Duzce University Journal of Science and Technology
Düzce Üniversitesi Bilim ve Teknoloji Dergisi
14 (1), 2026, 143-151.
https://doi.org/10.29130/dubited.1808533
e-ISSN: 2148-2446
https://dergipark.org.tr/en/pub/dubited



Research Article

Machine Learning Based Fault Diagnosis in Variable Speed Synchronous Generators

Mithat Önder[1*], Muhsin Uğur Doğan[2]

[1] Bolu Abant Izzet Baysal University, Faculty of Engineering, Department of Electrical and Electronics Engineering, Bolu, Türkiye. ror.org/01x1kqx83

[2] Bolu Abant Izzet Baysal University, Bolu Technical Sciences High School, Department of Electronic and Automation, Bolu, Türkiye. ror.org/01x1kqx83

*Corresponding author: mithatonder@ibu.edu.tr

Received: 22.10.2025
Revision: 19.11.2025
Accepted: 25.12.2025
Available Online: 21.01.2026

Abstract: In this study, the performance of five different machine learning algorithms, decision tree, random forest, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and logistic regression, was investigated for fault diagnosis in variable-speed synchronous generators. The dataset, consisting of real-world experimental data, includes both healthy and faulty generator operating states. Pre-processing steps such as normalization, Z-Score standardization, and feature selection were applied to the data, and the effects of these processes on classification performance were evaluated. According to the findings, the decision tree algorithm achieved the highest performance with an accurate rate of 99.43% and Matthews Correlation Coefficient (MCC) value of 0.975. While the random forest algorithm yielded similar results, the KNN, SVM, and logistic regression algorithms achieved lower accuracy values. It was determined that the pre-processing steps did not provide a significant increase in model performance, and the dataset was already balanced in terms of scale. The results revealed that the decision tree algorithm is the most suitable and reliable method for fault detection in variable-speed synchronous generators. This study demonstrates that machine learning-based approaches can be used effectively in early fault diagnosis in generators.

Keywords: *Synchronous generator, Machine learning, Fault diagnosis*

I. INTRODUCTION

Synchronous generators play a critical role in electrical energy production and are widely used in wind turbines and hydroelectric power plants, making it extremely important to ensure their reliable operation. The inevitable reality of faults occurring in generators due to mechanical, electrical, or magnetic causes necessitates the development of early detection and preventive maintenance methods. If faults are not detected early, serious consequences such as production loss, increased maintenance costs, and safety risks can arise (Çınar et al., 2020; Zhang et al., 2020). Fault diagnosis processes become even more challenging, especially in systems operating at variable speeds, due to the complexity of time-varying electrical signals (Çınar et al., 2020). Recent research shows that Machine Learning (ML) technologies make fault recognition processes more effective and carry the potential to reduce maintenance times (Nguyen & Medjaher, 2019; Yedla et al., 2024). Consequently, the application of ML-based methods in synchronous generators used for electrical energy production has become critically important not only for increasing system reliability but also for reducing costs (Silva et al., 2021; Zhang et al., 2020). Although traditional fault detection methods are carried out using techniques such as thermal, vibration, and acoustic analysis, the main disadvantages of these methods include being time-consuming, dependent on experience, and having low generalization capabilities (Chen et al., 2023; Tang et al., 2020a). Especially in synchronous generators operating at variable speeds, the unstable nature of electrical signals reduces the effectiveness of classical methods and leads to their inadequacy in early fault detection (Tama et al., 2022). The specified limitations of traditional methods cause a decrease in the quality of fault recognition processes, while new technologies, and particularly ML-based solutions, are expected to fill these gaps. ML applications can automatically extract features from datasets, thereby significantly speeding up fault detection (Gong et al., 2019; Ma et al., 2019). Nowadays, the use of Deep Learning (DL) techniques has led to significant advancements in the field of fault recognition for rotating machines and has contributed to a more accurate assessment of the current situation (Tama et al., 2022; Tang et al., 2020b). ML algorithms play a significant role in fault diagnosis processes thanks to their ability to model complex and non-linear relationships. By overcoming the limitations of traditional methods, considerably more successful results can be achieved through their capacity to learn from very large datasets (Lei et al., 2020). Particularly, the use of DL techniques has become more effective in detecting complex patterns and relationships, which enhances the ability to optimize machine maintenance processes and predict faults in advance (Zhao et al., 2019). ML algorithms automatically model based on data, providing more accurate predictions about the system's condition and thus offering the possibility of early fault detection. In this context, the implementation of ML-based solutions is critically important in the fields of industry and engineering for reducing maintenance costs and increasing system reliability (Agrawal & Ali, 2024).

