

POWER GRID FAULT PREDICTION METHOD BASED ON FEATURE SELECTION AND CLASSIFICATION ALGORITHM

XIAOYU YANG

ABSTRACT

In order to accurately predict the risk level of power grid with considering weather factors, a fault level prediction method based on Random Forest and Multi-classification Support Vector Machine is proposed. After data preprocessing and fault classification, the relevant meteorological features of distribution network fault are summarized. The features' weights can be calculated and sorted by using Random Forest algorithm, to obtain the optimal fault feature set, and then the fault level is predicted based on multi-classification SVM. Through the analysis of practical examples, the performance of the proposed method is better than that of SVM alone, and the prediction results can provide an effective basis for risk pre control of power grid, which has practical application value

Index Terms: *Random Forest, Support Vector Machine, Weather factors, Power Grid, Fault prediction.*

Reference to this paper should be made as follows:

XIAOYU YANG, (2021), "POWER GRID FAULT PREDICTION METHOD BASED ON FEATURE SELECTION AND CLASSIFICATION ALGORITHM" *Int. J. of Electronics Engineering and Applications*, Vol. 9, No. 2, pp. 34-44, DOI 10.30696/IJEEA.IX.1.2021.34-44.

Biographical notes:

XIAOYU YANG was born in Hebei, China, in 1996. She received the bachelor's degree in network engineering from Hebei Normal University in 2019. She is currently pursuing the master's degree with the School of Control and Computer Engineering, North China Electric Power University in Beijing, China. Her research interest includes Artificial Intelligence and power grid operation analysis

1. INTRODUCTION

Power grid security refers to the ability of power grid to resist disturbance events such as faults, which directly reflects the reliability of power grid. Distribution network, as an important part of the power system, is connected to the user side and power transmission. It has the characteristics of wide geographical distribution, multiple types of equipment, complex connection and variable operation modes. The safe operation of distribution network is an important part of the safe operation of the entire power network, and is the key link to improve the operation level of power supply system. [1]. Current research [2] shows that meteorological factors have a significant impact on the reliability of distribution network. For distribution networks, it is necessary to detect and isolate the faults as soon as possible when they occur in order to speed up the troubleshooting and recovery and reduce the losses. However, it is difficult to detect and control risks in time before a power outage. The active warning application function [3] provides a scientific basis for active warning, prevention and control of frequent power outage accidents.

Document [4] establishes a regression model for grid cells per square kilometer and possible hurricane-related outages in the region to predict the impact of hurricanes on distribution network outages and formulate emergency response policies. Document [5] discusses the effect of vegetation maintenance on overhead lines of distribution network, puts forward a failure rate prediction model based on the influence of vegetation growth factors, and tests the model using historical vegetation growth parameter data and feeder failure rate. Document [6] based on the knowledge acquisition method of rough set, predicts the characteristic gas state characteristics by gray model, and finds the latent fault types accurately as soon as possible, so that the transformer can be repaired in advance. Literature [7] presents two methods to model the failure rate of overhead distribution lines. The first one is based on Poisson regression model, which captures the counting nature of failure events on overheated distribution lines. The second is the Bayesian network model, which uses conditional failure probabilities under different weather conditions. These two methods are used to predict weather-related failure events on overhead lines each year.

At present, Big Data, Machine Learning and other technologies are hot spots in many fields, and they are also applied to the prediction of distribution network fault level. Document [8] uses the method of correlation analysis to excavate the implicit relationship between the operating trend of equipment indicators and the operating conditions of the equipment, so as to effectively predict the fault or impact and effectively predict the failure of complex power equipment. Document [9] establishes a weather-sensitive neural network combination prediction model, which uses the maximum correlation coefficient to adjust the main meteorological factors causing the change of the number of emergency repairs in real time, accurately predicts the number, type and change trend of the faults, and provides qualitative and quantitative guidance for the early deployment of personnel and materials for emergency repairs.

In this paper, the original data is pre-processed, the fault level is divided, then the random forest (RF) [10] is used as the feature selection method, and the optimal feature subset for distribution network fault prediction is selected based on the variable importance measure of the random forest itself. Then, fault risk level prediction is performed based on Multi-classification Support Vector Machine (MSVM). Finally, a numerical example is used to verify the predictive performance of the proposed method.

