Journal on Big Data
DOI:10.32604/jbd.2021.024892
Article

A Lightning Disaster Risk Assessment Model Based on SVM

Jianqiao Sheng[1], Mengzhu Xu[2], Jin Han[3,*] and Xingyan Deng[2]

[1]Information and Communication Branch of State Grid Anhui Electric Power Co., Ltd., Hefei, 230041, China
[2]Information and Communication Branch of State Grid Shanxi Electric Power Co., Ltd., Taiyuan, 030000, China
[3]School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing, 210044, China
[*]Corresponding Author: Jin Han. Email: hjhaohj@126.com
Received: 05 November 2021; Accepted: 18 November 2021

Abstract: Lightning disaster risk assessment, as an intuitive method to reflect the risk of regional lightning disasters, has aroused the research interest of many researchers. Nowadays, there are many schemes for lightning disaster risk assessment, but there are also some shortcomings, such as the resolution of the assessment is not clear enough, the accuracy rate cannot be verified, and the weight distribution has a strong subjective trend. This paper is guided by lightning disaster data and combines lightning data, population data and GDP data. Through support vector machine (SVM), it explores a way to combine artificial intelligence algorithms with lightning disaster risk assessment. By fitting the lightning disaster data, the weight distribution between the various impact factors is obtained. In the experiment, the probability of lightning disaster is used to compare with the actual occurrence of lightning disaster. It can be found that the disaster risk assessment model proposed in this paper is more reasonable for the lightning risk. It has been verified that the accuracy rate of the assessment model in this paper has reached 80.2%, which reflects the superiority of the model.

Keywords: Lightning disaster; lightning risk assessment; SVM

1 Introduction

In recent years, due to the global warming, extreme weather has occurred frequently. As a natural disaster, lightning disaster greatly threaten the daily life of human beings, causing a large number of casualties and economic losses [1–2]. Various industries such as aviation, power facilities, railway transportation, and telecommunications are all affected by it [3]. In the cases of casualties caused by meteorological disasters, lightning disasters account for almost the vast majority. Therefore, related researchers have aroused the exploration of the law of lightning activities. Lightning disaster risk assessment, as a method of statistics and research on the disaster risk of specified areas, has made considerable progress. Some countries have also designated relevant industry standards to guide the departments to carry out lightning disaster risk assessment.

At present, many researchers have conducted in-depth research on the risk assessment of lightning disasters. Most studies on the lightning disaster risk assessment in specified areas use the analytic hierarchy process (AHP) to classify the issues, as shown in Fig. 1, starting from the three aspects of disaster-causing factors, disaster-pregnant environment and disaster-resistance capabilities to calculate and assess the risk value of the area. These three aspects are composed of many different factors, and the choice of factors in the research of different researchers also varies. This paper lists several main influencing factors in each aspect in Fig. 1.