In recent years, a significant number of studies have been conducted on the applications of ML and DL for fault diagnosis in synchronous generators. In these studies, the ability of ML algorithms to model complex non-linear relationships has shown the potential to increase the effectiveness of fault detection processes, overcoming the

This is an open access paper distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial 4.0 International License.

limitations of traditional methods (Lei et al., 2020). Rahnama et al. (2020) succeeded in detecting internal faults through simulations of a synchronous generator, performing realistic numerical modeling in this context. Jiao et al. (2020) presented a comprehensive evaluation of Convolutional Neural Network (CNN) applications for fault recognition in rotating machines and emphasized the importance of DL techniques for fault diagnosis under variable speed conditions. Salomon et al. (2019) examine condition monitoring and predictive maintenance applications for synchronous generators using electrical monitoring techniques, revealing the benefits of ML techniques. Gopinath et al. (2020) stated that they significantly increased the performance of fault diagnosis systems for synchronous generators with developed feature mapping techniques, which is another example supporting the applicability of ML-based systems. Xu et al. (2025) showcased successful applications of a fault diagnosis model they developed using ResNet-18 architecture. In addition to ML applications, the use of DL techniques is particularly noteworthy. Peng et al. (2019) presented an innovative approach in the fault diagnosis of high-speed train wheelset bearings using deep CNN methods. Overall, these studies demonstrate that the use of ML and DL techniques in the fault diagnosis of synchronous generators is increasingly prevalent and that these approaches have high potential to enhance reliability and effectiveness (Mazaheri-Tehrani & Faiz, 2021).

In this study, the dataset used was obtained from the work conducted by Tominaga et al. (2024). The some of the data sets created for fault diagnosis of variable speed synchronous generator were used. Considering that measurements were taken at different speeds, the dataset includes two main conditions: synchronous generators operating in a healthy state and a faulty state. Three different fault types, namely phase-phase, same coil turns, and same phase coils, are included within the dataset. Another important feature is that this dataset has been pre-filtered. This indicates that additional preprocessing steps (Z-score, normalization, feature selection) did not contribute to the study, as the existing dataset already has a configuration suitable for analysis. In the study by Tominaga et al. (2025a), the model training was conducted entirely with simulation data, while real measurement data was only used in the model validation phase. In contrast, in the approach proposed in this article, both the training and testing phases were carried out entirely with real experimental data. Thus, the performance of the developed models was evaluated based on actual system behavior, not a simulation environment. In the study by Tominaga et al. (2025b), the authors tested their developed "Tree TCN" and "Tree LSTM" models on different datasets, and their performance is worse compared to the results obtained in this study. The main objective of this study is to automatically classify whether synchronous generators operating at different speeds are working in a faulty or healthy state using machine learning algorithms. The aspect of the study is performing a simple yet effective binary classification, classifying only as 'faulty' and 'healthy' without differentiating between fault types. In this study, the aim is to compare five different machine learning algorithms: decision trees, random forest, KNN, SVM, and logistic regression.

II. MATERIAL AND METHOD

This section describes the materials and methods used in the study. In this context, steps such as dataset description and classification are presented.

A. Dataset Description

The dataset employed in this study was sourced from the work conducted by Tominaga et al. (2024). The dataset comprises real electrical current and voltage series obtained from a test setup established in a laboratory environment. The dataset used contains measurements from a four-pole, salient-pole synchronous generator system operating at variable speeds. During the data acquisition process, physical quantities such as voltage, current, torque, speed, active, and reactive power were measured using high-accuracy sensors which are LEM LV 20-P, LEM LTS 25-NP, and the signals were passed through low-pass filters with a 2 kHz cutoff frequency. The variable-speed generator dataset consists of files, each representing different fault conditions, and a part of this dataset was used. There are 51 different features in the dataset; among these are parameters that characterize the generator's electrical and mechanical behavior, such as phase currents, phase voltages, field current, fault current, torque, rotor speed, and active and reactive power measurements. This data was collected under both healthy conditions and three different internal fault types which are phase-to-phase, short circuits between turns of the same coil, and short circuits between coils of the same phase. In this study, the dataset was labeled as "faulty" and "healthy".

