International Journal of Thermofluids 24 (2024) 100814

Contents lists available at ScienceDirect

International Journal of Thermofluids

journal homepage: www.sciencedirect.com/journal/international-journal-of-thermofluids

Design of intelligent detection method for electricity transmission line equipment defect based on data mining algorithm

Xiang Yin *, Ya Zuo, Gaoshan Fu

State Grid Xinjiang Electric Power Company, Urumqi 830063, Xinjiang, China

ARTICLE INFO

Keywords:
Electricity transmission line equipment
Data mining algorithm
Intelligent detection method
Cluster analysis

ABSTRACT

Electricity transmission line is the most significant way of power transmission. Regular detection of it can find and eliminate its defects and hidden dangers in time and prevent major accidents, which is of great significance to the power system. In order to find the problems in the electricity transmission line in time, this paper applied the data mining algorithm to the intelligent detection method of electricity transmission line equipment defects. An electricity transmission line equipment defect intelligent detection and monitoring system was constructed, and the differences between clustering analysis image recognition technology in data mining algorithms and the XGBoost algorithm were analyzed. The results showed that compared with using XGBoost algorithms, the highest accuracy rate of the intelligent detection method of electricity transmission line equipment defects using data mining algorithm in the detection results was 98 %, which was generally higher than that of XGBoost algorithms, and could reduce the consumption of time. From the perspective of replication rate, the overall average value of XGBoost algorithms was 52.38 % and the overall average value of data mining algorithms was 7.63 %. The replica rate of data mining algorithm was much lower than that of XGBoost algorithm and the performance of fault signal detection was better. Therefore, the application of data mining algorithm to the intelligent detection method of electricity transmission line equipment defects can be more suitable, thus significantly improving the efficiency of all aspects. At the same time, the method has the advantages of simple operation, fast, reliable and not affected by region.

1. Introduction

In the 21st century, along with the progress of science and technology, the new generation of information communication technology, such as artificial intelligence (AI), the Internet of Things (IoT), has been rapidly developed. With the development of AI technology, various new technologies and products emerge one after another. AI terminal devices have a variety of functions, can meet people's various needs in entertainment, life, learning and work and also change people's way of life from another point of view. Recently, as the information and communication technology develops rapidly, the power grid has been following the pace of the new generation of AI. It is expected to provide new opportunities for the daily maintenance and transformation of the smart grid. The main task of electricity transmission line detection is to check the quality of electricity transmission line equipment to ensure the safe and stable operation of power network. The current existing transmission line detection methods have limited speed and efficiency, especially in long-distance transmission lines, with limited coverage and detection frequency. Besides, defects can only be discovered after they occur, lacking real-time monitoring and early warning capabilities, making it difficult to take timely measures to prevent the expansion of faults. The data mining algorithm can extract important feature information and the correlation between devices by analyzing and processing a large amount of monitoring data of transmission line equipment, which can provide important basis for equipment defect detection. It can identify different patterns of transmission line equipment in normal and abnormal operating states. By establishing models to identify these patterns, it can help distinguish between the normal operation of devices and potential defect issues, achieving intelligent detection.

Intelligent equipment defect detection in the power system is currently a problem that has to be solved in transmission lines in order to effectively improve the operation and maintenance quality of transmission lines. Among them, Tao X [1] explored the automatic detection of insulator defects using aerial images and accurately located the defects in the input images collected from the actual detection environment. Accurate locating and identifying of insulator faults was achieved

* Corresponding author.
E-mail addresses: yinxiang524@sina.com (X. Yin), wzlzjut@126.com (Y. Zuo), Gridwang1218@126.com (G. Fu).

https://doi.org/10.1016/j.ijft.2024.100814

Available online 22 August 2024

by using a convolution neural network (CNN) cascade model based on in-depth learning. The experimental results showed that the method could detect the defect of the insulator effectively under different working conditions and it could detect the defect accurately and robustly. On the basis of previous work, using CEEMD (Complementary Ensemble Empirical Mode Decomposition) sensitivity coefficient and other methods, Zhong Y [2] analyzed the time-frequency signals of five types of faults and extracted their features. Experiments showed that the system had a high sensitivity to mechanical vibration signals. Zhao Z [3] regarded bolts as the fastener with the largest number and the most easily lost in electricity transmission lines. To solve this problem, an unsupervised clustering algorithm based on the visual morphology of bolts was proposed, and it was used to build a defect recognition model that can perceive the difference of visual morphology of bolts. The result of defect recognition was obtained by regression operation and regional feature classification. Experiments showed that this method could obtain satisfactory detection results. Luo Y [4] explored how to reduce the dependence on human experience in electricity transmission line detection, put forward intelligent electricity transmission line detection based on unmanned aerial vehicle (UAV) and proposed three core technologies, namely, UAV path planning, track tracking, fault detection and diagnosis. When line detection was intelligent, it could greatly enhance the flexibility and reliability of the system. Song C [5] presented an intelligent target detection method based on "Cloud Edge Synergy" and applied it to the fault detection of insulator strings in the power IoT. The results showed that the proposed algorithm had a high detection accuracy, reduced the computational load and improved the detection efficiency under the maximum cosine. Liu J [6] proposed an improved RetinaNet-based object detection algorithm and established a DenseNet-based feature pyramid network as the backbone network. The improved model was trained and tested by using a transmission line defect dataset for validation. The results indicated that this method had advantages and effectiveness in defect detection of transmission lines, and met the requirements of intelligent detection in terms of accuracy. Wang B [7] developed an edge intelligence based ice monitoring system that utilized a multi-scale object detection network to extract high-dimensional feature information for ice monitoring. Compared with other CNN-based object detection methods and traditional ice monitoring methods, the proposed method had a recognition accuracy of 74.5 % for ice images collected under extreme weather conditions. Tan H [8] proposed a new method for longitudinal protection of ultra-high voltage direct current transmission lines based on U-I characteristics. By establishing a U-I plane with DC voltage and current as variables, the real-time captured dynamic trajectory of the U-I plane was analyzed. Finally, the correctness of the results was verified through theoretical analysis and simulation. Therefore, for different types of faults, different intelligent detection methods can be used to achieve the maximum effect.

