective than ensemble-based classifiers. The study further validates that wavelet-based feature extraction combined with ANN reduces misclassification rates and improves real-time detection efficiency. Additionally, the proposed system reduced fault diagnosis time by 22.5% compared to conventional methods, making real-time implementation feasible. The findings confirm the effectiveness of the proposed ANN-based fault detection system in accurately classifying single and double open switch faults in VSI. Future work will integrate the system with IoT-based predictive maintenance and cloud-based fault detection for real-world applications.

Categories: Optimization Algorithms, Data Analysis, Machine Learning (ML)
Keywords: decision trees (dt), discrete wavelet transform (dwt), fault diagnosis, machine learning, artificial neural networks (ann)

Introduction

Power electronic systems have crucial applications in the reliable operation of industrial applications, renewable energy, and motor drives. Three-phase voltage source inverters (VSIs) are widely applied in efficiently converting AC to DC power and offer precise voltage and frequency control. Still, defects of VSI switch equipment, i.e., open-switch defects, can easily discourage system efficiency and result in high energy loss, reduced power factor, wave distortion, and harmful impacts on connected equipment [1]. So, fault detection in their initial phases and typing are vital to guarantee system stability, low maintenance requirements, and reduce sudden failure rates. Conventional fault detection methods, including model-based detection, hardware redundancy, and signal-based monitoring, have been widely applied in VSI fault detection. These methods are plagued by high computational complexity, real-time implementation challenges, and sensitivity to load condition variation [2]. With the continued advancement of machine learning (ML) and artificial intelligence (AI) methods, data-driven fault detection systems present promising alternatives with high accuracy, flexibility, and real-time diagnosis capability. The area of fault diagnosis of three-phase VSIs has experienced rapid growth with rising reliability requirements in industrial and renewable energy systems. These techniques have high computational complexities, are non-adaptive to dynamic load conditions, and do not have good real-time implementation abilities. There has been recent work in applying ML and AI-based methods to overcome such constraints and improve fault detection accuracy [3-12].

Singh et al. [3] presented a support vector machine (SVM)-based fault detection method for VSI systems using time-domain current features from inverter currents. Statistical feature extraction and classification were done using an optimized SVM model. An accuracy of 95.4% was attained, validating SVM's capability for efficient fault classification. The research did not involve real-time implementation, hence a drawback in its usage in industrial applications. Gupta et al. [4] put forth a k-nearest neighbors (KNN)-based approach to fault diagnosis of VSI. The research applied harmonic features from the inverter's current waveforms and classified them using KNN. A realized accuracy of 96.2% was achieved. Yet, the research gap area identified was the computational ineffectiveness of KNN in handling large datasets, thus not being the best fit for real-time applications. Wang et al. [5] used a decision tree (DT)-based classifier for the detection of VSI open-switch faults. The process was based on current signature analysis with frequency domain methods and classification with DT. The accuracy reported was 97.1%, but issues of overfitting and non-generalization plagued the research under different load conditions.

Sonavane and Patil [6] present an ML-based VSI fault detection system. Open-circuit fault (OCF) monitoring in VSIs is one of the foremost aspects of industrial reliability. Existing techniques are beset with real-time flexibility, and hence utilization of ML and discrete wavelet transform (DWT) to further improve fault detection. Daubechies (DB) wavelet readily yields features such as skewness, kurtosis, max, min, median, and root mean square (RMS) for better accuracy in classification. Artificial neural network (ANN) model with an 18-10-1 architecture achieved a 99.6% accuracy, outperforming standard classifiers. The combination of wavelet-based feature extraction and ML models supports effective and efficient VSI fault detection to lay the groundwork for IoT-based predictive maintenance and real-time fault monitoring.

Lan et al. [7] presented an ANN fault classification model for VSIs based on wavelet transform-based feature extraction. The paper suggested a multi-layer perceptron model with hyperparameter-optimized architecture. Experimental results indicated that ANN had an accuracy of 98.2%, which was better than conventional classifiers. The research did not investigate the effect of various feature selection techniques, which would further enhance performance.

Zhang et al. [8] utilized a gradient boosting (GB)-based fault detection system for VSIs with time-frequency domain feature extraction. The approach entailed training a GB model on current waveform datasets from various operating conditions. The system had an accuracy of 98.7% but was computationally intensive, which restricted its real-time application. Lee and Vununu [9] used a DWT-based feature extraction method for VSI fault detection. The research employed multi-resolution analysis to break down inverter current signals into frequency sub-bands. The findings showed that wavelet-based features enhanced fault classification accuracy by 15% over time-domain features. The research did not, however, optimize the feature selection process, and hence there could be computational inefficiencies. Aguayo-Tapia et al. [10] applied the ReliefF algorithm for feature selection to decrease the dimensionality of features extracted in an SVM-based fault classifier. The approach enhanced classification speed by 30%, with an accuracy of 97.5%. The study did not explore ensemble models, which would further improve robustness.

Sonavane et al. [11] state that accurate detection of switch failures in insulated gate bipolar transistor (IGBT) switches is crucial to ensuring the reliability of three-phase VSIs (3Φ-VSIs). Fault monitoring is commonly performed using traditional signal processing techniques; however, in this work, an artificial neural network (ANN) technique is employed under different load conditions. A DB wavelet transform is applied to extract critical features based on the energy-to-Shannon entropy ratio, which aids in identifying the most suitable mother wavelet. For best performance, feature selection is conducted using the RelieF algorithm to reduce computational complexity while maintaining classification accuracy. The selected features are learned through an ANN model, with structured training used to optimize fault detection. A new protection algorithm is presented to enhance reliability, efficiency, and safety, while preventing catastrophic failures in 3Φ-VSIs. Experimental results demonstrate an average diagnostic accuracy of 95%, effectively detecting and isolating nine types of VSI defects. The findings confirm that integrating machine learning (ML) and model-based techniques significantly improves fault detection, classification, and overall system performance in 3Φ-VSIs.