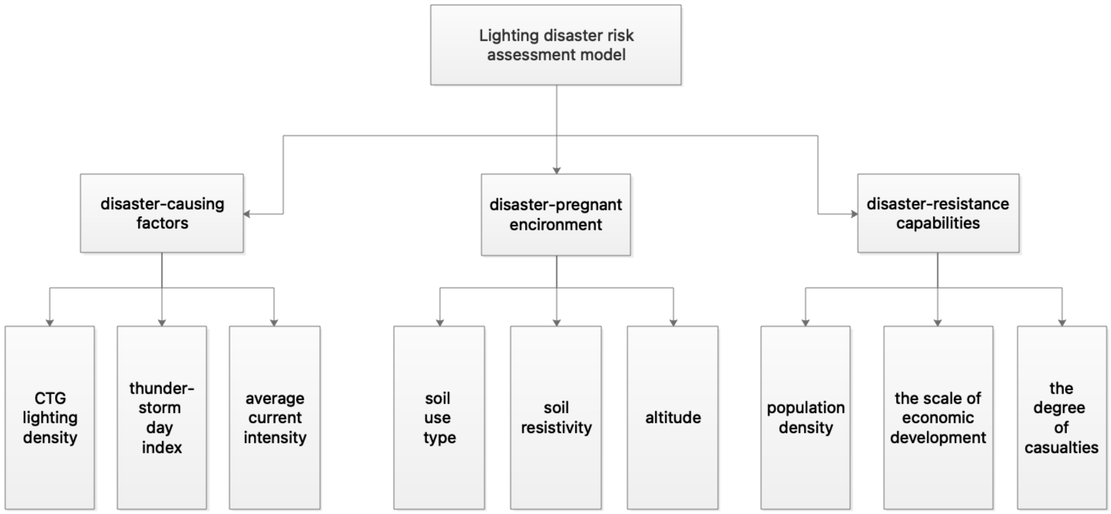

Figure 1: Hierarchical division of lightning disaster risk assessment

In terms of disaster-causing factors, the cloud-to-ground (CTG) lightning density, thunderstorm day index and average current intensity are the more prominent indicators. The cloud-to-ground lightning density refers to the number of lightning in a unit area; the thunderstorm day index refers to the days of lightning discharge in a specified area for one year. The lightning current intensity is the average value of the current intensity (unit: kiloampere) of the return strike in the specified time period and within the specified area. The disaster-pregnant environment is mainly composed of factors such as soil utilization type, altitude, and soil resistivity. Soil use type refers to the land use category of the area, which can generally be industrial land or agricultural land, etc.; altitude refers to the height difference between the area and sea level; soil resistivity is the product of average value of soil resistance per unit length and the cross-sectional area. In terms of disaster resistance, there are generally influencing factors such as population density, scale of economic development, and scale of loss of life. Population density is the number of people per unit area; the scale of economic development is the GDP value of the region; the scale of loss of life is the average annual number of casualties per unit area caused by lightning disasters in the region. This indicator can objectively reflect regional causes. The degree of casualties caused by lightning disasters also reflects the ability of people in the area to defend against lightning disasters. The core work after selecting the relevant factors is the establishment of the weight distribution method and the index evaluation system [4–5]. Li et al. [6] used a weighted comprehensive evaluation algorithm and used factors such as lightning density, lightning disaster frequency, property loss and life loss as evaluation indicators to carry out a research on the lightning disaster risk zoning of Sanming City. Cheng et al. [7] combined the weighted comprehensive evaluation method with statistical method, disaster analysis method and expert scoring method to comprehensively evaluate impact factors.

They used thunderstorm days and CTG lightning density as two factors to analyze which areas in Anhui Province are most vulnerable to lightning disasters. Cui et al. [8] used the weighted comprehensive evaluation method and the AHP to evaluate the disaster risk and vulnerability of Nanjing according to the standard mathematical formula of natural disaster risk and the conceptual framework of flood disaster risk. The AHP is based on the opinions of experts and researchers in various fields. Wang et al. [9] used prefectures and cities as analysis units to evaluate the vulnerability of lightning disaster areas using the information method, and then used the reverse derivation method to evaluate and zoning of the vulnerability of Yunnan Province. Recently, Chen et al. [10] and Liu et al. [11] conducted a more comprehensive zoning study on the basis of previous research results by introducing parameters such as population distribution and soil conductivity. Although there are few related studies in this field abroad, some international researchers have also put forward some meaningful studies. Biswas et al. [12] proposed a GIS and IDW statistical model for spatial vulnerability analysis of lightning disasters, which can determine the spatial heterogeneity of lightning risks. Nastos et al. [13] used precision lightning network (PLN) to process lightning data and analyzed the characteristics of lightning disaster risk from the perspective of time and space.

The above-mentioned related methods can make a relatively accurate assessment of the lightning disaster risk to a certain extent, but there are still three major problems: 1. The results of the assessment are not granular enough; 2. The weight setting of the impact factors is often more one-sided; 3. The assessment results are often only driven by impact factor data, and the connection between the setting of the evaluation model and the actual situation of lightning disasters is not close. Aiming at the above problems, this paper develops a rasterized lightning disaster risk assessment model based on SVM, which is driven by lightning disaster data, improving the scientificity of the model.

