Proceedings of the 2nd International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/74/20240484

Electrical fault detection and classification based on multiple machine learning algorithms

Hailang Liu[1,3,*], Xuanyu Liu[2]

[1]School of Electronic Information, Xi'An Polytechnic University, Shanxi, Xi'an, 710699, China
[2]School of Electronic Information, Xi'An Polytechnic University, Shanxi, Xi'an, 710699, China

[3]2278033549@qq.com
[*]Corresponding author

Abstract. Electrical faults in the power system refer to a variety of abnormalities in the power equipment, and these faults may lead to equipment damage or even cause dangerous events such as fires and explosions, which pose a serious threat to people's lives and property. Therefore, it is very important to find and deal with these faults in time. In this paper, based on the electrical fault dataset of power system, SVM, decision tree, KNN and random forest model are used to detect electrical faults. After the confusion matrix results are analyzed, most of the predictions of the four models are correct, and only a few instances that should not be predicted as electrical faults are predicted as electrical faults. All the four machine learning models achieved more than 99% accuracy in detecting electrical faults in the power system and were able to detect electrical faults very accurately. Among them, the SVM model has the highest prediction accuracy of 99.6%, followed by the KNN model prediction accuracy of 99.5%. In summary, this paper adopts machine learning method to detect electrical faults in power systems, and achieves better results in the experimental results. This method can help power system supervisors find and deal with faults in time, and improve the safety and stability of the power system. In the future, we can further study how to optimize the model to improve the detection accuracy and apply it in practical production.

Keywords: Electrical fault detection, SVM, KNN.

1. Introduction

Electrical faults in power systems refer to a variety of abnormalities that occur in power equipment, including short circuits, overloads, undervoltages, and overvoltages [1]. These faults may lead to equipment damage, or even cause dangerous events such as fire and explosion, causing serious threats to people's lives and properties. Therefore, it is very important to find and deal with these faults in time.

The traditional method of electrical fault prediction mainly relies on manual inspection and equipment monitoring. However, this method has problems such as low efficiency, high cost, and insufficient accuracy [2]. In recent years, machine learning algorithms have been introduced into electrical fault prediction and achieved good results. Machine learning algorithms can construct a model by learning a large amount of historical data and predict future data based on the model [3,4]. In electrical fault prediction, machine learning algorithms can use historical data to analyze the correlation between different fault types and conduct comprehensive analysis based on factors such as equipment operation status, environmental temperature, humidity, etc., so as to accurately predict the potential risk of failure [5].

Among them, deep learning algorithm is one of the most widely used machine learning algorithms. Deep learning algorithms can extract the feature information in the data through multi-layer neural networks, so as to realize more accurate prediction. For example, in the power system, deep learning algorithms can be utilized to monitor the operating status of equipment and detect abnormalities in a timely manner [6,7]. In addition to deep learning algorithms, algorithms such as support vector machines and random forests are also widely used in electrical fault prediction. These algorithms have high accuracy and robustness and can realize accurate prediction in different environments.

In conclusion, machine learning algorithms play an important role in electrical fault prediction. In this paper, based on the electrical fault dataset of power system, SVM, Decision Tree, KNN and Random Forest models are used to detect electrical faults, respectively. By analyzing and modeling the historical data, more accurate and efficient fault prediction can be achieved, thus ensuring the safe and stable operation of the power system.

2. Data presentation

An electrical power system is a complex network of multiple electrical devices and lines that normally operate in a state of equilibrium. However, when insulation fails at any point in the system or becomes unbalanced due to contact of energized conductors, a short circuit or fault occurs in the line. Faults can be broadly categorized into two types, symmetrical and asymmetrical faults.

In symmetrical faults, all phases are shorted to each other or to ground (L-L-L) or (L-L-L-G). The nature of this type of fault is a balanced fault. In this type of fault, the fault currents in all phases are symmetrical, i.e., they are equal in magnitude and they are displaced at equal angles, all 120 degrees. This type of fault is more serious, but rarely occurs.