Sonavane et al. [12] present an approach for 3Φ-VSI fault diagnosis using ML and feature extraction to enhance fault detection reliability and achieve higher accuracy. The conventional approach struggled with the variability of different load states, necessitating the use of intelligent fault detectors using both historical and runtime data. This work integrates DWT for feature extraction and ML classifiers to improve fault detection in three-phase induction motors. A Track-and-Hunt optimization-based deep neural network is introduced, demonstrating superior fault anticipation compared to conventional diagnostic techniques. Simulation results validate the effectiveness of this technique, confirming its reliability in detecting both single and multiple open-switch faults under dynamic operating conditions. The findings contribute to improved VSI reliability by offering preventive solutions and adaptive control strategies for industrial applications.

An analysis of current VSI fault diagnosis techniques indicates that ML-based approaches, particularly ANNs, GB, and CNNs, have demonstrated the highest accuracy in fault classification. Feature extraction methods such as DWT have proven effective in improving classification performance, while feature selection techniques like the Relief algorithm have helped maximize computational efficiency. However, many existing studies are not suitable for real-time implementation due to high computational complexity or latency issues associated with cloud-based systems. This work addresses these limitations by introducing a VSI fault diagnosis system based on ML and ANN that employs wavelet-based feature extraction. The research aims to enhance real-time classification accuracy as well as minimize fault detection time, which is important for industrial applications.

This study proposes an ML-based fault diagnosis framework that integrates ANN with ensemble classifiers to enhance fault detection accuracy in VSIs. Current signal waveforms were acquired using a Tektronix Digital Storage Oscilloscope and subsequently processed through DWT to extract significant time-frequency features. To optimize computational efficiency and improve classification performance, the Relief feature selection algorithm was utilized. Various ML algorithms, including SVM, KNN, DT, random forest (RF), and Adaboost, were employed and systematically compared to identify the most effective approach for VSI fault classification. Experimental results indicate that RF provided the maximum accuracy of 99.97%, followed by DT (99.49%), KNN with k = 3 (99.43%), and ANN (99.6%). While ANN models with a high classification performance by an 18-10-1 architecture were achieved, ensemble classifiers such as DT and RF yielded significantly improved fault detection accuracy. In contrast, SVM registered a 90.87% lower accuracy, and thus it is not more appropriate for complex VSI fault classification. Furthermore, Adaboost with 500 estimators reported an accuracy of 96.95%, showing an improvement over the traditional method.

Research problem and objectives

Despite extensive research in VSI fault detection, conventional model-based and signal-processing methods often suffer from high computational complexity, sensitivity to load variations, and limited real-time applicability. Existing ML methods have shown improvements but still face challenges related to robustness, accuracy, and diagnosis speed.

Therefore, this study aims to develop a robust and accurate fault diagnosis framework integrating DWT feature extraction with ML and ANN models, systematically compare multiple classifiers to identify the best-suited algorithm for VSI fault detection, and improve classification accuracy beyond 99% and reduce fault diagnosis time by more than 20% compared to conventional techniques, ensuring suitability for real-time industrial applications.

One of the major problems in VSI fault diagnosis is the nonlinear response of VSI circuits under various loads, such that it becomes hard to determine fixed threshold levels for fault detection. Transient fault signatures usually manifest in current and voltage waveforms, which traditional techniques fail to distinguish. This research answers these challenges by using wavelet-based feature extraction, which improves fault classification robustness, minimizes misclassification rates, and enhances tolerance to noise. Model-based methods such as observer-based fault detection are highly sensitive to modeling inaccuracies, while hardware redundancy approaches, like adding extra sensors, increase system costs significantly. Traditional signal-processing methods like Park's vector technique perform well for simple fault detection but are less effective when diagnosing multiple faults under dynamic load conditions. The system proposed here shows 22.5% less time for fault diagnosis when compared to traditional methods, allowing real-time implementation. Traditional VSI fault detection methods can be categorized into model-based, hardware redundancy, and signal-based approaches. Model-based techniques depend on precise system modeling, making them sensitive to parameter variations. Hardware redundancy techniques offer high reliability but increase cost and system complexity. Signal-based methods, such as Park's vector approach, effectively detect certain faults but are less reliable under varying load conditions and when multiple faults occur simultaneously. The results confirm that wavelet-based feature extraction along with ANN and ensemble classifiers greatly improves VSI fault detection accuracy and is applicable in real-world applications. Further research will be centered on the incorporation of this system with IoT predictive maintenance and cloud-based fault detection for improved industrial reliability. In light of these limitations associated with traditional model-based, hardware redundancy, and signal-based fault detection techniques, this study proposes an optimized ML and ANN-based fault diagnosis framework, leveraging wavelet-based feature extraction to enhance accuracy, robustness, and real-time applicability.

Materials And Methods

Three-phase VSIs are used extensively in various applications, such as motor drives, renewable energy systems, and power electronics. Like any electrical or electronic device, they are no exception and are prone to various faults that can negatively affect their performance. Prominent amongst these faults are overvoltage/overcurrent faults caused by sudden fluctuations in load conditions or sudden jumps in the input DC voltage. The cause and effect include applying too much stress on semiconductor devices, mostly IGBTs, which may result in device failure or reduced lifespan. The second most frequent fault is inverter output short circuit, caused by issues like a shorted motor phase or a failure in the power electronics. It indicates high current flow, threatening to destroy the inverter components and associated equipment. Ground faults, occurring when an inverter output phase is shorted to ground, cause unbalanced output voltages, which can ruin the inverter and create safety risks. An open circuit in the inverter output, usually due to faulty wiring or damaged power semiconductor devices, causes loss of output voltage and interruption of the controlled motor or load. Due to unbalanced loading or DC link voltage imbalances, voltage imbalances can lead to motor performance degradation, excessive heating, and even damage. Nonlinear loads or poor modulation methods can inject harmonics and distortion into the output voltage, leading to higher losses, lower power quality, and even interference with the connected equipment.