In order to find an algorithm suitable for intelligent detection of defects in electricity transmission line equipment, many researchers have used data mining algorithms to carry out research to see if the detection method can bring the maximum effect. Among them, Kumar T S [9] used the decision support system of data mining in financial and marketing applications to analyze the actual situation and see if it can improve the performance of the organization. The results showed that the marketing decision support system used effective data mining methods to reduce the burden of organizational analysis and strategic planning and improved organizational performance. El Mohadab M [10] discussed the automation of resumes for researchers in different research laboratories and disciplines. Using natural language processing and data exploration to make decision trees in classifiers was to predict each researcher's work area. It has been found that natural language processing could combine string-word vector filters with decision tree classifiers in automating the scientific researcher's resume. Nurhachita N [11] compared K-Means (k-means clustering algorithm) and Naive Bayes'impact on new students' enrollment. Data mining algorithm was used to study national examination scores, school origin and learning items in the database. The results showed that using K-Means could produce three clusters and the accuracy of Naive Bayes results was 9.08 %. Peng Y [12] took the browsing path of an e-commerce website as an example, studied the data processing in data mining and used information entropy and ant colony clustering algorithm to complete the process. He used the original data to infer and verify and compared the effect before and after the improvement. This algorithm could obtain more accurate and effective data information. Wang C [13] analyzed the working principle of the traditional Apriori algorithm and some problems in the algorithm. After analyzing the mining method and process of time series association rules, the improved time series Apriori algorithm of regular item set was applied to the mining of association rules based on time constraints. Experiments showed that this method had a larger memory capacity than conventional methods. From the above points, it can be seen that various technologies contained in data mining algorithms can be applied in all fields, which promotes the intelligent development of defect detection methods for electricity transmission line equipment.

This article aims to improve the recognition ability of transmission line equipment defects and reduce missed and false detections through advanced data mining algorithms. By building a system that can monitor the status of transmission line equipment in real-time, potential defects can be detected and warned in a timely manner, improving the safety and reliability of the power system.

The research contribution of this article is:

(1) Innovative data mining applications: By processing large-scale and complex transmission line equipment data through data mining algorithms, it is possible to more accurately identify defects in equipment, reduce missed and false detection rates, and improve detection accuracy.
(2) Efficient defect identification model: Based on data mining algorithms, detection is implemented to monitor and warn the real-time status of transmission line equipment, helping operation and maintenance personnel to make timely responses and prevent potential fault risks.
(3) Real-time monitoring and early warning system: An intelligent detection system has been developed, which can quickly respond to changes in the status of transmission line equipment, provide timely warning information, and enhance the safety of the power system. This promotes the development of intelligent and automated management of transmission line equipment, improves management efficiency and level, and provides technical support for the intelligent transformation of the transmission industry.

2. Intelligent detection method for electricity transmission line equipment defect based on data mining algorithm

2.1. Basic concepts of data mining and detection of electricity transmission line equipment defect

2.1.1. Definition of data mining

The volume of data contained in databases has grown quickly as a result of ongoing development of database technology [14] and the extensive adoption of database management systems. Behind these huge data, there is a large amount of valuable information hidden. If this information can be extracted from the database, enormous benefits can be brought. This technology that extracts information from large-scale databases is called data mining. Data mining, in its broadest sense, is the process of obtaining hidden and potentially valuable knowledge and information from vast, imperfect, noisy, fuzzily distributed, and random data. The definition contains the following four meanings: (1) the data source must be real, massive and noisy. (2) What is found is meaningful knowledge for the user. (3) The knowledge obtained must be acceptable, understandable, and applicable, and the results should preferably be expressed in a natural language. (4) This does not mean to find a general knowledge or a new, pure and scientific theory, but only a relatively necessary condition and limitation to apply to certain fields of theoretical knowledge.

From a business perspective, data mining can be defined as an advanced and effective way to discover and model hidden, unknown or known laws by exploring and analyzing a large amount of enterprise data based on business goals set by an enterprise.

2.1.2. Definition and characteristic classification of defect detection for transmission line equipment

Defects of electricity transmission lines [15] are caused by long-term operation, external damage, natural disasters, etc. In addition, due to changes in the environment around the electricity transmission line, such as increased pollution sources, new trees, new houses, etc., the line cannot work properly, which affects the performance of the equipment. Anomalies that threaten the safe operation of lines and equipment or increase the extent of line damage are considered as equipment defects of electricity transmission lines.

Electricity transmission lines have two main features: (1) higher voltage, thicker conductor and fewer branches; (2) a long transmission distance and a large transmission capacity. In addition, in general, for electricity transmission lines, the higher the voltage level, the larger the corresponding transmission capacity and the longer the transmission distance. If the transmission capacity is small at the same voltage, the transmission distance may be longer. If the transmission capacity is large, the transmission distance may be shorter.

According to different classification methods, the defects of electricity transmission line equipment are divided into three categories. (1) Defect of equipment body: It refers to the defects existing in equipment lines, poles and towers themselves and their accessories, mainly including pole and tower body, pole and tower base, ground flat iron, wires, overhead ground wires, insulators, various types of gold, etc. (2) Defect of ancillary facilities: It refers to the defects of electric towers, wires and ground wires installed outside the ancillary facilities, which mainly include signs, warning signs, lightning protection, vibration protection, bird protection and some monitoring devices. (3) External hidden danger defect: It refers to a hidden safety defect suffered by the main body of equipment due to the change of external environmental factors, mainly including the impact of construction, earth moving, mechanical construction, greening, trees, kites, etc., on electricity transmission lines in the protection area, resulting in hidden danger defect.

2.2. Factors influencing defects of electricity transmission line equipment

The influence on the equipment defect in electricity transmission lines is divided according to the cause, which mainly includes two aspects, one is natural factor and the other is artificial factor.

(1) Influence of natural factors

In China, electricity transmission lines are mostly erected outdoors. They are exposed to the outdoors for a long time and are very vulnerable to severe weather. Not only can they be unpredictable, but they also cause serious damage to electricity transmission lines.

Lightning strikes: According to statistics of power line accidents in China, lightning strikes have caused over 50 % of high voltage line trips in recent years, so lightning strikes have caused enormous damage to power lines and electrical equipment.