2 Methodology

Before setting up the model, this paper statistically displays lightning disaster data and lightning data from 2010 to 2018. From the Fig. 2, it can be roughly seen that the more lightning activities occur in the year, the more frequent lightning disasters occur. Therefore, it can be concluded that lightning activities are closely related to the occurrence of lightning disasters.

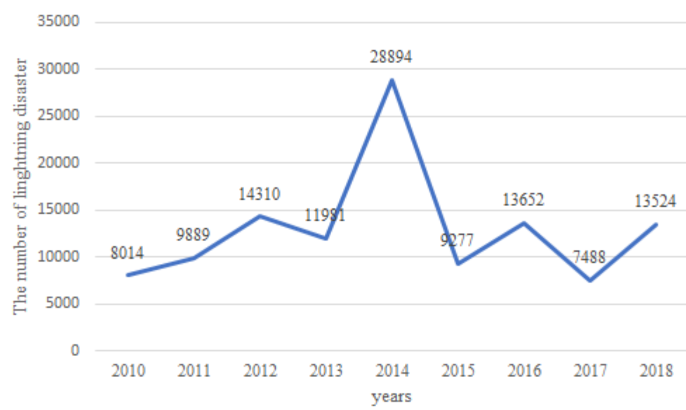

Figure 2: The number of lightning disaster varies with the years

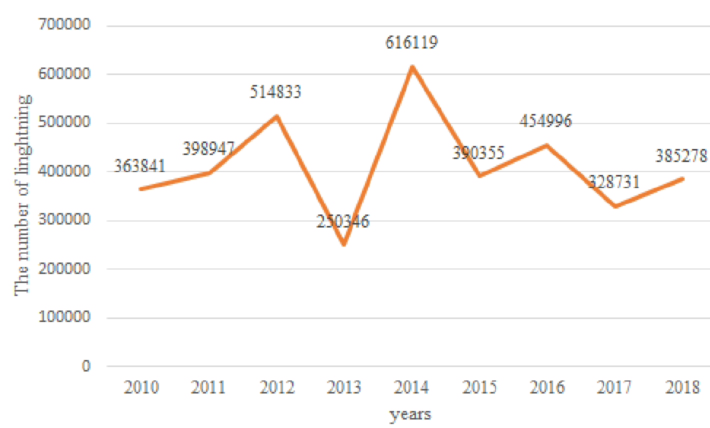

Figure 3: The number of lightning varies with the years

This paper first collects relevant data and rasterizes various data. The so-called rasterization refers to dividing the studied area into 1 km * 1 km geographic grids from the geographic level, and then putting the collected data into the corresponding grid. Secondly, we make the data in each grid dimensionless, which allows all the data to be unified in dimension so that subsequent calculations can be convenient. Then, the data is randomly shuffled, 70% of which is used as the train set and 30% as the test set. With the help of SVM [14–15], We draw the final conclusion. The specific calculation process is shown in Fig. 4.