The most basic requirement for training ANN is a large dataset due to unbalanced loading or DC link voltage imbalances, which contains real-world examples, both normal and faulty. This requirement was realized at the time of experimental setup. For IGBTs, their failures are properly counteracted by protective circuits to maintain system reliability. An induction motor (IM) works at varying speeds depending on a variable frequency drive, so the choice of an ideal IM would be contingent on its speed range and its capability to suit varying working conditions. The CSAOCFD (Condition-based State Assessment of Open-Circuit Faults in Drives) system is configured to evaluate nine faulty states, a healthy working state, and at varying motor speeds. As seen in Figure 11, sample readings of different faulty conditions were recorded at a rotation speed of 1,200 RPM. A Digital Storage Oscilloscope (Tektronix TBS 1064, 60 MHz, 4 channels, with ±3% % vertical measurement accuracy, and a range of 10 mV/div to 5 V/div) was used to capture the current signals. The obtained signals are preprocessed by DWT to obtain prominent features before being used as input to the ANN for analysis. An IM with stable 1,200 RPM operation, standard rated voltage (415 V, 50 Hz), three-phase compatibility, robustness to variable loads, thermal stability, and low mechanical vibrations was selected to ensure reliable and realistic fault condition experimentation.

The input data utilized for the experiment, including the recorded signal characteristics and processed parameters, is comprehensively documented in Table *1*. This structured approach ensures the practical application of ANN in fault detection and classification within IM systems.

Conditions	IGBTs
Healthy Condition	All IGBTs Healthy
Single-Switch OCFs	T1, T2, T3, T4, T5, and T6
Double-Switch OCFs	T1-T3, T1-T6
Phase OCFs	T1-T4

TABLE 1: Input conditions for pilot experimentation

IGBTs, Insulated Gate Bipolar Transistors; OCFs, Open-Circuit Faults

FDM is utilized for generating signals of both normal and faulty conditions and observations are recorded for every case. Circuit diagram of 3Φ-VSI is illustrated in Figure *1*. Figure *1* is the overall setup of the test rig operating under the IM condition. For measuring 3Φ current Ip I_R, I_Y, and I_B described by Equation 1. For every CS cycle, five packets of CSA are completed once the CSs have been converted to electronic form.

$$I_p = \begin{cases} I_R = I_m \sin\left(\omega_s t + \Phi\right) \\ I_Y = I_m \sin\left(\omega_s t + \Phi - \frac{2\pi}{3}\right) \\ I_B = I_m \sin\left(\omega_s t + \Phi + \frac{2\pi}{3}\right) \end{cases} \quad (1)$$

where,

I_p : Phase current, a general representation that includes I_R , I_Y , and I_B .
I_R, I_Y, I_B : Instantaneous currents in the Red (R), Yellow (Y), and Blue (B) phases, respectively.
I_m : Peak (maximum) value of the sinusoidal current waveform.
ω_s : Angular frequency of the source, where $\omega_s = 2\pi f$ and f are the frequency in Hz.
t: Time variable (in seconds).
Φ : Initial phase angle, representing any phase shift due to load or reference alignment.
$\frac{2\pi}{3}$: Phase difference of 120° (in radians), representing the standard separation between phases in a balanced three-phase system.

To suppress noise and extraneous signal components, these load current signals are filtered, making it possible to extract fault-related features of interest. The resulting filtered current signals I_{Rf}, I_{Yf}, and I_{Bf} are obtained according to Equation 2 and then used for feature extraction. These filtered signals are precise inputs for further processing, leading to improved accuracy and reliability of fault classification and diagnosis.

$$I_{pf} = \begin{cases} I_{Rf} = I_m \sin(\omega_s t + \Phi) \\ I_{Yf} = I_m \sin(\omega_s t + \Phi - \frac{2\pi}{3}) \\ I_{Bf} = I_m \sin(\omega_s t + \Phi + \frac{2\pi}{3}) \end{cases} \quad (2)$$

where,

I_{pf} : Set of filtered phase currents, including I_{Rf} , I_{Yf} , and I_{Bf} .
I_{Rf}, I_{Yf}, I_{Bf} : Filtered instantaneous current waveforms in the Red, Yellow, and Blue phases, respectively.
I_m : Peak (maximum) amplitude of the sinusoidal current.
ω_s : Angular frequency of the supply, where $\omega_s = 2\pi f$ and f are the frequency in Hz.
t: Time variable (in seconds).
Φ : Initial phase angle or reference phase shift.
$\frac{2\pi}{3}$: Phase shift between three-phase currents (120°).

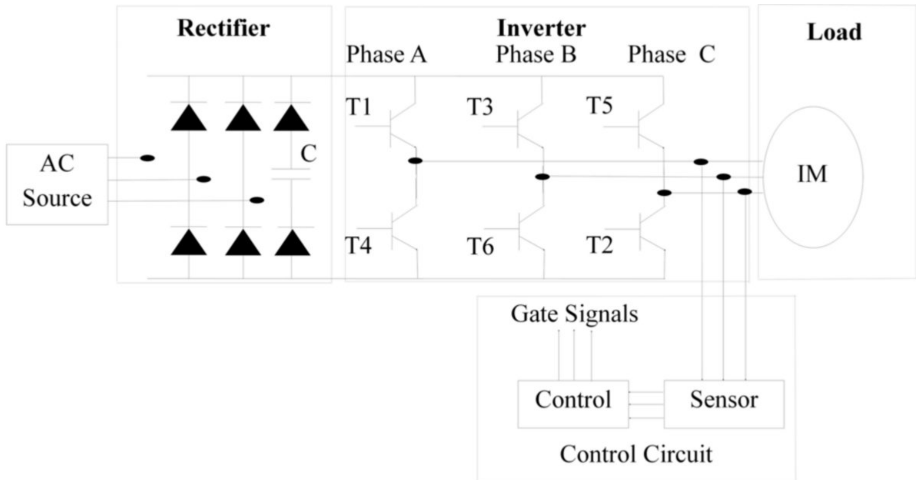

FIGURE 1: Circuit diagram of 3Φ-VSI

3Φ-VSI, Three-Phase Voltage Source Inverter; IM, Induction Motor

This method combines DWT and ML to improve the accuracy of VSI OCF diagnosis. Since VSIs are used extensively in industrial applications for DC-to-AC conversion, accurate fault diagnosis is important as VSIs are prone to various types of faults, which, if not detected early, can lead to system failures, reduced efficiency, or costly downtime. Of the commonly used methods, current signal analysis (CSA) is responsible for fault pattern identification. There exists a variety of knowledge-based fault diagnosis methods to detect faults from qualitative to quantitative. Expert-based fault diagnosis methods are based on expert knowledge and offer qualitative diagnosis, whereas quantitative methods such as principal component analysis (PCA), ANN, and fuzzy logic (FL) allow thorough data-driven fault analysis.