A fierce wind: With the expansion of power system scale and frequent occurrence of extreme weather events, the power line tripping events caused by strong wind are also increasing, which has a certain impact on the construction of power lines. In windy areas, the influence of wind on electricity transmission lines is more obvious, for example, the wind direction offset caused by squall wind can cause damage to electricity transmission lines.

Ice cover: For electricity transmission lines, the impact of ice cover on electricity transmission lines is mainly concentrated in North China. The ice thickness on the electricity transmission line is too thick, which may cause overload. In the case of uneven ice, it is very easy to cause accidents such as tower collapse and line interruption. This may cause great losses to people.

Bird damage: Birds' flight behavior does not have much effect on electricity transmission lines, but they often carry debris during flight. When they fly over electricity transmission lines, the scattering or attachment of debris may cause the electricity transmission lines to fail.

(2) The influence of human factors

In recent years, transmission and substation line tripping accidents caused by human factors are increasing. Within the protection scope of electricity transmission lines, illegal behaviors such as building houses privately, building roads wantonly, opening hills and firing guns, cofferdam digging ponds, setting fire and burning occur from time to time, which brings great harm to the safety and stable operation of electricity transmission lines and power networks. Therefore, the damage of electricity transmission lines caused by human factors cannot be ignored. The occurrence of these faults cannot be effectively predicted, which causes great harm and poses a direct threat to the safety of electricity transmission lines and people's lives and property.

2.3. Cluster analysis in data mining algorithms

2.3.1. Definition of cluster analysis

Cluster analysis [16] is a start-up technology for data mining application and a key technology for starting data mining. Cluster analysis is a very simple concept, which refers to a method of grouping samples based on data similarity without a given classification category. In the cluster analysis procedure, a collection of tangible or intangible things is divided into a number of categories made up of related objects for study. By classifying a group of objects, the same group of objects is more similar to other objects to some extent. The data also can be classified according to its inherent attributes, and elements in each cluster may have the same or similar characteristics under certain circumstances. Clustering models, an unsupervised learning approach, can be built on data without class labels, in contrast to classification models, which need training data made up of samples with class labels.

2.3.2. Importance of cluster analysis

Cluster analysis is an important aspect of exploring data mining. It is also a common technology in statistical analysis. It is widely used in many fields such as machine learning, pattern recognition, image analysis, information retrieval, bioinformatics, data compression, computer graphics and so on. Clustering analysis of large data is indispensable for any industry, both macro and micro. Therefore, when analyzing and mining data, the first problem to solve is to classify the data. Data clustering is a fundamental and common problem in data-intensive areas. Clustering can be considered the first step in the development of data resources.

In cluster analysis, the interval scale variable between similarity measures is a rough, linear, continuous, and uncertain continuous variable. Typical examples are weight, height, longitude, latitude, temperature, and so on. The selection of measurement units has a direct impact on the clustering results. Typically, choosing a smaller unit can result in a larger range of possible values for a variable, thereby better affecting the clustering results. Therefore, to avoid relying too much on cell selection for clustering results, data must be normalized. After normalization, the difference between the targets is calculated based on the distance. The most common distance measure [17] is Euclidean distance, which is defined as formula 1:

$$X(i,t) = \sqrt{\left(|a_{i1} - a_{t1}|^2 + |a_{i2} - a_{t2}|^2 + \ldots\ldots + |a_{in} - a_{tn}|^2\right)} \quad (1)$$

$i = (a_{i1}, a_{i2},\ldots, a_{in})$ and $t = (a_{t1}, a_{t2},\ldots, a_{tn})$ are two-dimensional data objects. In the application of Euclidean distance, particular attention should be paid to the selection of various measures in the sample, which should be able to effectively reflect the characteristics of class attributes. Two other well-known measures are the Manhattan Distance and Minkowski Distance. The Manhattan distance is shown in formula 2:

$$X(i,t) = |a_{i1} - a_{t1}| + |a_{i2} - a_{t2}| + \ldots\ldots |a_{in} - a_{tn}| \quad (2)$$

Minkowski Distance is shown in formula 3:

$$X(i,t) = \sqrt[q]{(|a_{i1} - a_{t1}|^q) + (|a_{i2} - a_{t2}|^q) + \ldots\ldots + (|a_{in} - a_{tn}|^q)} \quad (3)$$

It can be seen from Minkowski Distance that when $q = 1$, Manhattan distance is represented. When $q = 2$, Euclidean distance is represented. There are two states in the variable: 0 or 1. 0 means the variable is empty and 1 means the variable exists. If a variable has the same weight and has the same value in both states, the variable is symmetric. The most famous simple match factor is used to evaluate the difference between two objects, which is defined as formula 4:

$$X(i,t) = (r+s)/(q+r+s+t) \quad (4)$$

In a variable, if the output of two states is not equal, it is called variable asymmetry. For example, in one study, the most important result which is the lower probability of occurrence is coded as 1, and the rest is coded as 0. If two asymmetric variables are taken as 1 and it makes more sense than 0 at the same time, the t-value can be ignored. In this case, the dissimilarity is calculated using the evaluation factor Jaccard, which is defined as formula 5:

$$X(i,t) = (r+s)/(q+r+s) \quad (5)$$

The definitions of variables that appear in formulas 1 to 5 are shown in Table 1:

Therefore, to solve the problem of inefficiency in the process of defect collection due to artificial uncertainty in the existing technology, the algorithm of cluster analysis image recognition technology is applied to the detection and monitoring system, aiming to achieve the method of detecting the defects of electricity transmission lines, the defect detection equipment and system and to automatically identify the defects of electricity transmission lines.

3. Design of intelligent detection and monitoring system for defects of electricity transmission line equipment

3.1. Overall needs

Power electricity transmission lines occupy a certain proportion in the national power grid construction, are an indispensable part of economic development and residents' lives and are the conveyor belts connecting power stations, substations and users. Their normal work and operation are an important condition for circuit stability. The electricity transmission lines are all over the country. Most of the electricity transmission lines are exposed to the natural environment in the field. The environment in the field is very bad and complex. Under the long-term influence of natural environment such as hail, wind, sun and rainstorm, the lines themselves also aging with time. This easily lead to issues such as stock splitting, stock breaking and self-explosion of insulators. Therefore, it is necessary to continuously monitor and maintain them, aiming to ensure its normal operation.