Asymmetrical faults involve only one or two phases. In this type of fault, the three-phase line becomes unbalanced. There are three main types of these faults, namely line-to-ground (L-G), line-to-line (L-L), and double line-to-ground (LL-G) faults. Most of these types of faults occur in power systems. Among them, line-to-ground (L-G) is one of the most common asymmetrical fault types. It is usually caused by insulation damage, falling trees or bird short circuits. When a line-to-ground fault occurs, current flows from the phase to ground, causing an imbalance in the power system. Line-to-line (L-L) faults are usually caused by insulation failure between two phases or contact between two phases. This type of fault causes an increase in the voltage difference between the two phases, which leads to an unbalance in the power system. A double line-to-ground (LL-G) fault, on the other hand, is a short circuit of two lines to ground at the same time. This type of fault is usually caused by falling trees or other external disturbances. The various electrical faults are shown in Figure 1.

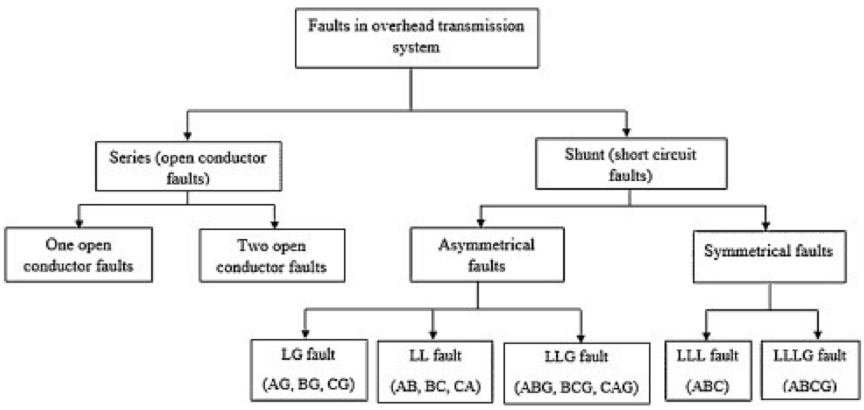


Figure 1. Classification of electrical faults.
(Photo credit : Original)

DOI: 10.54254/2755-2721/74/20240484

In the dataset Ia, Ib, Ic, Va, Vb and Vc are the input variables and GCBA is the output variable, each type of fault and the corresponding GCBA number is shown in Table 1.

Table 1. Data parameter.

G	C	B	A	Type
0	0	0	0	No faults
1	0	0	1	LG faults
0	0	1	1	LL faults
1	0	1	1	LLG faults
0	1	1	1	LLL faults
1	1	1	1	LLLG faults

3. Method

3.1. Support vector machine
Support Vector Machine is a binary classification model which is based on the principle of mapping data into a high-dimensional space and finding an optimal hyperplane to partition different classes of data [8].SVM model has strong generalisation ability and robustness, and performs excellently in dealing with high-dimensional data and nonlinear data.SVM can also deal with nonlinear problems by using kernel functions, and commonly used kernels are linear kernels, polynomial kernels and Radial Basis Function (RBF) kernel, etc.

3.2. Decision tree
Decision tree modelling is a method of decision analysis based on a tree structure. Its principle is to divide the dataset into multiple subsets by splitting the features, and finally get a tree structure. Decision tree model has the advantages of being highly interpretable, easy to understand and use, and performs well in dealing with classification and regression problems. However, decision trees are prone to overfitting problems and need to be optimised by methods such as pruning.

3.3. KNN
KNN model is a classification algorithm based on sample distance metric. Its principle is to find the K samples in the training set that are closest to the test sample and classify the test sample into the category with the most occurrences among the K samples.The KNN model is simple and easy to understand, requires no training, and other features, and it performs well when dealing with small-scale datasets [9]. However, the KNN model is computationally intensive for large-scale datasets and requires preprocessing operations such as data normalisation.

