

RESEARCH ARTICLE

Prediction Algorithm for Power Outage Areas of Affected Customers Based on CNN-LSTM

WEIXIANG HUANG, WEI ZHANG^{ID}, QIANYI CHEN, BO FENG, AND XIN LIElectric Power Research Institute, Guangxi Power Grid Company Ltd., Nanning 530023, China
Guangxi Power Grid Equipment Monitoring and Diagnosis Engineering Technology Research Center, Nanning 530023, China

Corresponding author: Wei Zhang (zw256257282931@163.com)

This work was supported in part by the Research on Flood Warning Technology and Development of Main Station System Functions under Grant GXKJXM20220109.

ABSTRACT Predicting outage areas for affected customers in a disaster requires real-time meteorological data, which can be challenging to obtain and process with high quality. In particular, missing or inaccurate data can occur in certain regions or during extreme weather events. Therefore, a Convolutional Neural Network-Long Short-Term Memory (CNN-LSTM) based algorithm is proposed to improve the accuracy and reduce the prediction time in the outage region. Ground-based automated weather observatories are used to obtain real-time weather data, including extreme weather data. These data are pre-processed to enhance data quality and accuracy by removing duplicates, filling in missing values, handling anomalies, normalizing input variables, and reducing dimensionality. Based on the results of the data pre-processing, the outage rate was calculated for different types of meteorological disasters and the geographical characteristics of the outages were analyzed. This analysis provides insights into the impact of different types of meteorological disasters on power outages and helps improve the accuracy of predictive models. The proposed algorithm employs a CNN neural network to capture spatial and temporal information from raw meteorological data by stacking convolution and pooling layers. The extracted features are then organized and output by fully connected layers, laying the foundation for subsequent time series modeling. An LSTM network is further utilized to construct a prediction model for the outage area, which takes as input the feature extraction results of the meteorological data. By integrating the temporal dimension information of meteorological data, the model outputs accurate predictions for the outage area. The experimental results demonstrate a consistent outcome with the actual test results, achieving high prediction accuracy with a short prediction time of 4.3 s and a maximum non-outage detection rate of 2 %. Therefore, the proposed algorithm proves to be significant for accurate and fast prediction of outage areas for affected customers in real world applications.

INDEX TERMS Affected customers, CNN-LSTM, convolutional layer, pooling layer, prediction of power outage areas.

I. INTRODUCTION

Stable power supply is crucial for economic development and residents' livelihoods. Power interruptions, especially large-scale blackouts, not only cause significant economic losses to power companies but also can disrupt critical infrastructure and result in substantial losses for the country and its people, even leading to public crises [1], [2]. In recent years, the frequency of extreme weather events has increased, making the

extensive coverage of power grids more susceptible to natural factors. To ensure efficient emergency responses for affected customers [3], [4], [5], it is necessary to forecast blackout areas and identify precise geographical locations. Therefore, stable power production and supply are vital for maintaining social stability, residents' well-being, and national security, and should be given adequate attention.

In order to improve the disaster resistance of the power grid, scholars in related fields have conducted research on power outage area prediction. Reference [6] proposed a statistical learning based method for predicting power outage space

The associate editor coordinating the review of this manuscript and approving it for publication was Chi-Seng Lam^{ID}.

under typhoon disasters, taking into account factors such as power grid, meteorology, and geography. After collecting and processing data, a power outage space prediction model was constructed using multiple algorithms. However, this method only considers factors such as power grid, meteorology, and geography, and other potential influencing factors may not have been fully considered, which affects the accuracy of the model's prediction. Reference [7] addresses the difficulty of predicting power outage events, analyzes the application scenarios of probability models in this field, and proposes a comprehensive model including admittance models and fault probability models. The experimental results show that this method can effectively predict power outage events and provide reference for power grid structure design. However, this study mainly focuses on the estimation of the probability of power outage events, and further expansion and improvement may be needed for the prediction of specific affected customer power outage areas. Reference [8] proposed a power outage classification and prediction method based on Bagging ensemble learning. This paper applies the Bagging ensemble model to predict power outage problems. Firstly, based on the geographical location relationships of different regions, construct spatial location matrices for different regions, and use Quick Response (QR) matrix decomposition to construct spatial features. Then, the Bagging ensemble learning framework is used to randomly resample the data, train different basic classifiers, and integrate the classifiers based on the combination strategy. Finally, this paper uses ensemble learning models to predict data. This algorithm has good performance. Reference [9] proposed an algorithm for predicting natural disaster power outages using synthetic distribution systems. This algorithm generates a comprehensive power system layout for any American city based solely on public data, and then uses a vulnerability function to simulate the power outage of a single building under hazardous loads. This algorithm provides a more localized building level estimate of the likelihood of power loss caused by natural disasters. However, the above two algorithms require a longer time to predict customer power outage areas, resulting in lower prediction efficiency.