Table 1
Definition of formula variables.

Sequence	Variable	Define
1	(i, t)	N-dimensional vector
2	q	A parameter that determines the measurement method of distance
3	r	The number of features with the same value of 1 in two vectors
4	s	The number of features with vector i being 1 and vector t being 0

The traditional method of power electricity transmission line detection mainly relies on the detection personnel to walk on the line regularly, observe their running condition and then climb to the height of the iron tower to inspect each component. Such detection method cannot overcome the shortcomings of high labor intensity, long detection time and high danger of detection personnel, especially in remote forests and mountains where roads and communications are poor. In case of severe weather, the inspectors cannot inspect and remove the potential safety hazards on the lines in time, which may result in serious economic and other losses. In addition, there are many kinds of contents to be detected in line detection. When checking line equipment with human eyes, some obvious and intuitive defects can be easily found and some minor defects in corners are likely to be ignored, so that small defects may gradually develop into large problems. Because inspectors observe a large number of equipment visually, and the longer the working time, the stronger the fatigue of the inspector, the more likely the defect identification is biased. Therefore, the image defect recognition algorithm [18] in cluster analysis technology can be used to analyze and process the line image data obtained by unmanned aerial vehicle detection, which intelligently identifies the defects of electricity transmission lines, effectively avoids the drawbacks of relying on human eye experience to judge the line defects and can find the problems in time and in an all-round way.

3.2. Overall framework of system

As the main carrier of power transmission, regular detection of electricity transmission lines is the basis to ensure the safe and reliable operation of power grid equipment. With the rapid development of economy, there are more and more electricity transmission lines in the power system, from economically developed urban areas to remote mountainous areas. Traditional manual detection inevitably have problems such as inadequate detection and hidden dangers cannot be eliminated in time. However, in recent years, the gradually rising unmanned detection has greatly made up for the shortcomings of manual detection and improved the accuracy and efficiency of detection. Unmanned aerial vehicle detection is a new detection mode which combines human and intelligence. Operators can detect electricity transmission lines by remotely controlling unmanned aerial vehicles. After the unmanned man reaches the specified hovering point at the specified tower, the electricity transmission lines that need to be detected are taken and captured, and then autonomously fly to another hovering point in turn until all the hovering points set are traversed once. UAV based on image ranging of markers can locate markers, monitor the distance between markers and surrounding markers in real-time and adjust its flight attitude in real-time to achieve the purpose of hovering in preset positions. Image processing technology is a technology that can extract, analyze and process static or dynamic targets from images. On this basis, the collected image information is transferred to the ground station using high-definition image transmission. The defect analysis algorithm based on image recognition technology based on cluster analysis is used to discover the defect from a large amount of image information and automatically warn and statistics the defect, aiming to further improve the information processing and statistical ability of detection. The overall structure of the system is shown in Fig. 1.

The whole system described in Fig. 1 measures the distance between the target and the UAV through two subsystems according to the algorithm of image ranging of the identifier, aiming to adjust the flight status of the UAV in time, so it achieves the autonomous obstacle avoidance of

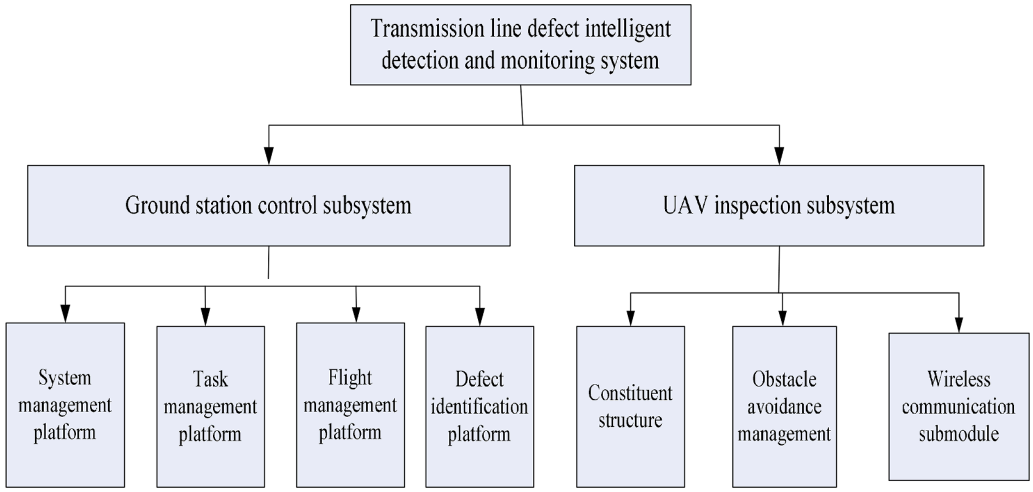

Fig. 1. Overall system structure.

the UAV. It can reach specific locations, take pictures of electricity transmission lines and complete the detection of the entire electricity transmission line. Then, using each platform is to set the login rights of the system, formulate, issue and manage detection tasks, acting as a channel for data transmission, storage and processing between systems. Real-time monitoring of flight posture, flight data and equipment operation during unmanned aerial vehicle detection is carried out to identify the defects of electricity transmission lines.

3.3. Implementation of intelligent detection and monitoring system for defects of electricity transmission line equipment

When image recognition is defective, changes in the light intensity and angle of the external environment may affect the brightness and