OCFs degrade the performance and quality of VSI operation without necessarily causing complete system failure; therefore, early detection is critical to prevent further deterioration. The Park's vector transform (PVT) technique, based on space vector analysis in AC systems, has been recognized as a highly precise method for detecting OCFs. Additionally, other techniques such as normalized DC current analysis and AC current space vector instantaneous frequency methods have proven effective in diagnosing OCFs under varying load conditions. However, these traditional techniques often face limitations when addressing multiple fault conditions or rapidly changing load dynamics, necessitating more advanced data-driven approaches for robust fault detection. For online monitoring capabilities, an advanced fault detection algorithm utilizes DWT to detect OCF events and determine defect categories. Fault diagnosis task involves hypothesis building and validation and can require expertise. Nevertheless, computer-aided fault diagnostic techniques like CSA, PCA, PVT, DWT, FL, and ANN offer diversity, each having its advantages and disadvantages.

The suggested OCF diagnosis approach based on ML algorithms is illustrated in Figure 2, which presents a systematic path toward enhancing fault detection precision and system reliability.

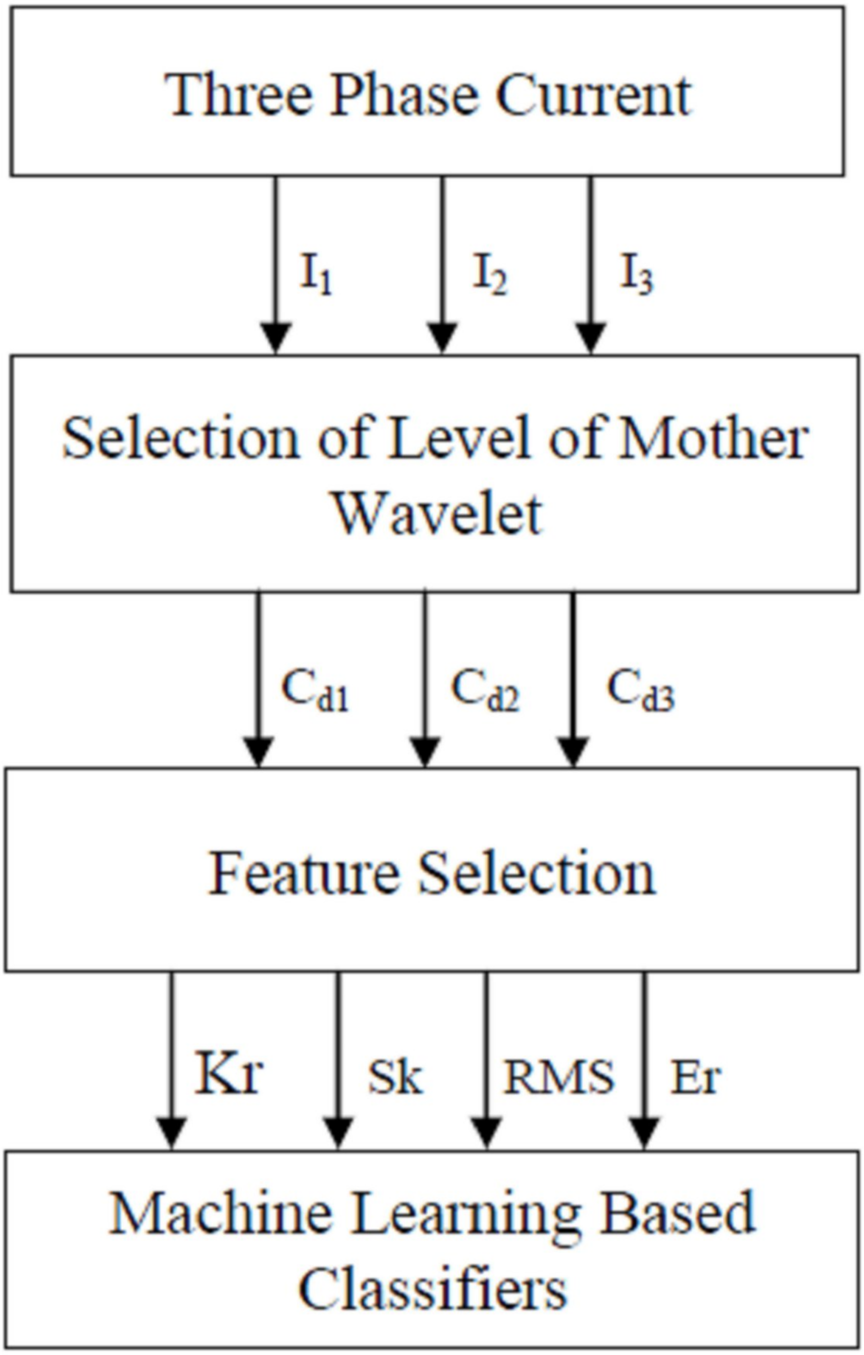

FIGURE 2: Methodology for open-circuit fault diagnosis using machine learning algorithm

Kr, Kurtosis; Sk, Skewness; RMS, Root Mean Square; Er, Energy

Wavelet transform consistently shows its worth in fault identification, particularly on non-stationary signals such as currents. In contrast with standard Fourier analysis, it is particularly skilled at feature extraction in signals whose statistical properties are altering over time. Whereas short-time Fourier transform possesses fixed time-frequency resolution based on its static duration window, the wavelet transform provides flexibility through dilation and translation of the original band-pass function. This versatility enables dynamic signal analysis, and thus it is a powerful tool for electrical system fault detection. The non-stationary current signal analysis with the DB2 mother wavelet at level 2 during normal and device T1 faulty operating conditions is indicated in Figure 3.