3.4. Random forest
Random forest model is an integrated learning method based on decision trees. Its principle is to improve the generalisation ability and robustness of the model by constructing multiple decision trees and using methods such as random sampling and feature selection. Random forest model has the advantages of high accuracy, robustness, and the ability to deal with high dimensional data, and performs well in dealing with classification and regression problems [10]. However, the random forest model is more sensitive to noisy data and requires preprocessing operations such as data cleaning.

4. Experiments and Results
After importing the data, the dataset is divided according to the ratio of 4:3:3, 40% of the data is used for model training, 30% for validation and 30% for testing. SVM, Decision Tree, KNN and Random

DOI: 10.54254/2755-2721/74/20240484

Forest models were imported respectively, binary classification was performed according to whether electrical faults occurred or not, the accuracy of each model prediction was calculated and the confusion matrix was output.

For the model parameter settings, the number of training rounds was set to 100, the learning rate was set to 0.0001, and the depth of the tree was set to 10. The accuracy of the prediction of each model is shown in Table 2 and Fig. 2. In addition, the tree structure of the decision tree model is output, as shown in Figure 4.

Table 2. Modelling evaluation.

Model	Accuracy
SVM	99.6
Decision tree	99.3
KNN	99.5
Random forest	99

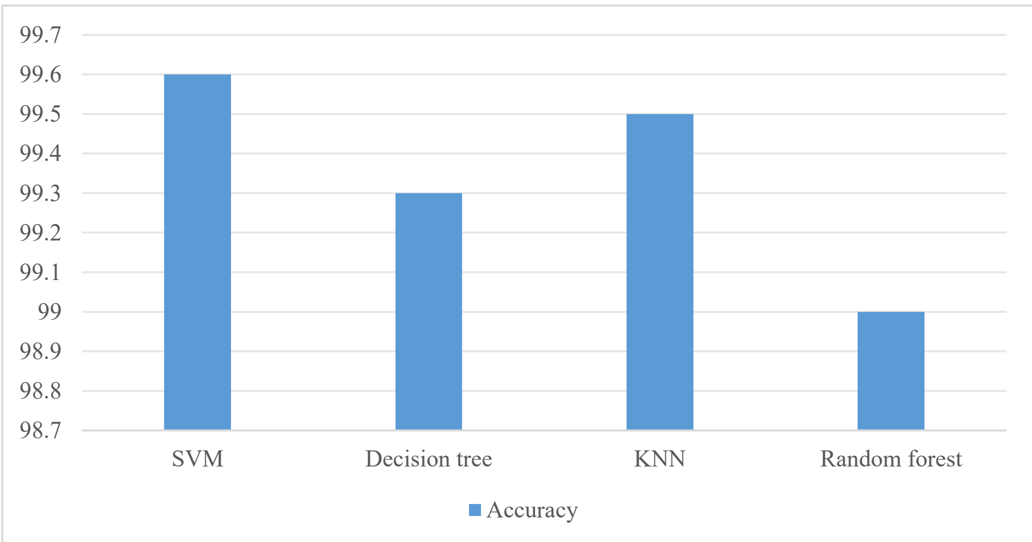
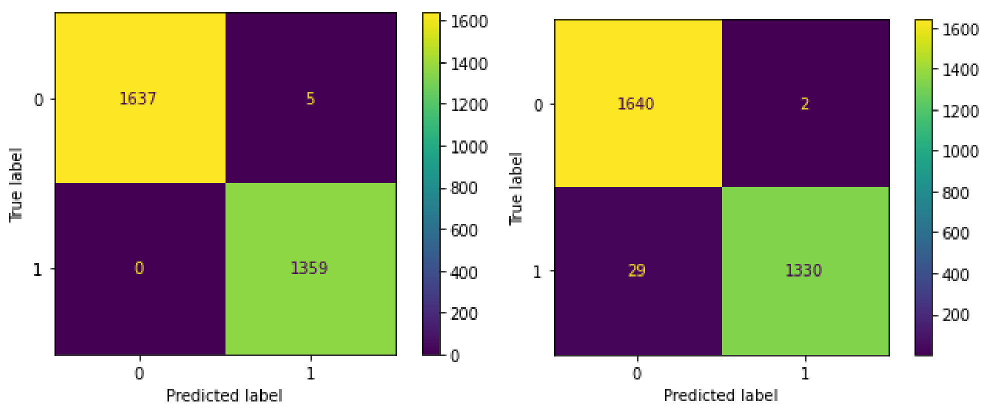