CNN-LSTM combines the spatial feature extraction capability of CNN for time series data with the temporal modeling capability of LSTM for dynamic changes over time. This method can capture the spatial features of time series data while handling the temporal dynamics. This combination offers high prediction accuracy and stability. Therefore, this paper proposes a CNN-LSTM based algorithm for predicting the power outage area of affected customers. The algorithm utilizes meteorological data obtained from ground automatic weather observation stations, and applies data preprocessing and analysis to determine the geographical characteristics of power outages. It employs CNN neural networks to extract effective features, and then utilizes LSTM networks to predict power outage areas of affected customers. Through simulation experiments, the algorithm can quickly and accurately predict the power outage area of affected customers, laying

a foundation for the stable operation of the power grid in affected areas.

II. PREDICTION OF POWER OUTAGE AREAS FOR AFFECTED CUSTOMERS

A. METEOROLOGICAL DATA ACQUISITION

Ground automatic meteorological observation stations are specialized equipment for collecting meteorological data, with high measurement accuracy and precision. This makes the meteorological data obtained through these observation stations more reliable and can serve as the basic data for experiments. The use of ground automatic meteorological observation stations can collect relevant initial data, in addition to real-time meteorological data obtained through direct observation (rainfall; maximum, average, and minimum temperatures; maximum, average, and minimum humidity; average wind speed, etc.), there are also extreme meteorological data. This paper obtains meteorological data from both real-time and extreme meteorological aspects.

1) REAL TIME METEOROLOGICAL DATA

Meteorological factors are variable at various times of the day and exhibit real-time volatility. Ground meteorological observation stations update meteorological data every hour, including station number (Station_Id_C), year (Year), month (Mon), day (Day), and hour (Hour). The 20 meteorological elements can be divided into 4 categories, specifically wind speed, pressure, humidity, and temperature.

Among them, the wind speed elements include: 2-minute average wind speed (WIN_SAvg_2mi), meters per second; Maximum wind speed (WIN_S-Max), meters per second; Maximum wind speed (WIN_S_Inst-Max), meters per second; 2-minute average wind direction (angle) (WINDAvg_2mi), degrees (degrees); Wind direction (angle) and degree (°) of maximum wind speed; Wind direction (angle) and degree (°) of maximum wind speed (WIND_INSTMAX); Wind power, level.

The atmospheric pressure elements include: sea level pressure (PRS-Sea), hPa; Air pressure (PRS), hPa; Maximum pressure (PRS.Max), hPa; Minimum air pressure (PRSMIn), hPa.

Humidity factors include: water vapor pressure (VAP), hundred pascals (hPa); Precipitation (PRE_1h), millimeters; Relative humidity (RHU), percentage (%); Minimum relative humidity (RHU-Min), percentage (%); Horizontal visibility (VIS), meters.

Air temperature elements include: apparent temperature (tigan), Celsius (°C); Temperature (TEM), Celsius (°C); Maximum temperature (TEM-Max), degrees Celsius (°C); Minimum temperature (TEM-Min), Celsius (°C).

Based on the above analysis, meteorological factors are obtained, including four indicators: actual temperature, temperature and humidity index, cold and humidity index, and human comfort. The calculation formula for the four comprehensive meteorological indicators is as follows: T represents

Celsius temperature, Rh represents relative humidity, and V represents wind speed.

Real temperature refers to the conversion of human perception under different V , Rh , T conditions to a comfortable temperature situation under static and saturated atmospheric conditions. The calculation formula is:

$$Te = 37 - \frac{37 - T}{0.68 - 0.14Rh + 1/(1.76 + 1.4V^{0.75}) - 0.29T(1 - Rh)} \quad (1)$$

The temperature and humidity index reflects the comprehensive sensory level of the human body under the two meteorological factors [10] Rh , T , and its calculation formula is:

$$THI = 1.8T + 32 - 0.55(1 - Rh)(1.8T - 26) \quad (2)$$

The cold humidity index is an indicator to measure the degree of coldness, and its calculation formula is:

$$Ee = (33 - T)(3.3\sqrt{V} - V/3 + 20)e^{0.005|Rh-40\%|} \quad (3)$$

The comfort index of the human body is an evaluation index for the comprehensive perception of meteorological factors acting on the human body, reflecting the comfort of the human body under the joint action of multiple meteorological factors [11]. Its calculation formula is:

$$k = 1.8T - 0.55(1.8T - 26)(1 - Rh) - 3.2\sqrt{V} + 3.2 \quad (4)$$

2) EXTREME METEOROLOGICAL DATA

Meteorological disaster factors are the main cause of power outages and faults among affected customers. According to the types of disasters caused by meteorological factors, they can be divided into the following causes:

