NOVATEUR PUBLICATIONS
INTERNATIONAL JOURNAL OF INNOVATIONS IN ENGINEERING RESEARCH AND TECHNOLOGY
[IJIERT] ISSN: 2394-3696 Website: ijiert.org
VOLUME 10, ISSUE 7, July -2023

DETECTION AND IDENTIFICATION OF ELECTRICAL FAULTS USING RANDOM FOREST CLASSIFICATION

Abdelsalm M.A Ehoedy
Department of Electrical and Electronic Engineering
Tobruk Higher Institute of Science and Technology
ehoedy@gmail.com

Ahmed M Khirala Mohamed
Department of Electrical and Electronic Engineering
Tobruk Higher Institute of Science and Technology
ahmedmkh88@gmail.com

Adhawi Ali Mohamad Elahiwel
Department of Electrical and Electronic Engineering
Higher Institute for Science and Technology Awlad-Ali
Adhawielahiwel@gmail.com

Abstract

Electric power generation and their transmission over electrical power grids and systems are an integral part of human development. It has led to the efficient and steady growth of economies and general human development. Electric power generation and its conveyance over transmission lines are however like every system of engineering prone to faults and errors. The development of machine learning systems has been very instrumental in the detection and classification of phenomena and scenarios in various fields. In this study, we propose the use of a machine learning technique known as random forest classification to carry out a process of electric fault detection and identification using an approach of the binary and multiclass classification process. Using adequate preprocessing and classification, the proposed method in this study achieved a binary classification of fault or no-fault classification of 99.6% accuracy, and a multiclassification of type of electric fault identification performance of 89.45% accuracy. The proposed method, tools, and analysis carried out in this study are presented in this paper comprehensively.

Keywords - electrical faults, electric fault classification, electric fault detection, random forest classifier

1 Introduction

Power transmission lines are a very crucial part of electrical power systems and grids. Since the beginning of the electric development its need and its use in human and economic development have been seen to develop significantly (Liu et al., 2020). Transmission lines are now used to convey electricity from the source of the electricity generation to the distribution network which will eventually reach the end users (Qoria, 2020). Electrical power systems are made up of very complex components and elements, and these systems like all other engineering systems have a likelihood of disturbances which in this context are known as electrical faults (Chaari & Al-Maadeed, 2020).

Electricity generation is done at a high capacity at electrical power plants, hence this needs the use of electrical power grids to distribute such power. Power plants and power grids require the use of fault detection in order to detect and track down faults adequately when they happen in order to ensure the safety of both people and equipment (Ehsani et al., 2021). Adequate fault detection in electrical grid systems required good detection processes and fault-type classification processes to make it optimal for processing and management (Liu et al., 2020). The ideal electrical grid and transmission fault detection systems are required to be fast, secure, reliable, and practical. Application of technologies such as computational pattern recognition has been used in recent times to help in the discrimination and detection of electrical power faults in transmission lines and electrical grids (Guo et al., 2022).

1.1 Aim of the Study

The aim and objective of this research are to contribute to the detection of electrical faults through the application of machine learning techniques and classification. This study proposes the use of a random forest classifier to detect electrical faults and to carry out specific classifications of the kind of electrical fault.

2 Literature Review

Electrical power systems are prone to disturbances and fault-causing elements. Machine learning and artificial neural networks (ANN) are very useful state-of-the-art computational tools that are resourceful in the identification and classification of patterns; including electrical fault patterns (Shifat & Hur, 2021). These computational tools have been developed and used in the detection and classification of electrical faults in power systems (Leh et al., 202; Frizzo et al., 2020, Cherif et al., 2020). Efficient and reliable fault detection and classification systems of electricity should be satisfactory relative to speed, accuracy, and practicality of usage. ANNs are computational systems that are regarded as very efficient under the underlined factors determining efficiency. ANNs are characterized by normalization, noise resistance, tolerance, and robustness as computational systems (Shifat & Hur, 2021). Different forms of ANNs have been applied for fault detection and classification on different kinds of datasets in several kinds of research (Saini et al., 2020; Kaur & Kaur, 2020; Jayamaha et al., 2019). Machine learning techniques have various kinds of architectures and models. These different kinds of models and architectures are designed for different purposes of detection and or classification problems (Liu et al., 2020).

3 Method

The proposed technique in this research is based on the classification of electrical faults using the random forest classifier. The classification process is applied to a dataset of electrical faults that were generated based on electrical fault simulations for six different classes of electrical faults. Figure 1 illustrates the proposed research model used in carrying out the classification experiments of this study, and the following subsections present the methods and techniques proposed and used in this study.