Figure 2. Modelling evaluation.
(Photo credit: Original)

The confusion matrix of SVM, Decision Tree, KNN and Random Forest models for predicting electrical faults is shown in Fig. 3.



DOI: 10.54254/2755-2721/74/20240484

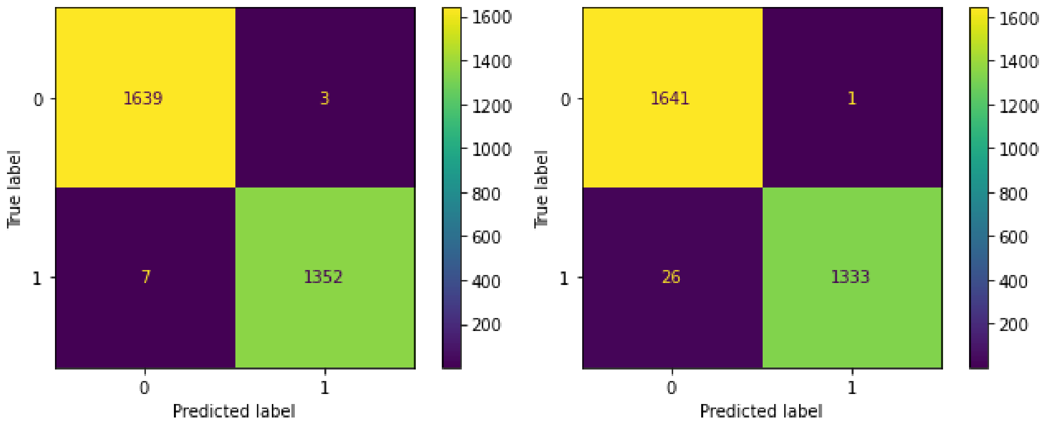


Figure 3. Confusion matrix.
(Photo credit : Original)

From the above figure, it can be seen that for the SVM model, 5 instances that should have been predicted as electrical faults were predicted as non-electrical faults. For the decision tree model, 2 instances that should have been predicted as electrical faults were predicted as non-electrical faults, while 29 instances that should have been predicted as non-electrical faults were predicted as electrical faults. For the KNN model, 3 instances that should have been predicted as electrical faults were predicted as non-electrical faults, while 7 instances that should have been predicted as non-electrical faults were predicted as electrical faults. For the Random Forest model, 2 instances that should have been predicted as electrical faults were predicted as non-electrical faults, while 29 instances that should have been predicted as non-electrical faults were predicted as electrical faults.

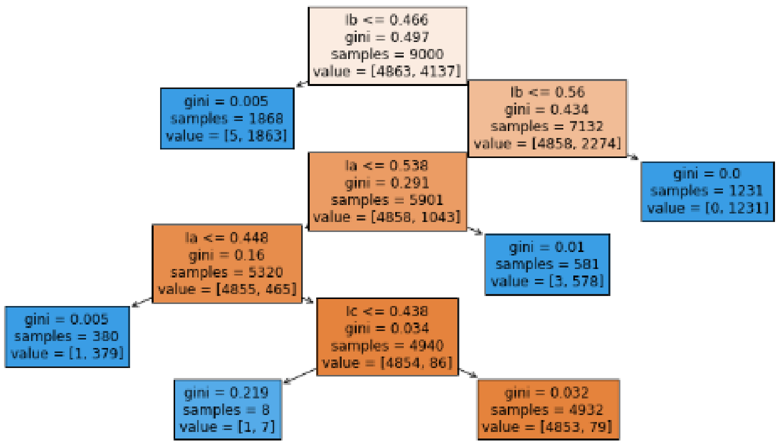


Figure 4. Decision tree structure.
(Photo credit: Original)