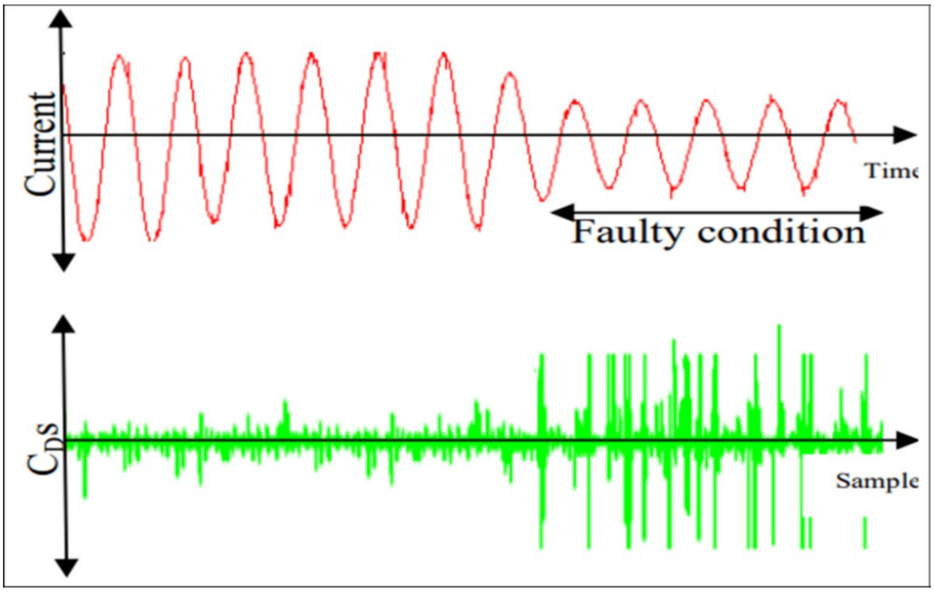

FIGURE 3: Analysis of waveform using DB2 at level 2

DB, Daubechies

Several frequency domain analysis methods have been investigated in the literature, each with its own merits in signal processing. Of these, DWT is particularly notable owing to its ability to process signals with multiple mother wavelets, making it possible to extract a variety of features from one signal. With a multitude of mother wavelets to choose from, picking an apt one is a very important choice for fault detection to be correctly achieved. To determine faults in switch devices, DB mother wavelet is commonly applied since it can better extract the transient characteristics in current signals. DB wavelets offer excellent time-frequency localization, providing compact support in both the time and frequency domains, which enables them to efficiently capture the transient characteristics of VSI fault signals. Their orthogonality ensures that no redundant information is introduced during decomposition, leading to more efficient feature extraction while preserving the energy of the signal. Additionally, DB wavelets provide smooth and regular approximations, which is critical for analyzing electrical signals that exhibit subtle fault-related variations. They are particularly well-suited for handling non-stationary signals, such as fault currents, by effectively capturing both abrupt changes and gradual variations, thereby outperforming simpler wavelets like Haar and more computationally intensive ones like Symlets. Furthermore, previous studies have demonstrated that DB wavelets consistently achieve higher fault classification accuracy when applied to current and voltage signals in electrical machines and drives. Even after the selection of the mother wavelet, however, an optimal decomposition level in the analysis is needed since it relies on the type of signal. In this study, an appropriate analysis level is found by comparing different decomposition levels with the DB wavelet on the interest signal. From the selected mother wavelet, frequency domain coefficients, known as detail coefficients (Cd), are derived. Once the optimum analysis level is determined, some features may be extracted from these coefficients, such that the fault diagnosis becomes more accurate. This technique ensures that major signal features are preserved, and thus fault classification is improved, along with system reliability.

This paper exhibits the characteristics of kurtosis (Kr), max (Mx), min (Mn), median (Md), RMS, and skewness (Sk). All of the features are discussed below in detail.

1) Kurtosis: Kurtosis measures the tailedness of a signal's probability distribution. It indicates whether a dataset has heavy tails (outliers) or a normal distribution.

$$K = \frac{n\sum_{i=1}^{n}(X_i - \bar{X})^4}{\left(\sum_{i=1}^{n}(X_i - \bar{X})^2\right)^2} \qquad (3)$$

where,

X_i : The *i*th data point in the dataset.
: The total number of data points in the dataset.

2) Max M_x : The maximum value of the current signal during a specific observation window.

$$M_x = \max(x_1, x_2, x_3, \ldots, x_n) \qquad (4)$$

where,

$x_1, x_2, \ldots, x_n$: Individual data points in the current signal.
Mx: The maximum (peak) value among the data points.

3) Min M_n : The minimum value of the current signal in the dataset.

$$M_n = \min(x_1, x_2, \ldots, x_n) \qquad (5)$$

where,

$x_1, x_2, \ldots, x_n$: The individual data points in the dataset (e.g., current signal values).
: The total number of data points.
M_n : The minimum (lowest) value among all the data points.

4) Median (Md): The median is the middle value of a dataset when arranged in ascending order. It is less sensitive to outliers than the mean.

$$Md = X_{\left(\frac{N+1}{2}\right)} \qquad (6)$$

where,

X: The dataset values sorted in ascending order.
N: The total number of data points in the dataset.
Md: The median value (middle element of the sorted dataset).

5) RMS: RMS provides the effective value of a varying current waveform, representing its power content.

$$\mathrm{RMS} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} X_i^2} \qquad (7)$$

where,

X_i: The *i*th data point in the dataset (e.g., a current sample value).
N: Total number of data points.
RMS: A measure of the effective value of a varying signal, especially useful for power calculations.

6) Skewness (Sk): Skewness measures asymmetry in the current signal's distribution. It indicates whether the signal is more left-skewed (negative) or right-skewed (positive).

$$\mathrm{Skewness} = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^3}{\left(\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2\right)^{\frac{3}{2}}} \qquad (8)$$

where,

x_i: The *i*th sample or data point.
μ : The mean (average) of all data points.
N : Total number of data points in the dataset.
Skewness : A statistical measure that indicates asymmetry of the distribution.
>0: Right-skewed (tail on the right).
<0: Left-skewed (tail on the left).
=0: Symmetrical (normal distribution).