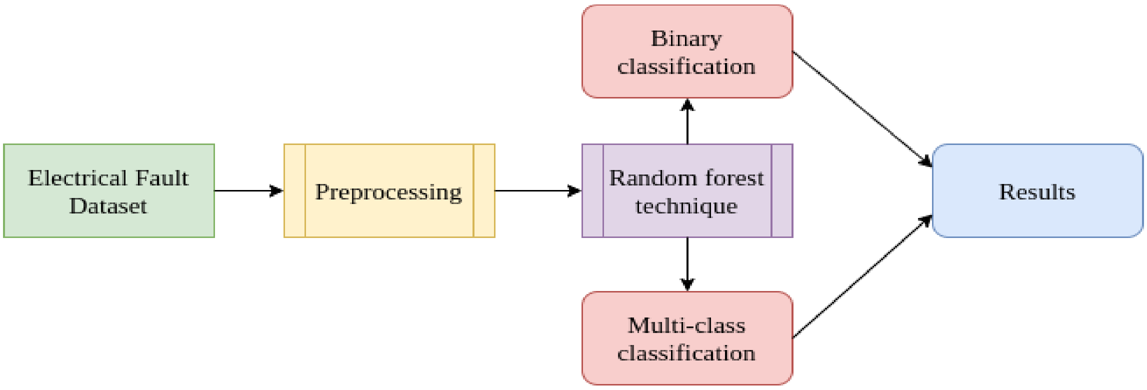

Figure 1: Proposed research model

3.1 Dataset

The dataset used in this study is a dataset of electrical faults that is publicly accessible to the Kaggle community (Prakash, 2021). The dataset is composed of the dataset from simulated data of a model's power grid system created for fault analysis. The simulated power system comprises four power generators and various transmission lines. The dataset consists of about 1200 data points labeled for 6 classes of electrical faults.

3.2 Preprocessing

Data preprocessing is a critical stage in any machine-learning process for classification. In the case of this study, the preprocessing step is carried out to ensure the dataset is free of noisy data such as data duplicates to avoid redundancy, null data cleaning, etc. The need for two different kinds of classification for binary; fault or no-fault, and multiclass; specific kinds of an electrical fault. To adequately ensure this is done the dataset is processed in two different sets, the first set is a grouping of all the data points with the labeling of being electrical faults or not, and the second group of data is for multicast depicting the specific classes of data. Figure 2 and Figure 3 show the distribution of data in both groups of the dataset used in the study. The different classes of electric fault in the dataset are also described as follows: No-Fault, LG fault (Between Phase A and Gnd), LL fault (Between Phase A and Phase B), LLG Fault (Between Phases A, B, and ground), LLL Fault(Between all three phases), LLLG fault(Three phases symmetrical fault).