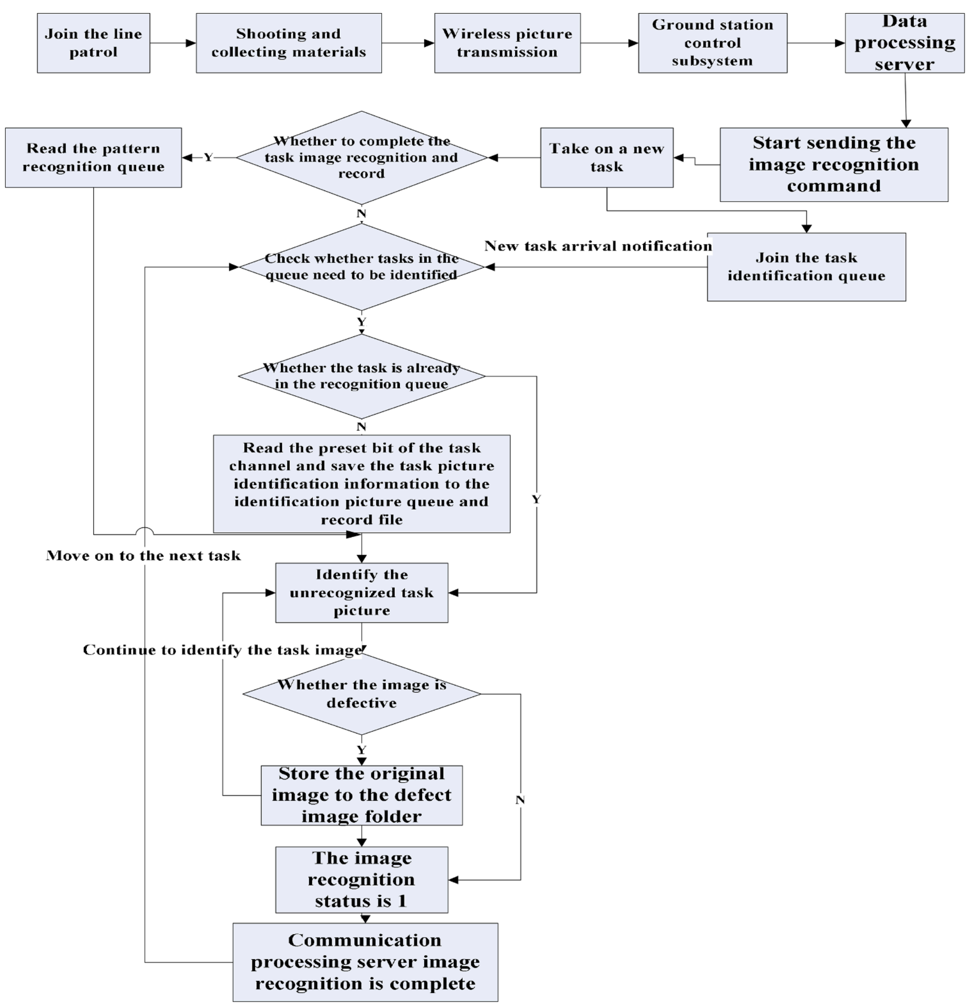

Fig. 2. Image defect recognition flow chart.

resolution of the picture. At the same time, noise and natural motion of surrounding objects may also reduce the quality of the picture, thereby affecting the effect of the picture. The same object can take on different forms at different angles and distances. Therefore, in image processing, especially in the extraction and recognition of target image, the intelligent detection and monitoring system for electricity transmission line equipment defects takes the clustering analysis image recognition algorithm as the basis. To improve detection accuracy and effectively preprocess the data, the mean interpolation method is first used to fill in missing values; then, the outlier identification is performed to directly remove the identified outliers; the Z-score standardization is used to standardize data, and the data of different scales is scaled to the same scale; finally, the data is integrated into a unified data storage to ensure the stability and effectiveness of model training. The collected data is sent to the ground monitoring station in the form of wireless transmission, and a continuous learning mechanism is established. The model is regularly updated, and the latest data and knowledge are integrated to adapt to equipment operation status and environmental changes, so as to improve the real-time and accuracy of the detection model. Ground observation systems receive, store, analyze and process real-time image data from unmanned reconnaissance aircraft detection. After the line defect is detected, the result of defect identification is displayed and the alarm is timely. Because the parameters required are different in different environments, they need to be reset each time they are executed. The process of defect image recognition for electricity transmission line equipment is given in Fig. 2.

After introducing several subsystems, there is a certain understanding of the intelligent detection and monitoring system for electricity transmission line equipment defects. The most critical step of image recognition process has been completed. Therefore, the entire intelligent detection system has been implemented, and the detection methods need to be experimented to obtain specific results. The interface diagram of the intelligent detection system for electricity transmission line equipment defect is shown in Fig. 3.

With the system described in Fig. 3, it is possible to discover the problems of electricity transmission lines in time and ensure the safe and stable operation of power networks and power supply systems with the features of long line length, complex geography, diverse environment and natural exposure.

4. Intelligent detection experiment of electricity transmission line equipment defect based on data mining algorithm

To validate the effectiveness of data mining algorithms, this paper conducted experimental analysis on intelligent defect detection of transmission line equipment. This paper focused on power equipment manufacturing companies, and the experimental dataset consisted of their daily monitoring data and data from regular inspection reports, mainly including voltage, current, temperature, vibration, and other data. To reduce the uncertainty of the model in this paper, the data in the dataset was preprocessed and used as input. In data processing, interpolation was used to process missing data. For time series data, the linear interpolation was used to fill in missing values and remove random noise from the data through smooth filtering. The data was scaled to between 0 and 1 to eliminate the influence of different feature dimensions. Finally, samples from existing data are randomly selected to generate new test data.

Before analyzing the experimental results, the sensitivity of the model was analyzed to parameter selection, as shown in Table 2:

From Table 2, it can be seen that the model had a certain sensitivity to changes in learning_rate, max_depth, and regulation_param.

4.1. Accuracy

Accuracy refers to the percentage of the total sample in which the correct predictions are made. Although accuracy can be used to judge the overall accuracy, it does not provide a good indicator of the outcome in case of sample imbalance. Therefore, this paper found the accuracy of the data mining algorithm used in the intelligent detection and monitoring system of electricity transmission line equipment defects compared with the XGBoost algorithm to measure the results. Fig. 4 shows the results of data mining algorithms compared with XGBoost algorithms on accuracy.