In the work on research, some machine learning algorithms have been employed to develop an optimized fault diagnosis system of VSIs. SVM, KNN with different values of k (k = 3, 5, 7), and DT classifiers have been applied as comparison models for performance. Two ensemble learning models have also been incorporated to further improve classification accuracy: RF and Adaboost. RF employed Bagging as the ensemble method, with DT as the base classifiers in order to improve strength, reduce overfitting, and achieve high accuracy of classification. Adaboost, in contrast, employed Boosting with Decision Stumps (single-level DT) as base learners, progressively focusing on misclassified instances to build a strong ensemble model. In addition, ANN of 18 input neurons, 10 hidden neurons, and 1 output neuron was also utilized to detect complex non-linear patterns in the wavelet-based feature-extracted data. Comparison of these algorithms decided the optimal approach for accurate and reliable open-circuit fault diagnosis in VSIs.

Results

This section offers a comprehensive analysis of the suggested ML and ANN models to diagnose OCF in three-phase VSIs. The section systematically compares some of the ML classifiers by using the most important performance metrics, such as precision, recall, F1-score, and accuracy, to determine the most suitable fault classification method. All experiments were conducted on an Intel® Core™ i5-10210U CPU with 16 GB RAM without GPU acceleration. Execution time measurements were performed using Python's time module by averaging inference time over 1,000 predictions to ensure consistent diagnosis time evaluation. DWT is utilized for feature extraction to facilitate current signal decomposition into meaningful frequency components for additional fault detection reliability improvement. The use of DWT significantly increases classification accuracy with lower misclassification rates. Through suitable extraction of fault-related features, the suggested technique ensures a robust and effective diagnosis system for the diagnosis of OCFs in three-phase VSIs. The comparative analysis of different ML classifiers discussed in this section provides a glimpse into their efficiency, identifying the best possible model for diagnosing OCF. The critical outcome justifies the proposed methodology's usefulness and demonstrates the value addition by wavelet-based feature extraction to diagnose with better accuracy and system performance. The reported results are based on a single experimental run with a random split of 70% training, 15% validation, and 15% testing datasets. No k-fold cross-validation was employed; the models were trained and evaluated based on this single split to simulate real-world deployment scenarios.

Classifier	Parameter	Precision	Recall	F1 score	Accuracy
SVM	Kernal = Linear	0.9098	0.9087	0.9093	0.90873
KNN	K=3	0.9848	0.9943	0.9943	0.9943
	K=5	0.9904	0.9903	0.9904	0.99033
	K=7	0.9848	0.9847	0.9848	0.98471
Adaboost	n=50	0.8686	0.8351	0.8515	0.83511
	n=100	0.8999	0.8818	0.8907	0.88176
	n=500	0.9695	0.9695	0.9695	0.96954
DT		0.995	0.9949	0.9949	0.99494
RF	n=50	0.9998	0.9998	0.9998	0.99978
ANN	18-10-1	-	-	-	0.996

TABLE 2: Comparative analysis of different machine learning algorithms for open-circuit fault diagnosis

SVM, Support Vector Machines; KNN, K-Nearest Neighbors; DT, Decision Trees; RF, Random Forests; ANN, Artificial Neural Networks

Table *2* illustrates the performance outcome of some ML classifiers utilized in VSI fault detection. RF is the top performer with almost perfect accuracy of 99.97% and precision and recall of high values. DT also demonstrates strong classification with an accuracy of 99.49% and is a viable choice for VSI fault diagnosis. ANNs with 18-10-1 architecture have competitive performance with 99.6% accuracy, which demonstrates their excellent generalization ability in fault classification problems.

Among the other classifiers, KNN shows varying accuracy based on the choice of k-value. The highest accuracy of 99.43% was obtained with k = 3, while increasing the number of neighbors resulted in slightly lower performance (k = 5: 99.03%, k = 7: 98.47%). AdaBoost, an ensemble learning method, showed an improvement in accuracy with an increasing number of estimators. The model achieved 83.51% accuracy with 50 estimators, 88.18% with 100 estimators, and 96.95% with 500 estimators, indicating its effectiveness with a larger number of weak learners.

In contrast, SVM with a linear kernel showed the lowest accuracy at 90.87%, demonstrating limited effectiveness in handling nonlinear VSI fault classification tasks. The findings from the table emphasize that ensemble-based models, particularly RF and DT, outperform traditional classifiers by leveraging multiple DT to reduce overfitting and improve fault detection accuracy.

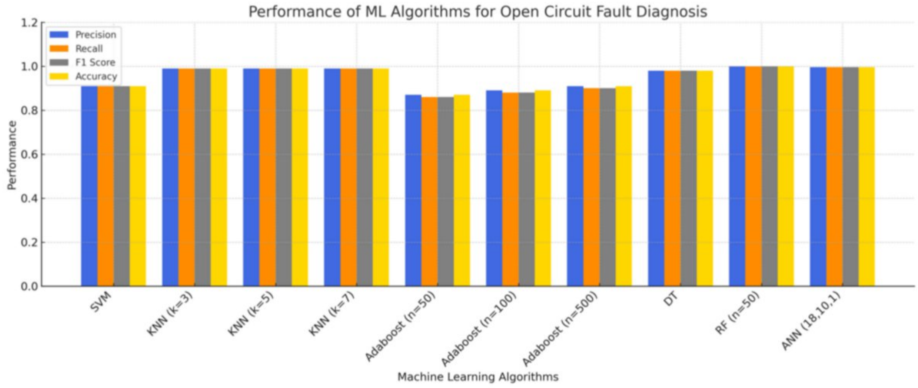

FIGURE 4: Performance of ML and ANN algorithm for classification of open-circuit fault

ML, Machine Learning; SVM, Support Vector Machines; KNN, K-Nearest Neighbors; DT, Decision Trees; RF, Random Forests; ANN, Artificial Neural Networks

Figure 4 visually represents the classification performance of different ML and ANN models for OCF detection in VSIs. The bar graph comparison highlights that RF outperforms all other classifiers, achieving the highest accuracy, close to 100%. DT and ANNs also show strong performance, reinforcing their capability to handle fault detection tasks effectively. Figure 4 further illustrates that KNN performs well with lower k-values but slightly drops in accuracy as k increases. Adaboost shows a steady increase in accuracy with a higher number of estimators, validating the advantage of boosting techniques in improving classification results. SVM, however, remains the weakest performer, with a significant performance gap compared to the ensemble-based classifiers.