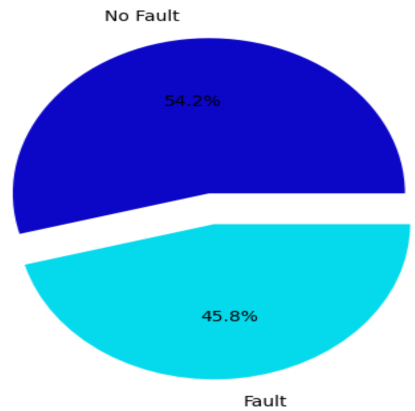

Figure 2: Distribution of binary class; fault and no fault in the dataset

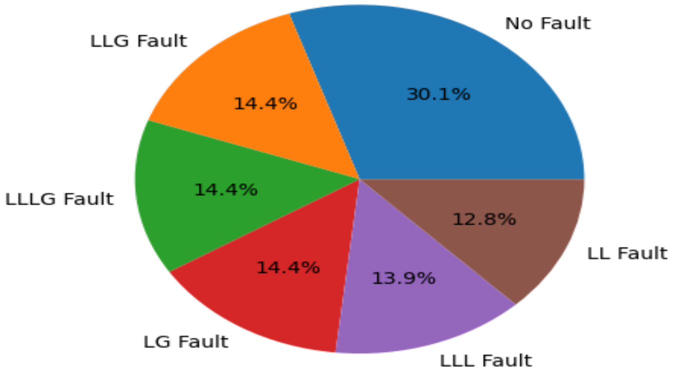

Figure 3: Distribution of all classes of faults in the dataset

3.3 Random forest classifier

Random forest is a machine learning classifier first introduced by Leo Breiman (2001). The classifier is made up of multiple decision trees that operate independently of one another, where an input's class is decided by a compiled voting of the classes assigned to the input by each decision tree in the classifier. The conceptual classification process of random forest is illustrated in Figure 4. Using the enormous amount of decision trees in the random forest classifier ensures an independent classification process for each decision tree in the model (Tyralis et al., 2019). Random forest classifier implements subset sampling and feature selection to improve the classification process, this is a process known as bagging (Breiman, 2001). The bagging process of the random forest technique ensures an adequate computation of the contribution of each feature subset in the ranking of the input data.

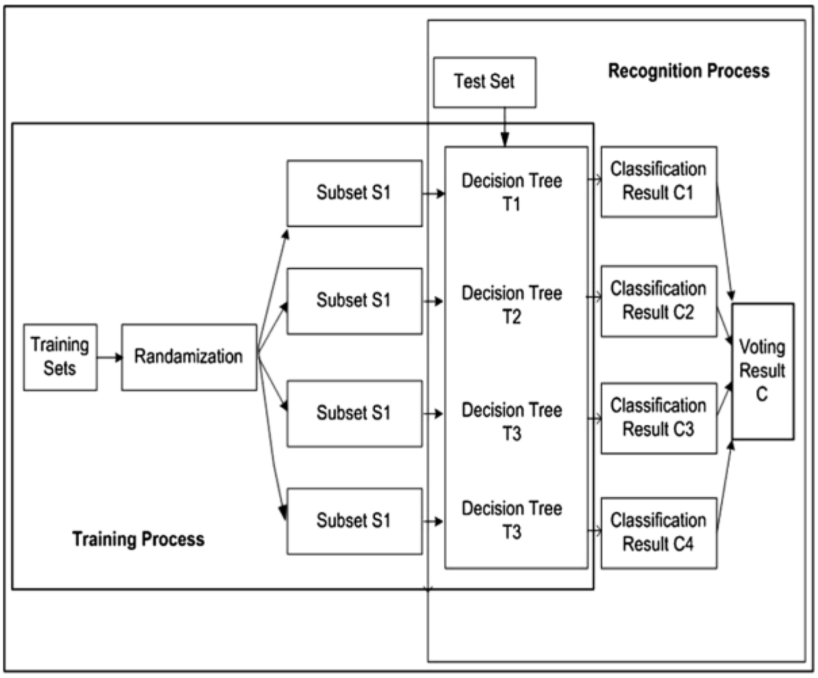

Figure 4: A conceptual framework of random forest classifier

3.4 Evaluation metrics

The classification experiments presented in this study are evaluated based on standard machine learning evaluation metrics. The evaluation criteria for the performance of the proposed method are accuracy, f1-score, recall, and precision. These evaluation metrics are calculated using the true positive (TP), true negative (TN), false positive (FP), and false negative (FN) measures (Jiddah et al., 2021; Jiddah & Yurtkan, 2023). Equations 1, 2, 3, and 4 show the mathematical formula for each evaluation criterion used.

Accuracy = (TN + TP) / (FP + TN + TP + FN) (1)

Recall (Sensitivity) = TP / (TP + FN) (2)

Precision (specificity) = TP / (FP + TP) (3)

F1-score = (precision * Recall) / (Precision + Recall) (4)

4 Results

The results of our experiments are presented for two distinct classification experiments. The first is a binary classification of the data set for fault or no-fault, and the second result is for the classification of all distinct classes of electrical faults in the dataset.

4.1 Binary classification

The binary classification experiments carried out in this study have been observed and reported to have a performance accuracy of 99.6%. The proposed model upon observation of the confusion matrix as illustrated in Figure 5 has misclassified a total of 9 data points. 3 no-fault data points were misclassified as fault data, and 6 fault data points were misclassified as no-fault data points. Table 1 shows the complete performance of the binary classification experiment based on the proposed method of this study.