S_1 : Power outage failure caused by lightning for affected customers;

S_2 : The harm of mountain fires caused by drought to the power grid, resulting in power outages for affected customers;

S_3 : Low temperature caused line icing, resulting in power outages for affected customers;

S_4 : Wind deviation of transmission lines caused by strong winds, resulting in power outages for affected customers;

S_5 : Geological disasters caused by rainstorm, such as flood and collapse, directly or indirectly damage the power grid and cause power failure of affected customers [12].

a: LIGHTNING STRIKE FACTOR

Calculate the risk component of power outage losses for affected customers caused by lightning strikes [13]:

$$S_1 = N_{thunder1} \times P_{thunder1} \times L_{thunder1} \quad (5)$$

In the formula, $N_{thunder1}$ represents the number of lightning strikes per year in the region, $P_{thunder1}$ represents the probability of lightning damage caused by power outages of affected customers, and $L_{thunder1}$ represents the power outage loss rate caused by lightning damage.

b: MOUNTAIN FIRE FACTOR

Mountain fires (also known as forest fires) caused by dry weather usually occur in the winter and spring seasons (the first quarter of each year) in Yunnan. The causes of fires include human activities, lightning, spontaneous combustion, etc. The probability of their occurrence is closely related to the level of forest fire risk. Forest fires have a significant impact on the power system, characterized by difficulty in rescue, long duration of harm, and generally permanent faults. According to statistics, in the first quarter of 2010, China Southern Power Grid experienced 262 faults and trips on lines with voltage levels above 220kV, including 128 trips and power outages caused by wildfires, accounting for 48.9% of the total number of trips and power outages. Calculate the risk component of power outage losses for affected customers caused by wildfires [14]:

$$S_2 = N_{fire2} \times P_{fire2} \times L_{fire2} \quad (6)$$

In the formula, N_{fire2} represents the number of occurrences of wildfire events, P_{fire2} represents the probability of wildfire damage caused by power outages of affected customers, and L_{fire2} represents the power outage loss rate of wildfire damage.

c: ICING FACTOR

Icing is a complex process, and the amount of icing is related to the conductor radius, Supercooled water drop diameter, air volume, wind speed, wind direction, air temperature, icing time and other factors. Icing leads to a doubling of the weight of the transmission line, an increase in sag, and subsequently a flashover accident; During the ice melting period, ice flashover is easily formed, and continuous arc may burn the insulator, causing a decrease in the insulation strength of the insulator; Iced wires are prone to galloping, which can cause damage to towers, wires, ground wires, hardware and components, and may also cause serious accidents such as frequent trips, power outages, broken wires, and tower collapses. Calculate the risk component of power outage losses for affected customers caused by icing [15]:

$$S_3 = N_{ice3} \times P_{ice3} \times L_{ice3} \quad (7)$$

In the formula, N_{ice3} represents the number of occurrences of icing events, P_{ice3} represents the probability of icing damage caused by power outages of affected customers, and L_{ice3} represents the power outage loss rate of icing damage.

d: WIND BIAS FACTOR

The reclosing rate of wind bias tripping is very low, which is due to the continuity of wind during reclosing, which reduces the gap between the conductor and the tower. At the same time, there is a high possibility of operating overvoltage in the system during reclosing, so the second breakdown may occur when the gap distance is large. Calculate the risk component of power outage losses for affected customers caused by wind

bias:

$$S_4 = N_{\text{wind4}} \times P_{\text{wind4}} \times L_{\text{wind4}} \quad (8)$$

In the formula, the number of occurrences of N_{wind4} wind bias events, P_{wind4} represents the probability of wind bias damage caused by power outages of affected customers, and L_{wind4} represents the power outage loss rate of wind bias damage.

e: RAINSTORM FACTOR

Due to increased rainfall, it is easy to cause floods to collapse, leading to secondary disasters such as mudslides and landslides, which have a significant impact on the transmission line towers and lead to regional power outages. Calculate the risk component of power failure loss of the affected customers caused by rainstorm:

$$S_5 = N_{\text{rain5}} \times P_{\text{rain5}} \times L_{\text{rain5}} \quad (9)$$

where, N_{rain5} is the number of rainstorm events, P_{rain5} is the probability of rainstorm damage caused by power failure of the affected customer, and L_{rain5} is the rate of power failure loss caused by rainstorm damage.