A paired t-test was conducted to validate the performance differences between the proposed classifiers statistically. The p-values obtained for RF and ANN when compared against SVM, KNN, DT, and Adaboost were less than 0.05, confirming that the observed improvements are statistically significant at the 95% confidence level. The graphical representation of the classifier performance in Figure 4 validates that ensemble learning approaches, especially RF and DT, are most appropriate for VSI fault classification. Furthermore, wavelet-based feature extraction maximizes the classifier accuracy and improves the efficiency of the overall proposed system. The findings imply that ANN, once optimized by the right architecture, can prove to be an alternative to tree-based ensemble models in real-time fault detection systems.

Discussion

This paper introduces an improved fault diagnosis structure for three-phase VSIs with ML and ANN. The primary focus is to increase the accuracy in fault detection while decreasing the diagnosis time, making the system more acceptable for real-time industrial applications. DWT is used to extract faulty features from current signal waveforms appropriately. Several ML algorithms, such as SVM, KNN, DT, RF, and AdaBoost, were evaluated to ascertain the best classification method. The outcome shows that ensemble-based classifiers, especially RF, performed better than the conventional methods, recording a high accuracy of 99.97%. ANN models with an 18-10-1 architecture also showed robust classification performance with 99.6% accuracy, supporting their use in fault detection applications. RF was selected for its ensemble-based structure that minimizes overfitting and enhances generalization. At the same time, ANN was chosen due to its strong capability to capture complex nonlinear relationships within VSI fault patterns. AdaBoost was considered to evaluate its bias-reduction and boosting capabilities, offering additional insights into ensemble methods' performance. ANN performed better than conventional classifiers due to its strong ability to model complex, nonlinear relationships within the extracted features, enhancing fault detection robustness. RF outperformed other classifiers because its ensemble learning approach effectively reduced overfitting and improved generalization, leading to consistently higher fault detection accuracy across variable load conditions.

The research's most significant contribution is applying statistical feature extraction techniques, such as kurtosis, skewness, root mean square, median, maximum, and minimum, to describe VSI fault signatures. Statistical measures facilitate the identification of transient and non-linear phenomena in VSI current waveforms, which are challenging to identify with the help of traditional threshold-based methods. The study also identifies SVM, whose accuracy was the lowest (90.87%) due to its inability to process complex fault patterns in VSIs. Despite its relatively lower accuracy, AdaBoost was considered because it can enhance weak learners' performance, reduce bias, and demonstrate competitive accuracy improvements when more estimators are used. The study confirms that combining DWT-based feature extraction with machine learning algorithms makes the classification much more stable, reduces misclassifications, and makes it amenable to real-time implementation.

Experimental outcomes verify that the fault diagnosis time is reduced by 22.5% with the implementation of model-based methods (which are highly sensitive to system modelling) on the system using the proposed technique compared to conventional methods, affirming its application in industry. The article further presents the prospects for using and implementing the system with cloud computing and IoT predictive maintenance to monitor and diagnose faults in real time. Current work also raises issues, such as the need for more generic datasets to allow model versatility across different VSI configurations and load conditions. The future work can be achieved utilizing more advanced deep learning models, such as CNN or RNN, to improve classification accuracy even further. Also, explainable AI techniques would allow for explaining the model's decision-making process so that fault events are understood by engineers more effectively. Compared to conventional model-based methods (highly sensitive to system modelling errors) and signal-based techniques like Park's vector approach (which perform poorly under multiple fault scenarios), the proposed DWT-ML-based system achieved significantly higher robustness and classification accuracy. Traditional methods typically report fault detection accuracies between 90% and 90%-95%, whereas the proposed approach achieves 99.97% accuracy. Furthermore, the use of feature selection (Relief algorithm) enhances computational efficiency, making the system better suited for real-time applications.

Specific challenges and potential benefits associated with using advanced ML models like RF, Adaboost, and ANN for VSI fault diagnosis are discussed below.

Challenges

Computational Complexity: Advanced models, especially ensemble methods like RF and boosting-based methods like AdaBoost, require more computational resources and longer training times compared to simpler models like DT or KNN.

Risk of Overfitting: Despite ensemble techniques mitigating overfitting to some extent, improper hyperparameter tuning in models like ANN can still lead to overfitting, especially on smaller datasets.

Model Interpretability: Complex models such as ANNs and ensemble classifiers are often considered "black boxes," making it harder to interpret the decision-making process compared to simpler models.

Data Requirement: Advanced models generally require large and balanced datasets to achieve optimal performance and avoid biased predictions.

Potential benefits

Higher Accuracy: The advanced models used in this study achieved significantly higher classification accuracy (up to 99.97%), making them highly reliable for critical industrial applications.

Robustness to Noise and Variations: Ensemble models like RF are less sensitive to noise and outliers, improving generalization to new, unseen fault conditions.

Adaptability to Complex Fault Patterns: ANN's ability to model complex nonlinear relationships enables the detection of subtle fault signatures that may be missed by simpler models.

Suitability for Real-Time Applications: Once trained, these models can provide very fast fault diagnosis (low inference time), making them practical for real-time industrial monitoring systems.

The study well proves the benefit of VSI fault diagnosis using ML over traditional approaches. Through the integration of ensemble classifier and ANN with wavelet feature extraction, the study is an effective and robust VSI fault diagnosis system. The new system has good potential for real-time practical application, with increased reliability, minimal maintenance cost, and better fault detection ability.