B. Pre-Processing Stage

Various preprocessing steps were applied to the data to measure the performance of ML algorithms. In this study, three fundamental preprocessing methods were evaluated: normalization, Z-Score standardization, and feature selection.

Normalization prevents variables with large value ranges from dominating the model by transforming all features to the same scale (0–1). The min-max normalization method used in this study was applied as given in Eq.1 (Guo et al., 2025):

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

where x presents the original feature value, and $x_{\min}$and $x_{\max}$denote the minimum and maximum values of the corresponding feature, respectively. The same normalization parameters obtained from the training set were subsequently applied to the validation and test sets to prevent data leakage and ensure fair model evaluation.
Thus, all input features were evaluated on the same scale, optimizing the algorithms' convergence speed and classification performance.

Z-Score standardization normalizes data distributions by transforming the meaning of each feature to zero and the standard deviation to one. The following equation was used with Eq.2 (Chen & Huang, 2025):

$$Z = \frac{X - \mu}{\sigma} \quad (2)$$

where Xdenotes the original value of the feature, μ represents the mean value of the corresponding feature calculated from the training dataset, and σ is the standard deviation of that feature. This standardization ensures that all features contribute equally to the learning process by removing scale-related biases, particularly in models sensitive to features.

The univariate feature selection method was used. This method independently evaluates each feature based on its relationship with the target variable and selects the features with the highest statistical significance. The criterion used in the application is based on the F-test, f-regression, which measures the strength of the linear relationship between each feature and the target variable (Xu et al., 2025). Based on the resulting F-scores, the top 10 features with the highest correlation were selected. This approach, without considering multivariate interactions, statistically analyzes the individual impact of each feature on the target variable, ensuring the elimination of redundant or low-impact variables and enhancing the model's computational efficiency. As a result of the feature selection process, the most significant parameters that maximize classification performance and best represent the system characteristics were identified. The best 10 features are the field voltage, grid d-axis current reference, grid d-axis current, generator q-axis voltage, generator A-phase current, grid active power, grid modulation index, generator-side converter A-phase voltage, generator-side converter C-phase voltage, and grid-side C-phase voltage.

C. Classification Stage

In this study, five different machine learning algorithms were used to classify the faulty or fault-free operation of variable-speed synchronous generators: Decision tree, random forest, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and logistic regression. The decision tree algorithm is a classification method with a trunk-branch-leaf (tree) structure that divides data into subgroups using decision rules. At each node, the data is split according to a specific feature, and this process continues until the highest information gain, or the lowest Gini impurity is achieved. The resulting leaves represent the class labels. This method is a model with high interpretability, fast, and capable of direct decision-making (Charbuty & Abdulazeez, 2021). In this study, the default "gini" criterion was preferred as the model's split criterion. The Random Forest is a method that works by creating multiple decision tree models and combining their results. Each tree is trained on subsets of the data obtained through bootstrap sampling. In the prediction phase, the final class decision is made by taking the votes of all trees. This structure reduces the risk of overfitting compared to a single decision tree and produces more stable, generalizable results (Biau & Scornet, 2016). In this study, 100 decision trees were created by default, and the "gini" criterion was used as the split criterion. The KNN algorithm performs classification based on the assumption that similar instances will be close to each other. When a new instance is to be classified, the K nearest neighbors in the training data are identified, and the class to which most of these neighbors belong is assigned to the new instance. The distance measurement is typically calculated using Euclidean or Minkowski metrics. Despite its simple structure, KNN is an instance-based method that can yield very successful results on well scaled data (Soucy & Mineau, 2001). In this study, the default Euclidean distance was used as the distance metric, and all neighbors were evaluated with equal weight (weights='uniform'). The SVM algorithm seeks to identify an optimal hyperplane that effectively separates data points belonging to two distinct classes. This hyperplane is established so that the margin between the classes is as wide as possible. When the data cannot be linearly separated, kernel functions are employed to project the input data into a higher-dimensional feature space, enabling linear separation in that space (Zhang, 2012). In this research, the RBF kernel was utilized as the default kernel type. The flexibility of the decision boundary was adjusted using the regularization parameter 'C', ensuring a balance between margin maximization and model complexity. The logistic regression, on the other hand, is a probabilistic linear classification approach that models the relationship between predictor variables and the dependent variable. It estimates the likelihood that a given sample belongs to a particular class using the logistic (sigmoid) function, and classification decisions are made based on these probabilities (Bisong, 2019). In this study, increasing the maximum iteration limit was effective in preventing convergence issues, while L2 (Ridge) regularization was adopted to control overfitting and improve model stability.