2. RELATED WORK

2.1 RANDOM FOREST

Before introducing the specific construction method of random forest, firstly, bagging idea is introduced. Let the set contain n different sample individuals. Now we take a sample from the set each time, and then put it back for n times to form a set. For a sample in the set, the probability that it cannot be extracted each time is:

$$P = 1 - \frac{1}{n}$$

The probability that the sample is not included in set is:

$$p = \left(1 - \frac{1}{n}\right)^n$$

Where, $n \rightarrow \infty$ the following equation holds:

$$\lim_{n \rightarrow \infty} p = \lim_{n \rightarrow \infty} \left(1 - \frac{1}{n}\right)^n = e^{-1} \approx 0.368$$

Therefore, after removing the duplicate samples, it is equivalent to about 63.2% of the original set samples in each new set. Each new set acts as a weak learner, and the sampling is in the form of put back, so there is no connection between the sets.

Bootstrap aggregation (Bagging) is the earliest combinatorial learning algorithm. Bootstrap aggregation (bagging) is the earliest combinatorial learning algorithm. Its specific process is as follows:

- 1) By bootstrap sampling, training sets S_1, S_2, \dots, S_N are randomly generated;
- 2) Each training set generates the corresponding decision tree T_1, T_2, \dots, T_N ;
- 3) Each decision tree is tested with test set samples;
- 4) By voting, the category of the test set data has the largest number of votes.

The Random Forest is composed of a large number of classification trees. Each classification tree acts as a weak learner and combines all the trees to produce voting results. For the classification algorithm, the prediction result is the category with the most votes cast by all weak learners; For the regression algorithm, the output of the model is the regression result obtained by all weak learners and the arithmetic average value.

At present, there are three kinds of decision tree generation algorithms, namely ID3 algorithm [11], C4.5 algorithm [12] and Cart classification algorithm [11]. For the algorithm of constructing classification tree based on random forest, C4.5 algorithm is selected here, because C4.5 algorithm is improved on the basis of ID3 algorithm. Through the calculation of information gain rate, ID3 algorithm can avoid the disadvantage that the result of ID3 algorithm tends to more numerical attributes due to the calculation of information gain. Experiments show that C4.5 algorithm is better than cart classification algorithm. The decision tree can be optimized by post pruning to improve its generalization ability and avoid over fitting.

The construction process of random forest algorithm model is shown in [Figure-1] as follows:

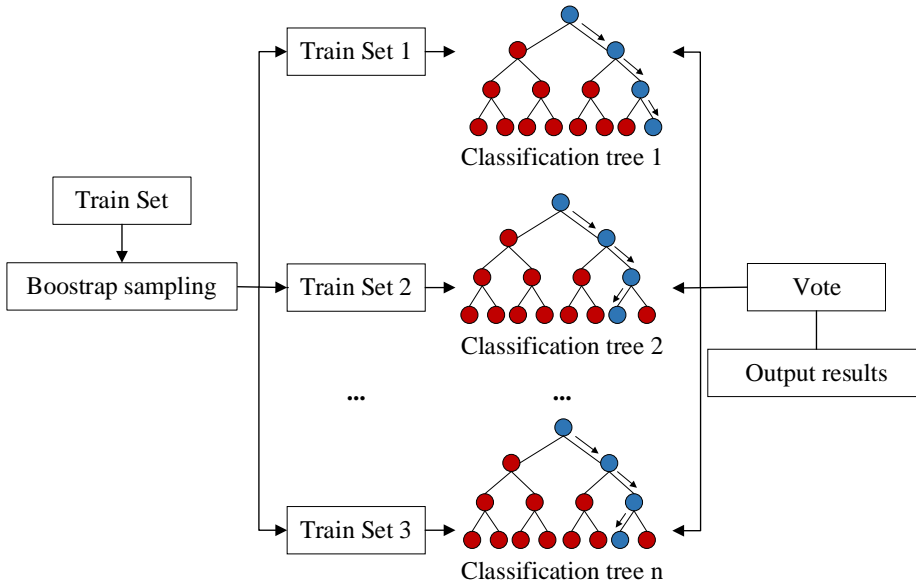


Figure: 1. Random Forest construction process

- Like Bagging algorithm, N training sets S_1, S_2, \dots, S_N are generated randomly by resampling with Bootstrap method;
- Each training set generates the corresponding decision tree T_1, T_2, \dots, T_N by C4.5 algorithm. The difference from bagging algorithm is that random forest algorithm randomly extracts m attributes from M attributes as the current node's split attribute set, and splits the node from top to bottom according to the optimal split mode of the m attributes set;
- After the natural growth of the classification tree is completed, the classification tree is constructed by post pruning;
- Each decision tree is tested by using test set samples. Through voting, the category with the largest number of votes is the category of test set data

2.2 SVM ALGORITHM

Support Vector Machine (SVM) algorithm was first proposed by Vapnik et al. [13], which is used to solve the classification problem of two categories. Now it is widely used in classification, recognition and regression problems.