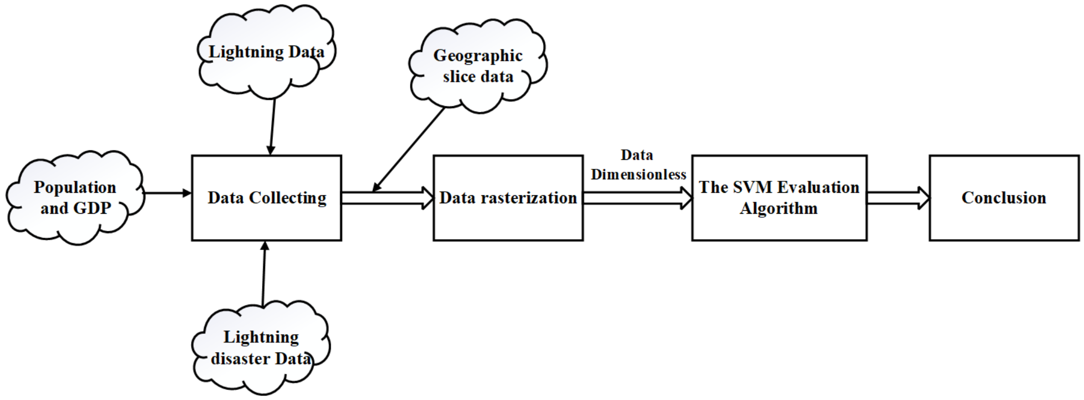

Figure 4: The calculation process of the model

2.1 Data Collection

The data in this paper collected the lightning disaster data (Table 1), ADTD lightning location data (Table 2), geographic slice data, population and GDP data of Hunan Province from 2010 to 2018.

Table 1: Lightning disaster data

Latitude	Longitude	Industry	Level	Casualty (Person)	Economic Losses (Thousand)	Type	City	County
26.6	112.4	housing	3	2	130	direct thunder	Hengyang	Hengnan
27.2	111.6	countryside	3	1	50	direct thunder	Shaoyang	Shaodong
28.3	113.1	flammability	1	1	600	direct thunder	Changsha	Liuyang
29.1	110.5	communication	1	0	16.5	direct thunder	Zhangjia Jie	Sangzhi

Table 2: ADTD lightning location data

Date	Time	Latitude	Longitude	Power	Steepness	Deviation	Position Way	Province	City	County
2015 /3/22	21:35:21	28.2	110.4	–23.7	–6	0	TSTDDF	Hunan	Huaihua	Yuanling
2015 /3/22	22:48:22	28.8	112.5	–40.7	–15.3	74	TSTDDF	Hunan	Yiyang	Yuanjiang

2.2 Data Processing

Step 1: Put the data into the corresponding geographic grid

According to the geographic spatial extent of each grid and the latitude and longitude of each lightning, the geographic grid to which each lightning belongs can be determined. Similarly, rasterize the lightning disaster data to determine its corresponding geographic grid.

Step 2: Average the intensity of lightning data in the grid

In actual situations, each geographic grid contains multiple pieces of lightning data, and the intensity of the lightning data in the grid is averaged. The average lightning intensity of each grid is:

$$G_k = \frac{\sum_{i=0}^{n} S_i}{n} \tag{1}$$

where k is the grid number, G_k represents the average lightning intensity in grid k, and n represents n lightning data in grid k. S_i is the i-th lightning intensity of the lightning data in grid k.

Step 3: Add labels to the geographic grid

According to the collected lightning disaster data, the geographic grids that have experienced lightning disasters are marked as 1, which is called a positive sample. The grids that have not experienced lightning disasters are marked as –1, which is called a negative sample.

Step 4: Make the data dimensionless

In order to calculate the data of different dimensions together, the "maximum and minimum dimensionless data standardization" method is selected in the model. The log function is used to normalize the data. The formula is:

$$X^* = \frac{\log_{10} x_i}{\log_{10} x_{mac}} \tag{2}$$

s the i-th sample data, x_{max} is the maximum value of the sample data.

After completing the above four steps, a standard data set that can be used in the experiment has been constructed.

2.3 Construction of Lightning Disaster Risk Assessment Model Based on SVM

We consider that lightning disaster risk assessment is a two-category problem in the experimental stage, that is, lightning disasters occur or not, so we introduce the idea of SVM to train and experiment the model.

This paper selects lightning data, lightning disaster data, and population GDP data as experimental data, and constructs a function F such that:

$$Y = F(w, l, S, G, P) \tag{3}$$

Among them, l is the lightning frequency data of the grid, S is the average current intensity data of the grid, G is the GDP data, P is the population data, and Y is the actual result, that is, whether there is a lightning disaster.