III. RESULTS AND DISCUSSIONS

In this section, the findings of the study are examined and discussed. The data related to the results are presented comparatively in tabular form, and the accuracy rates of the findings are explained in detail.

A. *Performance Measures*

In this study, the performance of the machine learning algorithms was assessed based on the confusion matrix. This matrix represents the comparison between the predicted and actual class labels and includes four essential components: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). The structure of the confusion matrix is presented in Figure 1. Additionally, the Area Under the Curve (AUC) metric, derived from the ROC (Receiver Operating Characteristic) curve, was employed to evaluate the models. The ROC curve illustrates how the True Positive Rate (TPR) varies with the False Positive Rate (FPR). A higher AUC value, approaching 1, indicates a stronger capability of the model to distinguish between the classes. Based on the confusion matrix values, several statistical indicators were computed to measure classification accuracy and reliability (Önder et al., 2023):

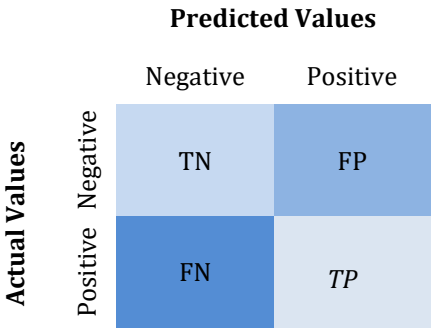


Figure 1. Confusion matrix.

$$Classification\ accuracy\ (\%) = \frac{TP+TN}{TP+FP+TN+FN} \times 100 \quad (3)$$

$$Sensitivity\ (\%) = \frac{TP}{TP+FN} \times 100 \quad (4)$$

$$Specificity\ (\%) = \frac{TN}{FP+TN} \times 100 \quad (5)$$

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

$$f\text{-}measure = \frac{2 \times precision \times recall}{precision+recall} \quad (8)$$

$$True\ Positive\ Rate\ (TPR) = \frac{TP}{TP+FN} \quad (9)$$

$$False\ Positive\ Rate\ (FPR) = \frac{FP}{TN+FP} \quad (10)$$

$$Area\ Under\ Curve\ (AUC) = \int TPR\ d(FPR) \quad (11)$$

$$Balanced\ Accuracy\ Score\ (BAS)\ (\%) = \frac{Sensitivity + Specificity}{2} \quad (12)$$

$$Matthews\ Correlation\ Coefficient\ (MCC) = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP+FP) \times (TP+FN) \times (TN+FP) \times (TN+FN)}} \quad (13)$$

B. *Results and Discussion*

In this section, the preliminary results obtained from the raw dataset are first presented. Subsequently, after applying the preprocessing procedures including normalization, Z-Score standardization, and feature selection the performance outcomes of the decision tree, random forest, KNN, SVM, and logistic regression algorithms are evaluated and systematically compared with those derived from the raw data. In all studies, the data set was randomly divided into 70% training and 30% test data.

Table 1. Performance metrics of the classification algorithms when applied to the raw dataset

Classification Algorithm	Accuracy (%)	BAS (%)	Precision	Recall	f-measure	MCC
Decision Tree	**99.434**	**98.743**	**0.997**	**0.997**	**0.997**	**0.975**
Random Forest	98.362	94.522	0.984	0.998	0.991	0.928
K-Nearest Neighbors	93.771	81.580	0.948	0.983	0.965	0.711
Support Vector Machine	93.923	79.490	0.941	0.992	0.966	0.714
Logistic Regression	91.071	68.100	0.910	0.995	0.951	0.548

Table 2. Performance metrics of the classification algorithms when applied to the normalized dataset

Classification Algorithm	Accuracy (%)	BAS (%)	Precision	Recall	f-measure	MCC
Decision Tree	99.432	98.741	0.997	0.997	0.997	0.975
Random Forest	98.361	94.520	0.984	0.998	0.991	0.928
K-Nearest Neighbors	93.771	81.582	0.948	0.983	0.965	0.711
Support Vector Machine	93.920	79.494	0.941	0.992	0.966	0.714
Logistic Regression	91.083	68.133	0.910	0.995	0.951	0.549