The basic idea of SVM is to find an optimal hyperplane in high-dimensional space through some kernel function, and separate the two kinds of data. Specifically, suppose that the training set on a feature space is $\{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$, where $x_i \in R^n$, $y_i \in \{+1, -1\}, i = 1, 2, \dots, N$, x_i is the i -th feature vector, y_i is the class label, and when it is equal to $+1$, it is a positive example, and when it is equal to -1 , it is a negative example. Assuming that the training data set is linearly separable, for a given data set and hyperplane $\omega \cdot x + b = 0$, the geometric interval of the hyperplane with respect to the sample point (x_i, y_i) is defined as:

$$\gamma_i = y_i \left(\frac{\omega}{\|\omega\|} \cdot x_i + \frac{b}{\|\omega\|} \right)$$

The minimum geometric interval of the hyperplane with respect to all sample points is:

$$\gamma = \min_{i=1,2,\dots,N} \gamma_i$$

The distance is the distance from the support vector to the hyperplane.

The classification problem with multiple features is usually nonlinear and separable. It can be transformed into a linear classification problem in a certain dimension feature space by nonlinear transformation, and learn linear support vector machine in high dimension feature space. Specifically, the kernel function $K(x, z)$ is used to replace the inner product to construct and solve the convex quadratic programming problem:

$$\begin{aligned} \min_{\alpha} & \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j K(x_i, x_j) - \sum_{i=1}^N \alpha_i \\ \text{s.t.} & \sum_{i=1}^N \alpha_i y_i = 0 \\ & 0 \leq \alpha_i \leq C, i = 1, 2, \dots, N \end{aligned}$$

Where α_i is Lagrange Multiplier and C is Penalty Factor. The classification decision function can be obtained by calculation:

$$f(x) = \text{sign}(\sum_{i=1}^N \alpha_i y_i K(x, x_i) + b)$$

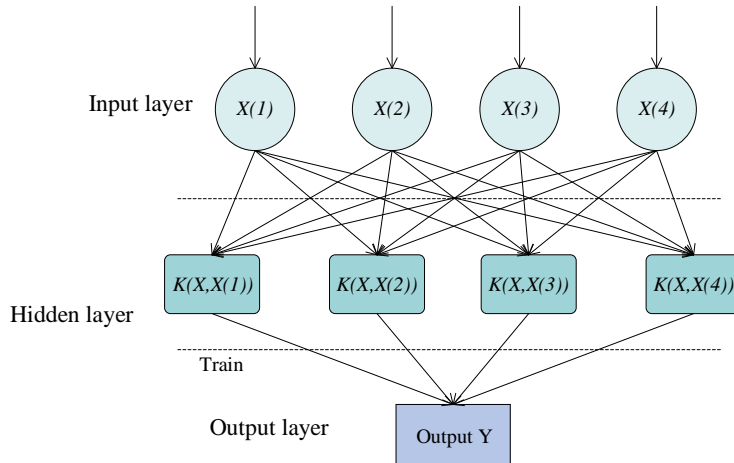


Figure: 2. Random Forest construction process

For different data sets, the classification effect of different kernel functions may be completely different. Common kernel functions include polynomial kernel function, radial basis function and sigmoid kernel function. The structure of SVM model is shown in **[Figure-2]**.

Traditional SVM is used to solve the problem of binary classification. When the multi classification problem is divided into different levels according to the fault degree, multi class support vector machine can be used, which can be realized by "one-to-one" [14], "one to many" and "directed acyclic graph".

In this paper, the "one-to-one" method is used to achieve multi classification, that is, a SVM sub classifier is designed between any two classes of samples, so a total of $\frac{k(k+1)}{2}$ sub classifiers are designed for k classes of samples. When predicting a new sample category, each sub classifier discriminates it and votes for the corresponding category. The category with the most votes is determined as the prediction category.