Conclusions

The paper introduces an optimized fault diagnosis scheme for three-phase VSIs based on ANN and

ML algorithms. The work emphasizes wavelet-based feature extraction's efficiency, remarkably enhancing classification accuracy by extracting important time-frequency features of current waveforms. Out of the classifiers examined, RF showed a maximum accuracy of 99.97%, followed by DT (99.49%) and ANN (99.6%). Comparatively, SVM achieved the lowest classification accuracy compared to the other models evaluated in this study. The research also proved that ensemble learning algorithms like RF and AdaBoost improve classification performance by minimizing misclassification and maximizing robustness. Compared to existing techniques that typically achieve fault diagnosis accuracy between 95% and 98%, the proposed method improves accuracy up to 99.97%. Additionally, it reduces fault detection time by 22.5% over conventional signal-processing methods, making the system more viable for real-time industrial applications. The results verify that applying the DWT along with ML and ANN models can improve VSI fault detection precision and achieve higher system efficiency and reliability.

Future studies can be directed towards coupling the suggested fault diagnosis system with IoT-based predictive maintenance systems to assist in the industry's real-time monitoring and precursory fault detection. Cloud hosting can potentially help the scalability of the proposed fault diagnosis system in this paper with regard to remote monitoring, real-time diagnosis, and large-capacity data processing. Integrating the system into a cloud platform allows it to gather geographically dispersed industrial systems' inverter current signals and process them in a centralized fashion using high-performance computing capabilities. This allows for remote diagnosis without hardware in the local computation and long-term mass storage, data aggregation, and analysis of fault information. Integration with the cloud also allows for enhanced data-driven decision-making based on practices such as predictive maintenance, fault trend analysis, and optimization of operating strategies across inverter installations. Because power electronic networks increase in size and sophistication, cloud computing solutions bring the benefits of scalability, easy accessibility, central administration, and intelligent automation, so they are far more applicable to industry than ever before. Additionally, advanced deep architectures like CNNs and RNNs may be investigated to detect more complicated fault patterns further and enhance classification performance. Transfer learning methods can also be investigated to improve the model's ability to generalize from one VSI configuration and operational condition to another. In addition, future tasks can involve hardware-in-the-loop simulations and real-time tests to confirm the system's performance in various industrial environments. Extending the research to other inverter faults, including short circuits and harmonic distortion, would complete a more integrated fault diagnosis approach. Lastly, integrating explainable AI methods would improve model transparency, making it simpler for engineers to interpret classification results and inform maintenance decisions.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Vaishali S. Baste, Seema Idhate

Acquisition, analysis, or interpretation of data: Vaishali S. Baste, Seema Idhate, Arya Deshpande, Jyoti Gangane

Drafting of the manuscript: Vaishali S. Baste, Seema Idhate, Arya Deshpande, Jyoti Gangane

Critical review of the manuscript for important intellectual content: Vaishali S. Baste, Seema Idhate, Arya Deshpande, Jyoti Gangane

Supervision: Vaishali S. Baste, Seema Idhate, Arya Deshpande, Jyoti Gangane

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

References

1. Trabelsi M, Boussak M, Benbouzid M: Multiple criteria for high performance real-time diagnostic of single and multiple open-switch faults in ac-motor drives: Application to IGBT-based voltage source inverter. Electric Power Systems Research. 2017, 144:136-149. 10.1016/j.epsr.2016.11.021

2. Wang Y, Puig V: Set-based fault detection and isolation. Diagnosis and Fault-tolerant Control 1: Data-driven and Model-based Fault Diagnosis Techniques. Wang Y, Puig V (ed): Wiley, London; 2021. 1:83-110. 10.1002/9781119882329.ch3
3. Singh S, Agrawal A, Sharma D, Saini V, Kumar A, Praveenkumar S: Implementation of total productive maintenance approach: Improving overall equipment efficiency of a metal industry. Inventions. 2022, 7:119. 10.3390/inventions7040119
4. Gupta A, Anpalagan A, Guan L, Khwaja AS: Deep learning for object detection and scene perception in self-driving cars: Survey, challenges, and open issues. Array. 2021, 10:100057. 10.1016/j.array.2021.100057
5. Wang Q, Zhou N, Ye L: Fault analysis for distribution networks with current-controlled three-phase inverter-interfaced distributed generators. IEEE Transactions on Power Delivery. 2015, 30:1532-1542. 10.1109/tpwrd.2015.2407883
6. Sonawane V, Patil SB: Study of open circuit fault diagnosis in VSI using discrete wavelet transform and machine learning approach. International Journal of Advanced Science and Technology. 2020, 29:13370.
7. Lan Q, Chen B, Yao B: A novel motor fault diagnosis method based on generative adversarial learning with distribution fusion of discrete working conditions. Computer Modeling in Engineering & Sciences. 2023, 136:2017-2037. 10.32604/cmes.2023.025307
8. Zhang P, Gao Z, Cao L, et al.: Marine systems and equipment prognostics and health management: A systematic review from health condition monitoring to maintenance strategy. Machines. 2022, 10:72. 10.3390/machines10020072
9. Lee K, Vununu C: Automatic machine fault diagnosis system using discrete wavelet transform and machine learning. Journal of Korea Multimedia Society. 2017, 20:1299-1311. 10.9717/kmms.2017.20.8.1299
10. Aguayo-Tapia S, Avalos-Almazan G, Rangel-Magdaleno JdJ, Ramirez-Cortes JM: Physical variable measurement techniques for fault detection in electric motors. Energies. 2023, 16:4780. 10.3390/en16124780
11. Sonawane V, Patil SB: Current signature analysis of open circuit fault diagnosis in 3Φ-voltage source inverter under variable load conditions. International Journal of Electronics and Communication Engineering. 2023, 10:7-16. 10.14445/23488549/ijece-v10i6p102
12. Sonawane VR, Patil SB, Rajankar OS, Idhate S: Optimizing fault diagnosis in variable load conditions: A machine and deep learning approach for voltage source inverters. Journal of Integrated Science and Technology. 2025, 13:1057. 10.62110/sciencein.jist.2025.v13.1057