Table 1 presents the performance metrics of the classification algorithms on the raw dataset. According to the results, the decision tree algorithm exhibited the highest performance, achieving an accuracy of 99.43% and MCC value of 0.975. The random forest algorithm demonstrated similarly high accuracy 98.36% and stability. The KNN and SVM models obtained moderate performance with accuracy rates around 93%, while logistic regression yielded the lowest results 91.07%. This table reveals that the DT and RF algorithms possess a robust generalization capability even with raw data. Table 2 presents the performance metrics obtained after the application of normalization. The resulting values are nearly identical to those from the raw data; no significant changes were observed in the accuracy rates or MCC values. This suggests that the differences in feature scales within the dataset do not have a substantial impact on classification performance, indicating that normalization did not improve the models' success. The results in Table 3, where Z-Score Standardization was applied, also almost completely aligned with the raw data results, like the normalization step. The accuracy, F-measure, and MCC values for all algorithms remained unchanged. This finding indicates that the dataset was already well-balanced and appropriately scaled, and consequently, Z-Score standardization provided no additional contribution. Following the application of feature selection (Table 4), minor alterations in the performance of some algorithms were observed. The decision tree and random forest algorithms remained the most successful models, achieving 99.24% and 99.14% accuracy, respectively. However, slight shifts were noted in their MCC values which are DT: 0.967; RF: 0.963. This situation indicates that while feature selection provided some degree of model simplification, it did not have a pronounced effect on overall accuracy. Conversely, the other algorithms which are KNN, SVM, logistic regression experienced a minor degradation in performance, implying that the reduced number of selected features may have limited the discriminative power of certain complex models.

Table 3. Performance metrics of the classification algorithms when applied to the Z-Score Standardized dataset

Classification Algorithm	Accuracy (%)	BAS (%)	Precision	Recall	f-measure	MCC
Decision Tree	99.434	98.742	0.997	0.997	0.997	0.975
Random Forest	98.363	94.531	0.984	0.998	0.991	0.928
K-Nearest Neighbors	93.782	81.593	0.948	0.983	0.965	0.711
Support Vector Machine	93.921	79.492	0.941	0.992	0.966	0.714
Logistic Regression	91.070	68.080	0.910	0.995	0.951	0.548

Table 4. Performance metrics of the classification algorithms after applying Feature Selection

Classification Algorithm	Accuracy (%)	BAS (%)	Precision	Recall	f-measure	MCC
Decision Tree	99.244	98.374	0.996	0.996	0.996	0.967
Random Forest	99.142	97.460	0.993	0.998	0.995	0.963
K-Nearest Neighbors	92.360	77.951	0.938	0.977	0.957	0.640
Support Vector Machine	91.671	69.821	0.915	0.997	0.954	0.586
Logistic Regression	90.552	66.043	0.905	0.995	0.948	0.514

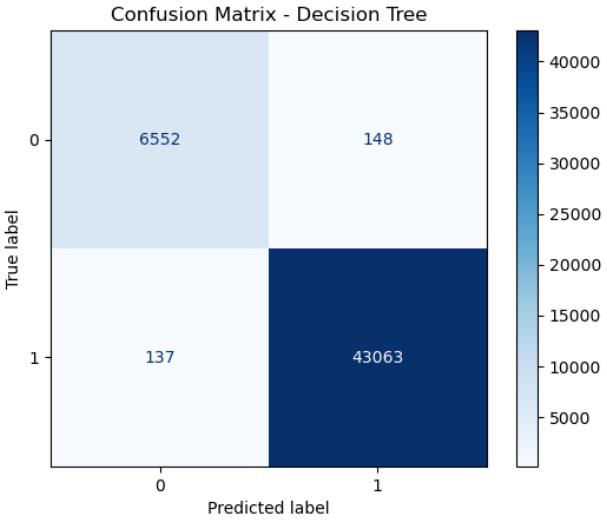


Figure 2. Confusion matrix of the best classification algorithm (decision tree with raw dataset)

Figure 1 shows the confusion matrix for the decision tree algorithm trained on the raw dataset. The matrix reveals that both the 'healthy' and 'faulty' classes were classified almost flawlessly. The extremely low count of misclassifications, minimal FN and FP, indicates that the model possesses a high generalization ability. This outcome demonstrates that the decision tree, particularly with this dataset, was able to learn the decision rules effectively and perform accurate classifications. Figure 2 shows the ROC curve for the decision tree algorithm. The proximity of the curve to 1 indicates that the AUC value is very close to 1, demonstrating that the model can achieve a near-perfect distinction between the classes. This finding shows that despite the simple structure of the decision tree, it exhibits a highly balanced performance in terms of both sensitivity and specificity.