3 PROPOSED WORK

3.1 DATA ACQUISITION AND PROCESSING

Starting from the predictability of the causes of the distribution network failure and the practical engineering usability of the prediction results, this paper forecasts the feeder failure level of the distribution network involving the local meteorological factors, and collects the basic data related to the distribution network failure. Because the current distribution network information system is still in the immature stage, the data will inevitably appear incomplete, repeated or even abnormal. In order to provide accurate and reliable reference for subsequent fault prediction, we must first preprocess the original data. This process includes the following parts:

- Data cleaning: delete irrelevant data and duplicate data in the original data, smooth noise data, filter out data irrelevant to mining topics, and deal with missing values and outliers;
- Data integration: multiple data sources are merged and stored in a unified database;
- Data transformation: normalize the data according to the following formula to eliminate the adverse effects caused by singular sample data:

$$x^* = \frac{x - \mu}{\sigma}$$

Where μ is the mean value of all sample data and σ is the standard deviation of all sample data.

- Outlier sample data elimination: after preprocessing the original data, due to the data coming from different classes, natural variation, data measurement and collection errors, abnormal data may still appear in the multi-dimensional data samples. This kind of data is called outlier sample data. Outlier sample data may directly affect the fitting accuracy of the model, or even lead to wrong prediction results, Therefore, outlier data must be eliminated. Outlier detection mainly uses the idea of clustering, clustering all objects first, and then evaluating the outlier score of each object.

In this paper, the K-means based clustering algorithm is used as follows:

- The samples are clustered into k clusters by K-means algorithm to determine the centroid of each cluster;
- The distance and relative distance from each object to its nearest centroid are calculated respectively;
- The relative distance is compared with the threshold. If it is greater than the threshold, the object is considered as outlier and removed.

3.2 FAULT LEVEL CLASSIFICATION

The power grid fault is divided into fault grades to facilitate subsequent prediction. The distribution network fault level is divided into three levels: mild risk, moderate risk and severe risk based on weekly fault frequency and outage duration, as shown in [Figure-3].

Risk level	Risk status	Number of failures per week	Proportion of outage duration per week
1	Mild	Less than 2 times	Less than 30%
2	moderate	3-6 times	30%~80%
3	severe	More than 6 times	More than 80%

Figure 3. Classification of distribution network fault risk level.

3.3 OPTIMAL FAULT FEATURE SET SELECTION BASED ON RF FAULT LEVEL CLASSIFICATION

In order to eliminate irrelevant features and reduce the consumption of computing resources, RF algorithm is used to sort the importance of features and select the optimal fault feature set for subsequent calculation.

The calculation method of feature importance [15] is as follows. Supposed there is a bootstrap sample $b=1,2,\dots,B$, where B represents the number of training samples, and the variable importance measure \overline{D}_i of feature x_i based on classification accuracy is calculated according to the following steps:

- Let $b = 1$, the decision tree T_b is created on the training sample, and the data outside the bag is marked as L_b^{oob} ;
- 2) T_b is used to classify data L_b^{oob} on the data outside the bag, and the number of correct classification is counted, which is recorded as R_b^{oob} ;
- For feature x_i , $i=1,2,\dots,N$, the value of feature x_i in L_b^{oob} is perturbed, and the perturbed data set is recorded as L_{bj}^{oob} . T_b is used to classify the data L_{bj}^{oob} , and the number of correct classification is counted, which is recorded as R_{bj}^{oob} ;
- For $b=2,3,\dots,B$, repeat steps 1) ~ 3);
- The variable importance measure \overline{D}_i of feature x_i is calculated by the following formula:

$$\overline{D}_i = \frac{1}{B} \sum_{b=1}^B (R_b^{oob} - R_{bj}^{oob})$$

The sequential backward search method is used to remove a feature with the least importance score from the feature set each time, iterate step by step, and calculate the classification accuracy. Finally, the feature set with the least number of variables and the

highest classification accuracy is obtained as the feature selection result. Using cross validation method, the data set is divided into equal parts in each iteration, and used for training and testing according to the ratio of 7:3. The variable importance ranking generated by the highest classification iteration in the test set is selected as the basis for feature deletion, and the average classification accuracy of multiple iterations is taken as the classification accuracy of the iteration, and the final variable feature set is obtained.