We record the i-th geographic grid data as the vector x_i, $x_i = (l,\ s,\ G, P\ )$, and record y_i as whether there has been a lightning disaster in the i-th geographic grid, and record it as 1, if it has occurred. There is the entire data set D, and the expression of D is shown in Eq. (4):

$$D = \{(x_i, y_i), (x_i, y_i), \ldots, (x_i, y_i)\ \} \tag{4}$$

We believe that there is a certain correspondence between the data vector x_i in each geographic grid and its label y_i. In high-order dimensions, this problem is a linearly separable problem. There exists a hyperplane in high-order dimensions for data points that do not have lightning disasters and data points that will have lightning disasters [16–17]. The data points we record above the hyperplane will not have lightning disasters, that is, $y_i > 1$. The data points we record below the hyperplane will have lightning disasters, that is, $y_i > 1$. The distance we record between the data vector x_i of each geographic grid in the high-order dimension and the hyperplane is r, then the expression of r is shown in Eq. (5):

$$r = \frac{w^T x + b}{||w||} \tag{5}$$

Then there is an expression for the distance from a point above the hyperplane to the hyperplane as:

$$r = \frac{w^T x + b}{||w||} > 0 \tag{6}$$

In the same way, the expression for the distance from the point below the hyperplane to the hyperplane is:

$$r = \frac{w^T x + b}{||w||} < 0 \tag{7}$$

For the support planes on both sides, there is an expression for the distance from a point above the support plane to the support plane:

$$r = \frac{w^T x + b}{||w||} > 1 \tag{8}$$

In the same way, the expression for the distance from a point above the support plane to the support plane is:

$$r = \frac{w^T x + b}{||w||} < -1 \tag{9}$$

The point on the support plane is the support vector. By scaling the w and b values of the hyperplane, the function distance from the support vector to the hyperplane is 1. The support vector is the point closest to the hyperplane, so other vector points are to the hyperplane. So the function distance of must be greater than or equal to 1. In fact, the initial model can be established at this time, and the objective function F is:

$$F = max \frac{2}{||w||} \tag{10}$$

$$s.t.\ y_i(w^T x + b) \geq 1 \tag{11}$$

Through data training, we can finally get the optimal hyperplane, which is the linearly separable function we require.

3 Experimental Results

3.1 Evaluation Criteria and Accuracy

We record the model output as the probability P_i of the occurrence of lightning disasters in the grids, and binarize the output of the model during the model reliability assessment stage, that is, if $P_i \geq 0.5$, it is considered that the grid is prone to lightning disasters, then this value is defined as 1. If $P_i < 0.5$, it is considered that the grid is not prone to lightning disasters, then the value is defined as 0, and then the result of the binarization process is compared with the real situation to evaluate the reliability of the model. It is defined as follows:

Definition: After binarization, if the evaluation value of the geographic grid corresponding to the actual lightning disaster is 1, the evaluation made by the model is a correct evaluation. Otherwise, it is an erroneous evaluation. The overall accuracy $P_{accuracy}$ of the model is calculated as follows:

$$P_{accuracy} = \frac{N}{T} \tag{12}$$

Among them, N represents the number of correct evaluations, and T represents the total number of lightning disaster grids in the test set. The accuracy $P_{accuracy}$ reflects the reliability of the model used for lightning disaster risk assessment.

In order to improve the accuracy of the experimental process and experimental results, this model uses a cross-validation method to test the accuracy of the model, that is, all data is randomly scrambled, and each time 70% of the data set is used as the test set and 30% as the training set. Let P be the final accuracy rate, and P_i is the accuracy rate of the i-th experiment. The specific verification calculation formula is shown below:

$$P = \frac{1}{n}\sum_{i=1}^{n} P_i \tag{13}$$

In the process of experiment, the value of n is 5, and the accuracy rate of each experiment is shown in Table 3.

Table 3: Experimental accuracy

Number	N	T	The accuracy rate
1	119	150	79.3%
2	105	148	70.9%
3	143	168	85.1%
4	99	122	81.1%
5	99	115	85.6%
Average	565	703	80.3%

After cross-validation, the evaluation accuracy of this model is 80.2%, which reflects the superiority of the model.

3.2 Display the Evaluation Results

In this paper, the data of 2010 to 2018 years in Changsha is used for calculation, the output of the model is visually displayed, and the output value of each raster is segmented and colored by the Jenks natural breakpoint algorithm. In Fig. 5, we divide the entire risk area into five levels. The red area indicates the area with a higher probability of lightning disasters, and the green area has the lowest risk.