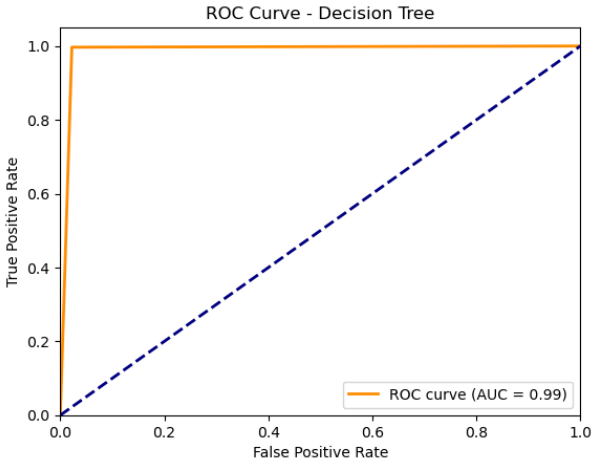


Figure 3. ROC curve of the best classification algorithm (decision tree with raw dataset)

The decision tree algorithm appears to perform best on both raw and preprocessed datasets. Because, it's ability to model nonlinear relationships and complex feature interactions without requiring assumptions about data distribution. Its built-in feature selection and robustness to feature scaling enabled more effective learning from the dataset, leading to improved predictive performance. The random forest model offers a similarly robust alternative, while the performance of KNN, SVM, and logistic regression algorithms is relatively lower. Scaling methods such as normalization and Z-Score standardization did not significantly impact performance; the dataset was already balanced. While feature selection provided a small degree of performance stability by reducing model complexity, it did not significantly improve overall accuracy. These findings demonstrate that the decision tree algorithm is the most suitable and reliable method for fault diagnosis in variable-speed synchronous generators.

IV. CONCLUSION

This study compared the performance of five different machine learning algorithms: decision tree, random forest, KNN, SVM, and logistic regression, for fault detection in variable-speed synchronous generators. The dataset used in the study consisted of experimental measurement data and included both healthy and faulty generator operating conditions. The results showed that the decision tree algorithm had the highest accuracy rate of 99.43% and the strongest correlation coefficient of MCC=0.975 across all datasets. The random forest model also demonstrated similar high performance. In contrast, the accuracy rates of the KNN, SVM, and LR algorithms were found to be relatively lower. The preprocessing steps of normalization and Z-Score standardization did not significantly improve the performance of the dataset, indicating that the dataset was already balanced in terms of scale. This is thought to be due to signal filtering during dataset generation. Feature selection provided a small degree of stability by reducing model complexity but did not significantly alter overall accuracy. ROC curve and confusion matrix analyses demonstrate that the decision tree model can distinguish between two classes which are healthy and faulty almost without error. Consequently, the decision tree algorithm stands out as the most suitable method for fault detection in variable-speed synchronous generators, offering both high accuracy and low computational complexity. Future studies will focus on extending the proposed framework to multi-class classification using advanced data preprocessing techniques and advanced classification algorithms, aiming to accurately model and discriminate among multi-class fault classification.

DECLARATIONS

Acknowledgements: The authors would like to express their sincere thanks to the editor and the anonymous reviewers for their helpful comments and suggestions.

Author Contributions: Writing original draft, M.O. and M.U.D, writing, review and editing, M.O. and M.U.D. All parts of the study were conducted by the Author.

Conflict of Interest: No potential conflict of interest was declared by authors.

Supporting Institutions: This research received no external funding.

Ethical Approval: This study does not involve human or animal participants. All procedures followed scientific and ethical principles, and all referenced studies are appropriately cited.

Plagiarism Statement: This article has been evaluated for plagiarism and no instances of plagiarism were detected.

Data Availability: Data sharing is not applicable to this study.

Use of AI Tools: The author/authors declare that no Artificial Intelligence (AI) tools were used in the creation of this article.