From Fig. 4, it can be seen that compared to XGBoost algorithms, the defect detection of transmission line equipment under data mining in this paper had more ideal accuracy results. From the specific comparison results, the average accuracy of the algorithm in this paper reached 91.02 % in the detection tasks from January to December; the average accuracy result of XGBoost algorithms was 72.23 %. Compared to XGBoost algorithms, the average accuracy under data mining in this

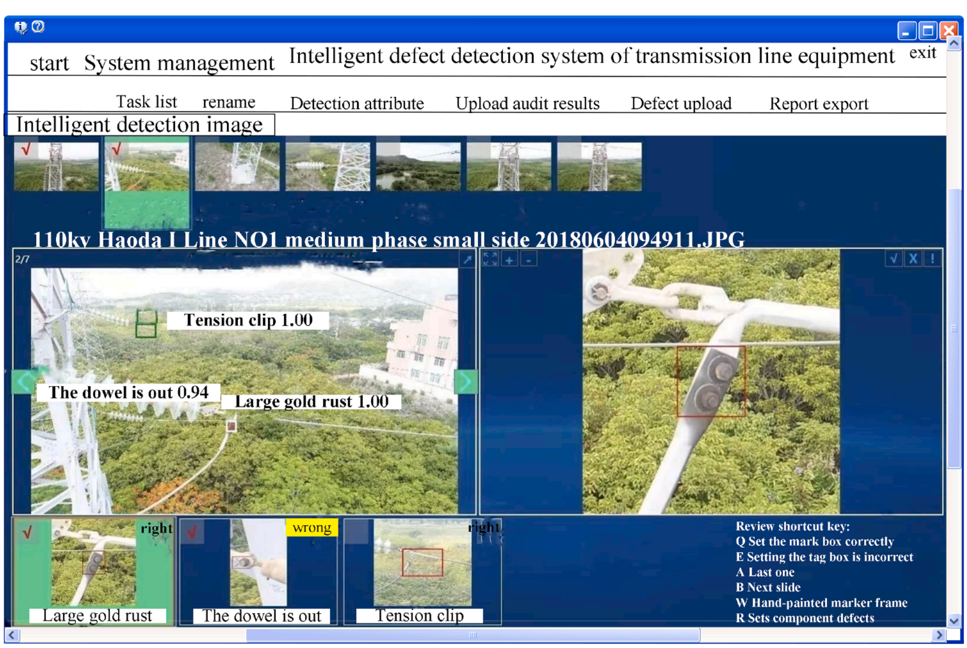

Fig. 3. Interface diagram of electricity transmission line equipment defect intelligent detection system.

Table 2
Sensitivity analysis.

Parameter	Value	Accuracy	Parameter changes	Change in accuracy	Sensitivity coefficient
learning_rate	0.001	0.84			
	0.005	0.86	0.004	0.02	5.0
	0.01	0.87	0.005	0.01	2.0
max_depth	10	0.91			
	20	0.95	10	0.03	0.003
	30	0.92	10	0.01	0.001
regularization_param	0.01	0.82			
	0.1	0.88	0.09	0.02	0.22
	1.0	0.93	0.9	0.01	0.011

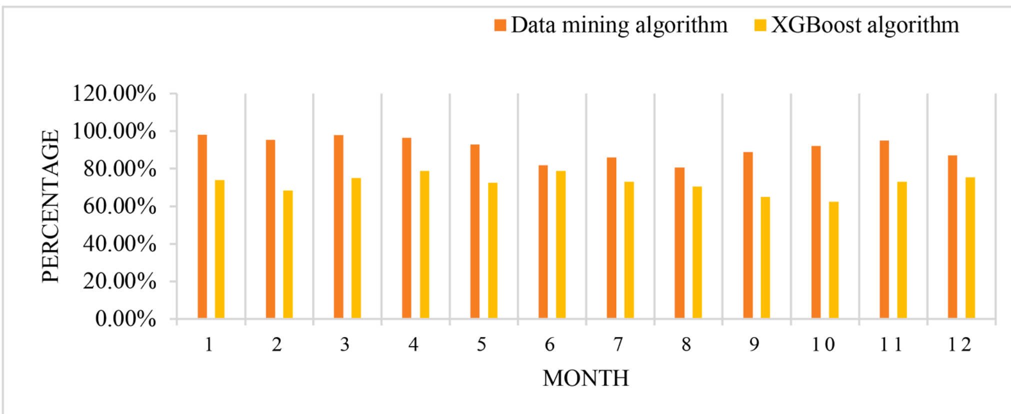

Fig. 4. Comparison results of accuracy of two algorithms.

paper was 18.79 % higher. By comparing the two algorithms, it can be seen that in the application of intelligent defect detection for electricity transmission line equipment, the data mining algorithm had a higher accuracy and stability than the XGBoost algorithm, which provided a better reference value for intelligent defect detection. Data mining algorithms can process large amounts of historical and real-time data, capture small anomalies and trend changes during device operation through analysis of massive data, and improve the accuracy of defect recognition. In equipment defect detection, complex features are automatically extracted from data to provide more comprehensive information and help identify equipment defects. Therefore, combining the data mining algorithm with the intelligent detection method of electricity transmission line equipment defects could better improve the accuracy of electricity transmission line equipment defects diagnosis results.

To objectively analyze the accurate effect of transmission line equipment defects, the detection error percentage ε of two types of algorithms is calculated, as shown in formula 6:

$$\varepsilon = \frac{V_p - V_t}{V_t} \times 100\% \tag{6}$$

In formula 6, V_p is the predicted value; V_t is the true value. According to formula 6, the detection error percentages of different algorithms are compared. The results are shown in Table 3:

From Table 3, it can be seen that in the error comparison, the algorithm based on data mining in this paper generally had lower error results, with an average error percentage of 4.41 % in the intelligent detection of transmission line equipment defects from January to December; the average error percentage of XGBoost algorithms in intelligent defect detection of transmission line equipment from January to December was 7.78 %. This result indicates that the intelligent defect detection under the algorithm proposed in this paper is more accurate and objective.

Table 3
Comparison of error percentages.

Month	Data mining algorithm (%)	XGBoost algorithm (%)
1	3.31	7.56
2	5.32	7.50
3	5.71	8.50
4	3.26	5.40
5	4.06	7.95
6	5.45	7.67
7	3.89	8.78
8	3.18	8.03
9	4.60	8.72
10	5.32	7.66
11	3.31	7.56
12	5.32	7.50

4.2. *Time*

In the experimental process, different samples, different methods and different locations result in different results in time. In order to solve this problem, this paper used data mining algorithm and XGBoost algorithm to expand the detection and comparison analysis to find out the difference between the two in the same sample, method, location, and so on. The comparison results are shown in Fig. 5.