3.4 FAULT LEVEL PREDICTION MODEL BASED ON MSVM

The training set containing the optimal features is sent to the built MSVM model for learning and training, and the fault level prediction is realized. Among them, the choice of kernel function of MSVM model is very important

Linear kernel function:

$$K(x_i, x_j) = x_i^T x_j$$

Polynomial (poly) kernel function:

$$K(x_i, x_j) = (x_i \cdot x_j + 1)^d$$

Radial basis function (RBF) kernel function

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$$

Where d is the number of polynomial functions; γ is a constant in the radial basis function.

Considering the distribution characteristics of the data used in the prediction, the radial basis function is generally selected as the kernel function. RBF kernel function can map a sample to a higher dimensional space. Compared with polynomial kernel function, RBF needs to determine fewer parameters. The number of kernel function parameters directly affects the complexity of the function. In addition, when the order of the polynomial is higher, the element values of the kernel matrix tend to infinity or infinitesimal, while RBF is higher, which will reduce the difficulty of numerical calculation.

When training MSVM with radial basis kernel, there are two parameters to consider: penalty coefficient C and RBF parameter. The penalty coefficient C has a certain influence on the fitting effect of the model. Specifically, the larger C , the better the fitting effect; The smaller C , the worse the fitting effect. In this paper, Grid Search method is used to optimize the penalty coefficient and kernel function parameters.

4 EXAMPLE RESULTS AND DISCUSSION

The data used in the experiment is the original data of distribution network related faults in 14 fault areas in 2013-2015 in a city. After eliminating the failure data caused by factors unrelated to meteorology such as construction impact and equipment aging, the pre-processing operations such as cleaning, conversion and integration are realized by using Python language, and the original data set can be used for fault prediction is obtained. The data set contains 1073 fault data, including 710 data with fault level 1, 276 data with fault level 2, 87 data with fault level 3, and training and test according to 7:3.

After analysis, 18 characteristics of the fault characteristics including temperature, humidity, rainfall, month and region classification are preliminarily determined. Through the iterative calculation of RF algorithm, the feature set with the least number of variables and the highest classification accuracy rate is selected. The final ranking results of feature importance are shown in [Figure-4]. The selected features include 11 features, including regional classification, monthly classification, weekly lightning strike rating, wind days above level 5, snow falling, etc.

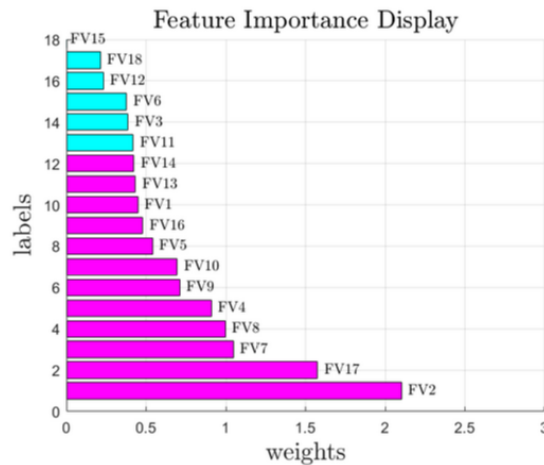


Figure: 4. Ranking results of feature importance by RF.

The new training set composed of the selected features and tags is sent to the MSVM model for training. In order to verify the rationality of the selected kernel function, three common basis functions are used, and five data sets are randomly divided into five groups for comparative experiments. The changes of the three training results and prediction accuracy are shown in [Figure-5].

After verification, the most stable and less complex RBF should be selected as the kernel function. The grid search algorithm is used to optimize the parameters. When the experimental training is about 25 times, the model can converge in most cases. Although there are some extreme cases leading to individual fluctuations in accuracy, the overall trend of high accuracy still remains stable, and the loss value also decreases to 0.05.

The change of model prediction accuracy with the number of training iterations is shown in [Figure-6].

The experiment shows that the prediction accuracy of the method can reach 96% on the data set used in this paper, which has practical application value.