Table 1: Binary classification performance report

Accuracy	Precision	Recall	F1-score
99.6%	99.72%	99.45	99.65

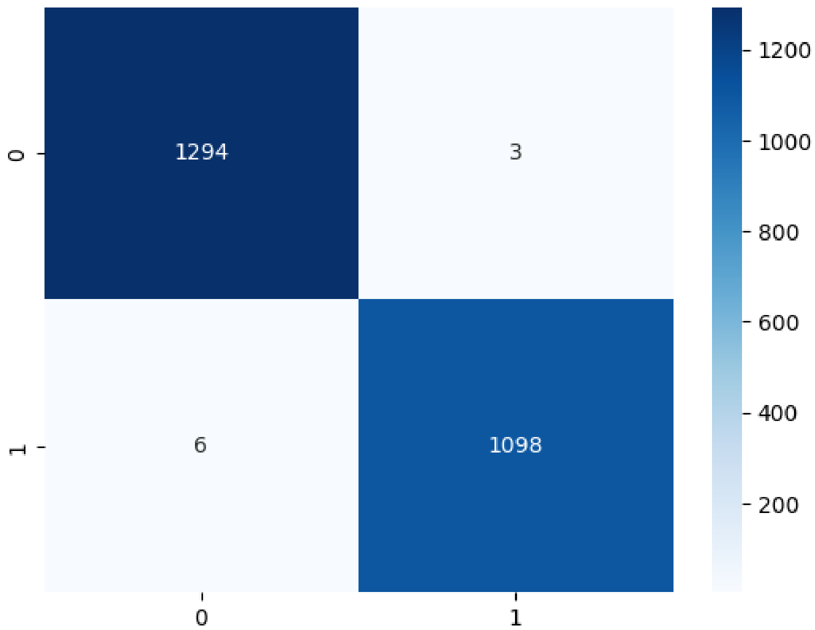

Figure 5: Binary experiment confusion matrix

4.2 Multiclass classification

The multiclass classification experiments carried out in this study have been observed and reported to have a performance accuracy of 89.45%. The multiclass classification carried out in this study has a higher misclassification rate compared to the binary classification. Figure 6 illustrates the confusion matrix of this classification experiment. Table 2 shows the overall performance metrics evaluation report for the multiclass classification using the proposed method.

Table 2: Multiclass classification performance report

Accuracy	Precision	Recall	F1-score
89.45%	87.5%	87.5%	87.5%

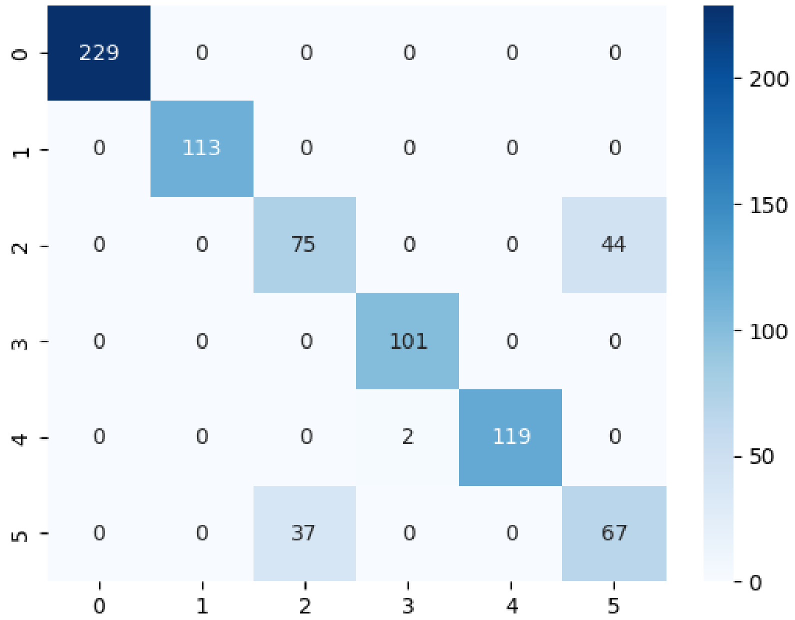

Figure 6: Multiclass experiment confusion matrix

4.3 Discussion

Based on the reported results for both classification experiments carried out in this study, the results of the binary classification have shown a higher accuracy in classification and this can be associated with the lesser number of classes. However, the complexity of the multiclass labels and the need for distinguishing the specific types of electric faults have given less performance accuracy. Observation of the multiclass confusion matrix also reveals the high rate of errors to be the result of misclassification between electric fault types LLL and LLG electric faults. The similarities between these two data points have resulted in over 90% of the misclassification errors in this experiment.