In Fig. 5, the algorithm proposed in this paper spent less time on intelligent defect detection of transmission line equipment. In terms of time comparison, the average time spent by the algorithm in this paper on defect detection tasks from January to December was 12.45, while the average time spent by XGBoost algorithms on defect detection tasks from January to December was 32.50. Compared with XGBoost algorithms, with the support of data mining, the detection time of this model has been reduced by 61.69 %. Data mining algorithms can automatically process and analyze large amounts of data, and extract and select key features related to defect detection, so as to reduce interference from other factors, improve model processing speed, and identify and process

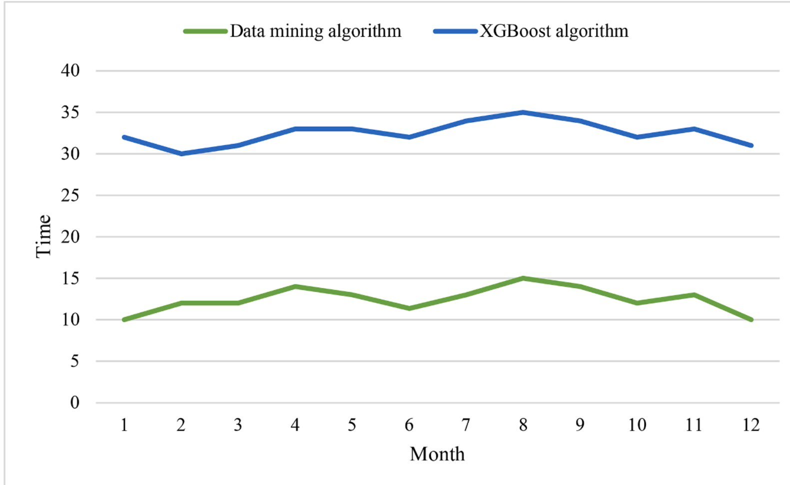

Fig. 5. Comparison results of time between the two algorithms.

potential problems faster and more efficiently.

4.3. Repetition rate

The repetition rate refers to the repeatability required to perform multiple tests on the same type of defect when it is intelligently detected. At a high repetition rate, more time is required, which affects the efficiency of detection. This paper compared the data mining algorithm used in the intelligent defect detection method of electricity transmission line equipment with the XGBoost algorithm in the repetition rate. The comparison results are shown in Fig. 6.

The result of the data mining algorithm compared with the XGBoost algorithm on the repetition rate is described in Fig. 6. It can be concluded from the figure that the highest, the lowest and the overall average repetition rate of the XGBoost algorithm were 60 %, 46 % and 52.38 %, respectively. The highest repetition rate was 10 % and the lowest was 5 %, with overall average of 7.63 % under different experimental times. The repetition rate obtained by data mining method was much lower than that by traditional method. It can utilize historical data and experiential knowledge to timely detect and provide feedback on defect information, avoid duplicate detection of identified problems, and optimize and adjust detection models through feedback mechanisms to improve detection efficiency and accuracy. The decrease of the repetition rate means the decrease of the detection speed and the improvement of the detection efficiency. The experimental analysis showed that the data mining algorithm used in the intelligent detection of electricity transmission line equipment defects could reduce the repetition rate in the detection process, and had higher detection efficiency and better application effect.

4.4. Detection performance

In order to further verify the validity of the intelligent detection method for electricity transmission line equipment defect based on data mining algorithm, this paper compared the performance of the data mining algorithm with the XGBoost algorithm for defect signal detection. The more stable the defect signal detected by the method, the larger the detection range and the better the effectiveness of the representation method. The comparison results are shown in Fig. 7.

The result of comparison between data mining algorithm and XGBoost algorithm on detection performance is described in Fig. 7. As shown in the figure, the defect signal detected by the intelligent defect detection method of electricity transmission line equipment using the XGBoost algorithm had a large fluctuation and a small detection range

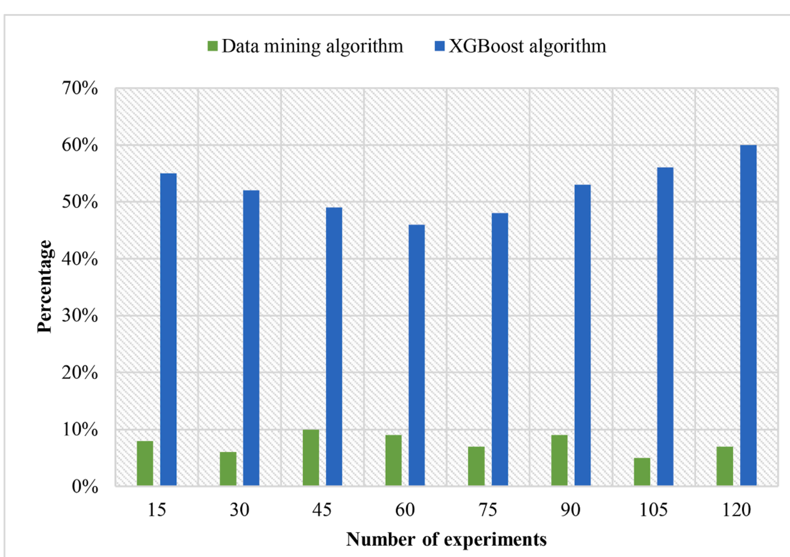

Fig. 6. Comparison results of repetition rates.

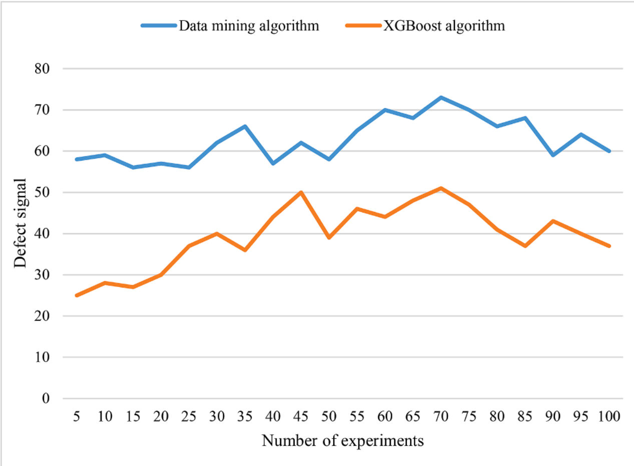

Fig. 7. Comparison results of algorithm detection performance.

with a maximum of 51 and a minimum of 25 under different experimental times. However, the defect signal detected by the intelligent detection method of electricity transmission line equipment defects using data mining algorithm had stronger signal stability than the XGBoost algorithm, and could be inspected in a larger range. The highest defect signal detected was 73, and the lowest was 56. From the perspective of the whole data, the data mining algorithm was more suitable for the intelligent detection method of electricity transmission line equipment defects. It can learn stable patterns from a large amount of data, has strong anti-interference ability, can identify true defect signals, and improve the stability and reliability of detection results. This also proved from the side that the intelligent detection method of electricity transmission line equipment defects based on data mining algorithm may have better performance in the detection of defect signals, and could effectively realize the intelligent detection of electricity transmission line equipment defects.