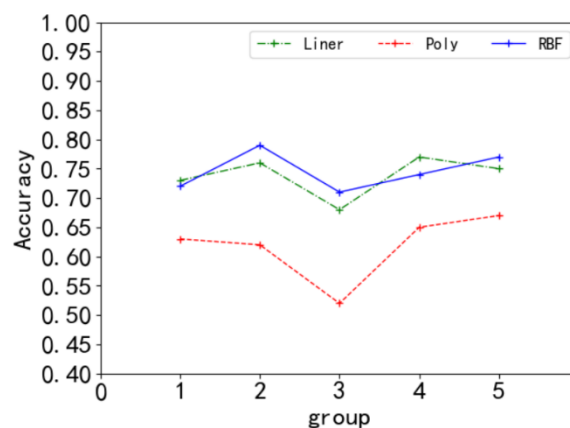


Figure: 5. Comparison of prediction accuracy of different kernel functions

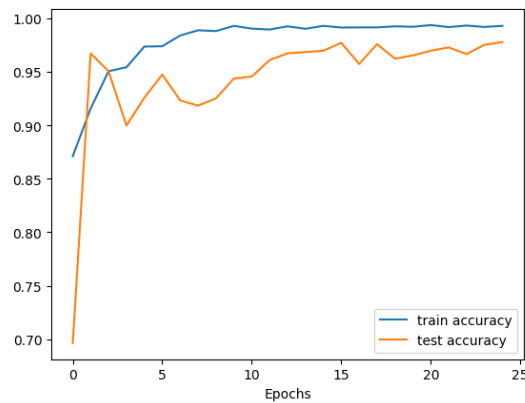


Figure: 6. Change of model prediction accuracy

[5] CONCLUSION

This paper presents a method of power grid fault prediction considering weather factors. Firstly, the optimal feature selection is based on RF, and the features are sorted by using the variable importance measure of Random Forest algorithm. The feature subset that can train the optimal performance classifier is found by using the backward sequence search method. Secondly, the prediction model of MSVM is established, and iterative training and parameter optimization are carried out. The trained model is verified to be reliable and stable. The proposed method provides an effective idea for fault prediction in the scenario of growing distribution network data, and has a good application prospect.

REFERENCES

- [1] S.Y. Ge, Z.H. Zhu, Hong Liu. "Comprehensive evaluation method for distribution network fault risk". In the Journal Of Proceedings of the CSU-EPSA 26(07) (2014).
- [2] B. Wen, Z.Y. Liu, B.Z. Hong. "Discussion on impact of weather on reliability of distribution networks". In the Journal Of Guangdong Electric Power 25(09) (2012).
- [3] W. Zhang, W.X. Sheng, K.Y. Liu. "A prediction method of fault risk level for distribution network considering correlation of weather factors". In the Journal Of Power System Technology, 42(08) (2008).
- [4] H. Liu, R.A. Davidson, A.M. Asce. "Negative Binomial Regression of Electric Power Outages in Hurricanes". In the Journal Of Journal of Infrastructure Systems, (2005).
- [5] D.T. Radmer, P.A. Kuntz, R.D. Christie. "Predicting Vegetation Related Failure Rates for Overhead Distribution Feeders". In the Journal Of IEEE Power Engineering Review, 22(9) (2007).
- [6] S.W. Fei, Yu Sun. "Fault prediction of power transformer by Combination of Rough Sets and Grey Theory". In the Journal Of Proceedings of the CSEE, 28(016) (2008).
- [7] A. Pahwa. "Modeling Weather-Related Failures of Overhead Distribution Lines". In proceedings of Power Engineering Society General Meeting, (2017).
- [8] Dian He, D.D. Tan, H.L. He. "Quantity prediction of distribution network faults based on Big Meteorological Data". In the Journal Of Journal of Jishou University, 040(001) (2014).
- [9] Wei Xi, Peng Li, X.B. Guo. "Application of Correlation Analysis Method based on multidimensional time series in prediction of equipment failure". In the Journal Of Advances of Power System & Hydroelectric Engineering, 000(012) (2014).
- [10] L. Breiman. "Random Forests". In the Journal Of Machine Learning, 45(1) (2001).
- [11] A.P. Sanil. "Principles of Data Mining". In the Journal Of Publications of the American Statistical Association, 98(461) (2007).
- [12] J.W. Han, M. Kamber, J. Pei. "Data mining: Concepts and techniques". In the Journal Of The Morgan Kaufmann Series in Data Management Systems, (2011).
- [13] J.A. Suykens, J. Vandewalle. "Least squares support vector machine classifiers". In the Journal Of Neural processing letters, (1999), pp. 293–300.
- [14] A. Mathur, G.M. Foody. "Multiclass and Binary SVM Classification: Implications for Training and Classification Users". In the Journal Of IEEE Geoscience and Remote Sensing Letters, 5(2) (2008).
- [15] D.J. Yao, Jing Yang, X.J. Zhan. "Feature selection algorithm based on random forest". In the Journal Of Journal of Jilin University, 44(01), 2014.