REFERENCES

Agrawal, E., & Ali, S. (2024). Advancements in machine learning algorithms for enhanced fault analysis and categorization in power systems. *International Journal for Multidisciplinary Research,* *6*(4), 1-18. https://doi.org/10.36948/ijfmr.2024.v06i04.25682

Biau, G., & Scornet, E. (2016). A random forest guided tour. *TEST, 25*, 197–227. https://doi.org/10.1007/s11749-016-0481-7

Bisong, E. (2019). Logistic regression. In *Building machine learning and deep learning models on Google Cloud Platform* (pp. 243–250). Apress. https://doi.org/10.1007/978-1-4842-4470-8_20

Charbuty, B., & Abdulazeez, A. (2021). Classification based on decision tree algorithm for machine learning. *Journal of Applied Science and Technology Trends, 2*(1), 20–28. https://doi.org/10.38094/jastt20165

Chen, J., & Huang, W. (2025). GNSS-R based sea ice classification using track normalized observables. In *Proceedings of*

OCEANS 2025 Brest (pp. 1–4). https://doi.org/10.1109/OCEANS58557.2025.11104660

Chen, X., Yang, R., Xue, Y., Huang, M., Ferrero, R., & Wang, Z. (2023). Deep transfer learning for bearing fault diagnosis: A systematic review since 2016. *IEEE Transactions on Instrumentation and Measurement, 72*, 1–21. https://doi.org/10.1109/tim.2023.3244237

Çınar, Z., Nuhu, A., Zeeshan, Q., Korhan, O., Asmael, M., & Safaei, B. (2020). Machine learning in predictive maintenance towards sustainable smart manufacturing in Industry 4.0. Sustainability, 12(19), Article 8211. https://doi.org/10.3390/su12198211

Gong, W., Chen, H., Zhang, Z., Zhang, M., Wang, R., Guan, C., & Wang, Q. (2019). A novel deep learning method for intelligent fault diagnosis of rotating machinery based on improved CNN-SVM and multichannel data fusion. *Sensors, 19*(7), Article 1693. https://doi.org/10.3390/s19071693

Gopinath, R., Kumar, C., Vishnuprasad, K., & Ramachandran, K. (2020). Feature mapping techniques for improving the performance of fault diagnosis of synchronous generator. *International Journal of Prognostics and Health Management, 6*(2), 1-12. https://doi.org/10.36001/ijphm.2015.v6i2.2256

Guo, J., Li, X., Wang, Y., Zhang, H., & Liu, J. (2025). Robust multi-task adversarial attacks using min-max optimization. In *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1–5). https://doi.org/10.1109/ICASSP49660.2025.10888541

Jiao, J., Zhao, M., Lin, J., & Liang, K. (2020). A comprehensive review on convolutional neural network in machine fault diagnosis. *Neurocomputing, 417*, 36–63. https://doi.org/10.1016/j.neucom.2020.07.088

Lei, Y., Yang, B., Jiang, X., Jia, F., Li, N., & Nandi, A. (2020). Applications of machine learning to machine fault diagnosis: A review and roadmap. *Mechanical Systems and Signal Processing, 138*, Article 106587. https://doi.org/10.1016/j.ymssp.2019.106587

Ma, P., Zhang, H., Fan, W., Wang, C., Wen, G., & Zhang, X. (2019). A novel bearing fault diagnosis method based on 2D image representation and transfer learning-convolutional neural network. *Measurement Science and Technology, 30*(5), Article 055402. https://doi.org/10.1088/1361-6501/ab0793

Mazaheri-Tehrani, E., & Faiz, J. (2021). Airgap and stray magnetic flux monitoring techniques for fault diagnosis of electrical machines: An overview. *IET Electric Power Applications, 16*(3), 277–299. https://doi.org/10.1049/elp2.12157

Nguyen, K., & Medjaher, K. (2019). A new dynamic predictive maintenance framework using deep learning for failure prognostics. *Reliability Engineering & System Safety, 188*, 251–262. https://doi.org/10.1016/j.ress.2019.03.018

Önder, M., Dogan, M. U., & Polat, K. (2023). Classification of smart grid stability prediction using cascade machine learning methods and the internet of things in smart grid. *Neural Computing & Applications, 35*, 17851–17869. https://doi.org/10.1007/s00521-023-08605-x

Peng, D., Liu, Z., Wang, H., Qin, Y., & Jia, L. (2019). A novel deeper one-dimensional CNN with residual learning for fault diagnosis of wheelset bearings in high-speed trains. *IEEE Access, 7*, 10278–10293. https://doi.org/10.1109/access.2018.2888842

Rahnama, M., Vahedi, A., Mohammad-Alikhani, A., & Takorabet, N. (2020). Diagnosis of brushless synchronous generator using numerical modeling. *Compel - The International Journal for Computation and Mathematics in Electrical and Electronic Engineering, 39*(5), 1241–1254. https://doi.org/10.1108/compel-01-2020-0018