5. Conclusions

With the continuous deepening of China's reform and opening up and the continuous development of the national economy, the scale, capacity and coverage of modern transmission networks are expanding and the position of electricity transmission lines in the national economy has become more important. On this basis, this paper combined the intelligent detection method of electricity transmission line equipment defects with data mining algorithm and made an experimental analysis on it from four aspects: time, accuracy, repetition rate and detection performance, obtaining the influence of the two methods in detection. In this paper, the image defect recognition technology in cluster analysis was used to diagnose and analyze the defect of electricity transmission line equipment. The results showed that the intelligent detection method of electricity transmission line equipment defect using XGBoost algorithm has some problems, such as low accuracy, long time, high repetition rate, small fluctuation of detection performance range, etc. The data mining algorithm has high accuracy and wide detection performance range in the process of operation. Therefore, it was proved that the data mining algorithm was more suitable for the intelligent detection method of electricity transmission line equipment defects, which significantly improved the effectiveness of all aspects of the role. However, this paper also has certain limitations. The defect data of transmission line equipment is usually unbalanced, that is, there may be significant differences in the proportion of normal samples and defect samples. This paper does not conduct in-depth analysis of sample classification in the training and detection process, and the data of transmission line equipment has complex features and diversity. In feature extraction, this paper does not comprehensively consider its influencing factors. In future research, further improvements in sample classification and feature extraction processes an be considered to enhance model performance and enhance its application in transmission line equipment defects.

Funding

No funding were used to support this study.

Data availability statement

No data were used to support this study.

CRediT authorship contribution statement

Xiang Yin: Writing – original draft. **Ya Zuo:** Writing – original draft. **Gaoshan Fu:** Writing – original draft.

Declaration of competing interest

These are no potential competing interests in our paper. And all authors have seen the manuscript and approved to submit to your journal. We confirm that the content of the manuscript has not been published or submitted for publication elsewhere.

References

[1] X. Tao, D. Zhang, Z. Wang, Detection of power line insulator defects using aerial images analyzed with convolutional neural networks, IEEE Trans. Syst. Man. Cybern. 50 (4) (2018) 1486–1498.
[2] Y. Zhong, J. Hao, R. Liao, Mechanical defect identification for gas-insulated switchgear equipment based on time-frequency vibration signal analysis, High Volt. 6 (3) (2021) 531–542.
[3] Z. Zhao, H. Qi, Y. Qi, Detection method based on automatic visual shape clustering for pin-missing defect in transmission lines, IEEE Trans. Instrum. Meas. 69 (9) (2020) 6080–6091.
[4] Y. Luo, X. Yu, D Yang, A survey of intelligent transmission line inspection based on unmanned aerial vehicle, Artif. Intell. Rev. 56 (1) (2023) 173–201.
[5] C. Song, W. Xu, G. Han, A cloud edge collaborative intelligence method of insulator string defect detection for power IIoT, IEEe Internet. Things. J. 8 (9) (2020) 7510–7520.
[6] J. Liu, R. Jia, W. Li, et al., High precision detection algorithm based on improved RetinaNet for defect recognition of transmission lines, Energy Rep. 6 (2020) 2430–2440, 2020.
[7] B. Wang, F. Ma, L. Ge, et al., Icing-EdgeNet: a pruning lightweight edge intelligent method of discriminative driving channel for ice thickness of transmission lines, IEEE Trans. Instrum. Meas. 70 (2020) 1–12.
[8] H. Tan, Y. Wang, Z. Li, et al., A novel pilot protection method for UHVDC transmission lines based on UI characteristics, Energy Rep. 9 (2023) 1850–1860.
[9] T.S Kumar, Data mining based marketing decision support system using hybrid machine learning algorithm, J. Artif. Intell. 2 (03) (2020) 185–193.
[10] M. El Mohadab, B. Bouikhalene, S. Safi, Automatic CV processing for scientific research using data mining algorithm, J. King Saud Univ. - Comput. Inf. Sci. 32 (5) (2020) 561–567.
[11] N. Nurhachita, E.S Negara, A comparison between naïve bayes and the k-means clustering algorithm for the application of data mining on the admission of new students, Jurnal Intelektualita: Keislaman, Sosial dan Sains, 9 (1) (2020) 51–62.
[12] Y. Peng, X. Yang, W Xu, Optimization research of decision support system based on data mining algorithm, Wirel. Pers. Commun. 102 (4) (2018) 2913–2925.
[13] C. Wang, X. Zheng, Application of improved time series Apriori algorithm by frequent itemsets in association rule data mining based on temporal constraint, Evol. Intell. 13 (1) (2020) 39–49.
[14] C. Guijas, J.R. Montenegro-Burke, X. Domingo-Almenara, METLIN: a technology platform for identifying knowns and unknowns, Anal. Chem. 90 (5) (2018) 3156–3164.
[15] Y. Hadad, J.C. Soric, A.B Khanikaev, Self-induced topological protection in nonlinear circuit arrays, Nat. Electron. 1 (3) (2018) 178–182.
[16] G. Galimberti, A. Manisi, G. Soffritti, Modelling the role of variables in model-based cluster analysis, Stat. Comput. 28 (1) (2018) 145–169.
[17] H.A. Abu Alfeilat, A.B.A. Hassanat, O Lasassmeh, Effects of distance measure choice on k-nearest neighbor classifier performance: a review, Big Data 7 (4) (2019) 221–248.
[18] S. Niu, B. Li, X Wang, Defect image sample generation with GAN for improving defect recognition, IEEE Trans. Autom. Sci. Eng. 17 (3) (2020) 1611–1622.