5 Conclusion

The generation, transmission, and distribution of electricity are as important as electricity itself to human development. It is paramount that these processes are ensured to be done with much adequacy and limited or no errors and faults. However, the reality is all systems are prone to errors and faults. Hence, it is important for systems to be put in place to be able to detect and resolve such errors efficiently. This study proposed the use of the machine learning technique known as the random forest classification technique to classify electrical faults. This study carried out the classification of two scenarios; detection of electric faults using binary classification, and identification of electric faults using multiclass classification. The binary classification has shown significant performance accuracy while the multiclass classification shows some significant errors due to the similarities in the electric faults.

5.1 Future work

The observed misclassification error in multiclass electric fault identification experiments requires the development of classification methods that will rid the systems of such misclassification errors. hence, this study proposes future work involving the use of the class cluster technique to further narrow down the classification of electric faults based on similar classes to potentially reduce the error rates between similar classes.

References

1. Liu, J., Zhao, Z., Ji, J., & Hu, M. (2020). Research and application of wireless sensor network technology in power transmission and distribution system. Intelligent and Converged Networks, 1(2), 199-220.
2. Qoria, T. (2020). Grid-forming control to achieve a 100% power electronics interfaced power transmission systems (Doctoral dissertation, HESAM Université).
3. Chaari, M. Z., & Al-Maadeed, S. (2020, February). Wireless power transmission for the internet of things (iot). In 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT) (pp. 549-554). IEEE.
4. Ehsani, M., Singh, K. V., Bansal, H. O., & Mehrjardi, R. T. (2021). State of the art and trends in electric and hybrid electric vehicles. Proceedings of the IEEE, 109(6), 967-984.
5. Guo, E., Jagota, V., Makhatha, M. E., & Kumar, P. (2022). Study on fault identification of mechanical dynamic nonlinear transmission system. Nonlinear Engineering, 10(1), 518-525.
6. Shifat, T. A., & Hur, J. W. (2021). ANN assisted multi sensor information fusion for BLDC motor fault diagnosis. IEEE Access, 9, 9429-9441.
7. Leh, N. A. M., Zain, F. M., Muhammad, Z., Abd Hamid, S., & Rosli, A. D. (2020, August). Fault detection method using ANN for power transmission line. In 2020 10th IEEE international conference on control system, computing and engineering (ICCSCE) (pp. 79-84). IEEE.
8. Frizzo Stefenon, S., Waldrigues Branco, N., Nied, A., Wildgrube Bertol, D., Cristian Finardi, E., Sartori, A., ... & Bartnik Grebogi, R. (2020). Analysis of training techniques of ANN for classification of insulators in electrical power systems. IET Generation, Transmission & Distribution, 14(8), 1591-1597.
9. Cherif, B. D. E., Bendiabdellah, A., & Tabbakh, M. (2020). An automatic diagnosis of an inverter IGBT open-circuit fault based on HHT-ANN. Electric Power Components and Systems, 48(6-7), 589-602.
10. Saini, M., Sultan, A. R., Mustafa, M. W., & Rahimuddin, R. (2020). Faults detection and classification on parallel transmission lines using modified Clarke’s transformation-ANN approach. PRZEGLĄD ELEKTROTECHNICZNY, ISSN 0033-2097, R. 96 NR 4/2020, 96(NR 4), 23-27.
11. Kaur, H., & Kaur, M. (2020). Fault classification in a transmission line using Levenberg–Marquardt algorithm based artificial neural network. In Data Communication and Networks: Proceedings of GUCON 2019 (pp. 119-135). Springer Singapore.
12. Jayamaha, D. K. J. S., Lidula, N. W. A., & Rajapakse, A. D. (2019). Wavelet-multi resolution analysis based ANN architecture for fault detection and localization in DC microgrids. IEEE Access, 7, 145371-145384.
13. Prakash, E. S. (2021, May 22). Electrical fault detection and classification. Kaggle. https://www.kaggle.com/datasets/esathyaprakash/electrical-fault-detection-and-classification?resource=download&select=detect_dataset.csv
14. Breiman, L. (2001). Random forests. Machine learning, 45, 5-32.

15. Tyralis, H., Papacharalampous, G., & Langousis, A. (2019). A brief review of random forests for water scientists and practitioners and their recent history in water resources. Water, 11(5), 910.
16. Jiddah, S. M., Abushakra, M., & Yurtkan, K. (2021). Fusion of geometric and texture features for side-view face recognition using svm. Istatistik Journal of The Turkish Statistical Association, 13(3), 108-119.
17. Jiddah, S. M., & Yurtkan, K. (2023). Dominant and complementary emotion recognition using hybrid recurrent neural network. Signal, Image and Video Processing, 1-9.