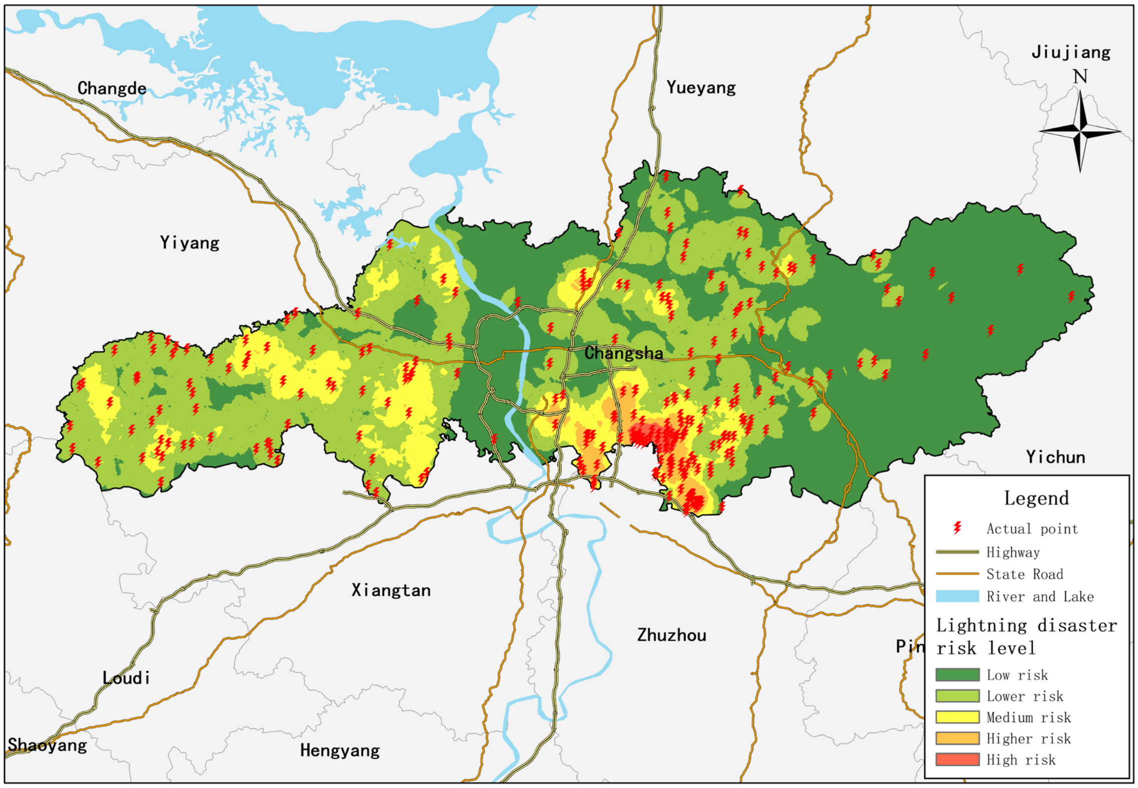

Figure 5: The map of lightning disaster risk assessment results and actual location

4 Conclusion

On the basis of previous research, this paper introduces the idea of geographic rasterization to further improve the evaluation model. By using the SVM, driven by lightning disaster data, in the training process of the algorithm, we continuously adjust and optimize the weight of each impact factor until the entire evaluation model is constantly approaching the final result, and finally a model with the best generalization and the highest evaluation accuracy is fitted. In future research, we can continue to explore the advantages of artificial intelligence algorithms for lightning disaster risk assessment, and try to introduce more impact factors to develop a more complete and accurate lightning disaster risk assessment mechanism.

Funding Statement: This work is supported by Network Security Inspection and Monitoring System Project of State Grid Corporation of China.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] W. Xiao and H. Chen, "National lightning disaster analysis and economic loss forecast," *Meteorological and Environmental Sciences*, vol. 33, no. 4, pp. 21–26, 2010.