Salomon, C. P., Ferreira, C., Sant'Ana, W. C., Lambert-Torres, G., Borges da Silva, L. E., Bonaldi, E. L., de Oliveira, L. E. d. L., & Torres, B. S. (2019). A study of fault diagnosis based on electrical signature analysis for synchronous generators predictive maintenance in bulk electric systems. *Energies, 12*(8), Article 1506. https://doi.org/10.3390/en12081506

Silva, A., Cortez, P., Pereira, C., & Pilastri, A. (2021). Business analytics in Industry 4.0: A systematic review. *Expert Systems, 38*(7), Article e12741. https://doi.org/10.1111/exsy.12741

Soucy, P., & Mineau, G. W. (2001). A simple KNN algorithm for text categorization. In *Proceedings 2001 IEEE International Conference on Data Mining* (pp. 647–648). https://doi.org/10.1109/ICDM.2001.989592

Tama, B. A., Vania, M., Lee, S., & Lim, S. (2022). Recent advances in the application of deep learning for fault diagnosis of rotating machinery using vibration signals. *Artificial Intelligence Review, 56*(5), 4667–4709. https://doi.org/10.1007/s10462-022-10293-3

Tang, S., Yuan, S., & Zhu, Y. (2020a). Convolutional neural network in intelligent fault diagnosis toward rotatory machinery. *IEEE Access, 8*, 86510–86519. https://doi.org/10.1109/access.2020.2992692

Tang, S., Yuan, S., & Zhu, Y. (2020b). Deep learning-based intelligent fault diagnosis methods toward rotating machinery. *IEEE Access, 8,* 9335–9346. https://doi.org/10.1109/access.2019.2963092

Tominaga, R. N., Sousa, L. A., Rocha, R. V., Monaro, R. M., Ávila, S. L., de Camargo Salles, M. B., & Carmo, B. S. (2024). Electrical signals dataset from fixed-speed and variable-speed synchronous generators under healthy and faulty conditions. *Data in Brief, 57*, Article 111018. https://doi.org/10.1016/j.dib.2024.111018

Tominaga, R. N., Rocha, R. V., Avila, S. L., Monaro, R. M., Salles, M. B. C., & Carmo, B. S. (2025a). Simulation-based neural network for robust short-circuit detection in wind turbines. In *2025 International Conference on Clean Electrical Power (ICCEP).* https://doi.org/10.1109/ICCEP65222.2025.11143686

Tominaga, R. N., Rocha, R. V., Avila, S. L., Monaro, R. M., Salles, M. B. C., & Carmo, B. S. (2025b). Modular and compact neural network framework for internal fault detection in generators using current signature data. In *2025 IEEE*

International Electric Machines & Drives Conference (IEMDC). https://doi.org/10.1109/IEMDC60492.2025.11061103

Xu, J., Gao, L., Liu, Z., & Yun, Q. (2025). Research on fault diagnosis method of synchronous generator based on ResNet-18. *Journal of Physics Conference Series, 2963*(1), Article 012021. https://doi.org/10.1088/1742-6596/2963/1/012021

Xu, J., Wang, X., Lu, J., & Hou, H. (2025). Xu weight is all you need! Short-term power load forecasting based on a novel adaptive feature selection method and Xu weight. In *2025 IEEE 4th International Conference on Computing and Machine Intelligence (ICMI)* (pp. 1–8). https://doi.org/10.1109/ICMI65310.2025.1114133

Yedla, H., Koppada, L. R., & Bodala, R. S. (2024). Advanced battery management: Forecasting health, state of charge & maintenance needs using AI & ML models (LSTM, gradient boosting, SVR, random forest). *Asian Journal of Research in Computer Science, 17*(8), 46-57. https://doi.org/10.9734/ajrcos/2024/v17i7489

Zhang, S., Zhang, S., Wang, B., & Habetler, T. (2020). Deep learning algorithms for bearing fault diagnostics: A comprehensive review. *IEEE Access, 8*, 29857–29881. https://doi.org/10.1109/access.2020.2972859

Zhang, Y. (2012). Support vector machine classification algorithm and its application. In C. Liu, L. Wang, & A. Yang (Eds.), *Information computing and applications* (pp. 179–186). Springer. https://doi.org/10.1007/978-3-642-34041-3_27

Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., & Gao, R. X. (2019). Deep learning and its applications to machine health monitoring. *Mechanical Systems and Signal Processing, 115*, 213–237. https://doi.org/10.1016/j.ymssp.2018.05.050