From the confusion matrix results, it can be seen that most of the prediction results of the four models are correct, and only a few instances that should not be predicted as electrical faults are predicted as electrical faults. From the prediction results of the four machine learning models, it can be seen that the accuracy of all four models in detecting electrical faults in the power system reaches more than 99%, and they can detect electrical faults very accurately, among which, the SVM model has the highest prediction accuracy of 99.6%, followed by the KNN model prediction accuracy of 99.5%.

5. Conclusion

Electrical faults in power systems are abnormal conditions, including short circuits, overloads, undervoltages, and overvoltages. These faults may lead to equipment damage or even cause dangerous events such as fires and explosions, posing a serious threat to people's lives and property. Therefore, it is very important to find and deal with these faults in time.

In this paper, electrical faults are detected using SVM, Decision Tree, KNN and Random Forest models based on the electrical fault dataset of power system. The confusion matrix results show that most of the predictions from the four models are correct, and only a few instances that should not be predicted as electrical faults are predicted as electrical faults.

From the prediction results of the four machine learning models, it can be seen that all four models have achieved more than 99% accuracy in detecting electrical faults in power systems. Among them, the SVM model has the highest prediction accuracy of 99.6%, followed by the KNN model prediction accuracy of 99.5%.

In summary, in this paper we have used a variety of machine learning algorithms to detect electrical faults and concluded that all of these algorithms can detect electrical faults very accurately. These results are of great significance for the stable operation and safety of the power system, and also provide a reference for future research.

References

[1] Shenquan W ,Yunfei J ,Pu X , et al.Fault detection using generalized autoencoder with neighborhood restriction for electrical drive systems of high-speed trains[J].Control Engineering Practice,2024,143105804-.
[2] Jing C ,Huiqin L ,Na X , et al.Fault diagnosis of electrical equipment based on virtual simulation technology[J].Nonlinear Engineering,2023,12(1):
[3] Moysidis A D ,Karatzinis D G ,Boutalis S Y , et al.A Study of Noise Effect in Electrical Machines Bearing Fault Detection and Diagnosis Considering Different Representative Feature Models[J].Machines,2023,11(11):
[4] You Z ,Wu Q,Zhou H, et al.Fault Detection of Electrical Automation Remote Equipment Based on Data Network[J].ELECTRICA,2023,23(3):458-465.
[5] C. E P ,Gary H ,Raymond H B , et al.Electrical substation grid testbed for DLT applications of electrical fault detection, power quality monitoring, DERs use cases and cyber-events[J].Energy Reports,2023,101099-1115.
[6] Mahmoud E ,Abdelkader I ,Wael S S .Principal component analysis approach for detecting faults in rotary machines based on vibrational and electrical fused data[J].Mechanical Systems and Signal Processing,2023,200
[7] Kong S ,Oh J ,Yoon D , et al.Integrating Deep Learning and Deterministic Inversion for Enhancing Fault Detection in Electrical Resistivity Surveys[J].Applied Sciences,2023,13(10):
[8] Yakun F .Electrical Line Fault Detection and Line Cut-Off Equipment and Control[J].Journal of Control Science and Engineering,2022,2022
[9] D. A B ,A. A E ,M. G A , et al.Internal electrical fault detection techniques in DFIG-based wind turbines: a review[J].Protection and Control of Modern Power Systems,2022,7(1):
[10] Cheng Y .Thermal Fault Detection and Severity Analysis of Mechanical and Electrical Automation Equipment[J].International Journal of Heat and Technology,2022,40(2):