[2] S. Yang, G. Zhao, S. Pan, C. Zhao, Y. Niu *et al.,* "Spatiotemporal distribution characteristics and early warning of lightning disaster in China," *Journal of Natural Disasters*, vol. 19, no. 6, pp. 153–159, 2010.

[3] E. Renni, E. Krausmann and V. Cozzani, "Industrial accidents triggered by lightning," *Journal of Hazardous Materials*, vol. 184, no. 1–3, pp. 42–48, 2010.

[4] Y. Tian, M. Wu, F. Shi and X. Wei, "Evaluation and regionalization of lightning disaster vulnerability over Hebei," *Meteorological Science and Technology*, vol. 40, no. 3, pp. 507–512, 2012.

[5] G. Chen, X. Cui, X. Tian, "Quantitative assessment and zoning of lightning hazard in Jiangsu Province," *Journal of Catastrophology*, vol. 32, no. 1, pp. 32–35, 2017.

[6] Y. Li, L. Wu, C. Yu, W. Lin and T. Wang, "Lightning characteristics and lightning disaster vulnerability zoning in Sanming of Fujian Province," *Journal of Nanjing University of Information Science & Technology (Natural Science Edition)*, vol. 9, no. 2, pp. 220–226, 2017.

[7] X. Cheng, W. Xie, K. Wang and H. Tian, "Method research of the risk zoning of thunder and lightning and its application over Anhui Province," *Journal of the Meteorological Sciences*, vol. 32, no. 1, pp. 80–85, 2012.

[8] X. Cui, Y. Zhuang and H. Wang, "Combined evaluation method-based risk zoning of lightning disaster in Jiangsu Province," *Journal of Natural Disasters*, vol. 24, no. 6, pp. 187–194, 2015.

[9] Y. Deng, L. Yin, Y. Xu and Y. Jing, "Vulnerability analysis and zoning of lighting disaster in Yunnan Province," *Meteorological Monthly*, no. 12, pp. 83–87, 2007.

[10] L. Chen, S. Long, H. Yu, Q. Li, X. Chen *et al.,* "Research on risk zoning of lightning disaster in Hebei Province," *Journal of Catastrophology*, vol. 34, no. 3, pp. 189–195, 2019.

[11] P. Liu, T. Zhang and Q. Zhou, "Analysis of the causes of lightning disasters in Kunming City from 2006 to 2016 and vulnerability risk zoning," *Journal of the Meteorological Sciences*, vol. 38, no. 6, pp. 824–831, 2018.

[12] R. N. Biswas, M. N. Islam, M. J. Mia and M. N. Islam, "Modeling on the spatial vulnerability of lightning disaster in Bangladesh using GIS and IDW techniques," *Spatial Information Research*, vol. 28, no. 5, pp. 507–521, 2020.

[13] P. T. Nastos, I. T. Matsangouras and T. G. Chronis, "Spatio-temporal analysis of lightning activity over Greece–Preliminary results derived from the recent state precision lightning network," *Atmospheric Research*, vol. 144, pp. 207–217, 2014.

[14] X. Wu, W. Zuo, L. Lin, W. Jia and D. Zhang, "F-SVM: Combination of feature transformation and SVM Learning via convex relaxation," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 11, pp. 5185–5199, 2018.

[15] S. Ding, B. Qi and H. Tan, "An overview on theory and algorithm of support vector machines," *Journal of University of Electronic Science and Technology of China*, vol. 40, no. 1, pp. 2–10, 2011.

[16] W. Hu, C. Zhao, Y. L. Sun and J. Jun, "Evaluation of whole span lightning shielding flashover risk based on 3-D Laser scanning technology," *IEEE Transactions on Power Delivery*, vol. 29, no. 2, pp. 794–801, 2013.

[17] E. Volpov and R. Linder, "Application of local environmental data to the lightning performance assessment of the IECo transmission-line network," *IEEE Transactions on Power Delivery*, vol. 32, no. 6, pp. 2546–2554, 2017.