B. METEOROLOGICAL DATA PREPROCESSING

Although sufficient meteorological observation data has been obtained in this paper, due to uneven distribution of meteorological observation stations, sudden network instability of observation station monitoring equipment, equipment failures, and other issues, the obtained meteorological data is missing or omitted in time, causing certain interference to the later analysis of meteorological data. In addition, the data collected by the meteorological observation station may have outlier, wrong values, duplicate values and other problems that do not conform to the current situation. Therefore, in order to ensure the accuracy and rationality of meteorological data, and facilitate the later prediction and analysis of meteorological data, it is necessary to carry out relevant data preprocessing work on the collected original meteorological data. The data preprocessing work includes the following steps:

(1) Delete duplicate values

Due to the presence of a series of duplicate values in the collected meteorological dataset, this column has no significance for predicting future power outage areas. Therefore, this column has been deleted.

(2) Fill in missing measurement values

Due to sudden failures, signal interruptions, and recording negligence in meteorological data collection equipment, there is a phenomenon of data missing in the dataset. The larger the amount of data, the higher the probability of this phenomenon occurring, and auxiliary methods need to be used to fill in the data. Due to the use of Python language tools in this paper, the `isnull` function in the Pandas library is used to determine missing values, and the `fillna` function is used to fill in the median values of the missing data.

(3) Handling outlier

Some element values recorded by meteorological observation stations are large or small, which obviously deviate from the actual situation. This is called outlier. However, in the face of massive datasets, using manual processing will inevitably result in excessive workload. Therefore, the use of mathematical means and the use of computers to deal with outliers is undoubtedly a very necessary and practical way.

According to the characteristics of the original meteorological data set selected in this paper, the Rayda criterion is selected to detect outlier. Assuming that the meteorological element data set X follows normal distribution, the outlier is judged according to the following formula:

$$P(|x - u| > 3\delta) \leq 0.0027 \quad (10)$$

In the equation, x represents the original meteorological dataset data, u represents the mean, and δ represents the standard deviation. The Laida rule indicates that if the value of x exceeds the interval of $(u - 3\delta, u + 3\delta)$, the data can be treated as abnormal data. In this paper, the mean value is used to fill in the outlier detected by the Laida rule.

(4) Normalized input meteorological element variables

If the values of these different meteorological elements are directly input into the network model, it will cause the model to crash. Therefore, in order to avoid the occurrence of this phenomenon, this paper needs to use a normalized calculation equation to scale the values of each data to the 0-1 range, and the formula is as follows:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (11)$$

In the formula, x represents the original meteorological dataset data, x' represents the cleaned sample data, and x_{\max} , x_{\min} represents the maximum and minimum values of each meteorological element in the original dataset. After unified normalization processing of meteorological data samples, it can be ensured that each data is within the 0-1 range.

(5) Dimension reduction processing

The massive amount of data increases the complexity and computational complexity of the model algorithm. Therefore, the Principal Component Analysis (PCA) algorithm is selected to extract the main feature components, retain only the main components, and recombine these new feature components into a new set of principal component variables, reducing the dimensionality of the model input data, removing useless noise data from the meteorological dataset, and achieving the effect of compressing the dataset size, improve model prediction accuracy and efficiency.

Assuming that the original meteorological datasetsample has m sample data, i.e. $(x_1, x_2, x_3, \dots, x_m)$, each sample contains n attribute variables, and the entire dataset is represented by $X_{m \times n}$.

The steps of the principal component analysis algorithm are as follows:

- (1) Standardize sample data processing. Centralize each column of meteorological element data in meteorological dataset X to obtain a standardized matrix A ;
- (2) Calculate the covariance matrix R :

$$R = \frac{1}{m}AA^T \quad (12)$$

- (3) For the covariance matrix R , solve the characteristic equation $|\lambda E - R| = 0$ to obtain the eigenvalues λ and w ;
- (4) Sort the eigenvalues of the covariance matrix $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_n$, select the k eigenvalues with the highest numerical value, and organize the corresponding eigenvectors $w_1, w_2, w_3, \dots, w_k$ for each eigenvalue;
- (5) Output projection matrix P , which maps the original meteorological dataset matrix to a new sample space.

$$Y = P^T X \quad (13)$$

C. ANALYSIS OF REGIONAL CHARACTERISTICS OF POWER OUTAGES CAUSED BY METEOROLOGICAL DISASTERS AMONG AFFECTED CUSTOMERS

Based on the meteorological data processed above, analyze the regional characteristics of power outage failures by calculating the power outage failure rates under different types of meteorological disasters.

Firstly, calculate the failure rate of power outages for affected customers based on the type of meteorological disaster:

$$\lambda_x = \frac{n_x}{T_x} \quad (14)$$

In the formula, n_x is the number of power outages caused by category x meteorological disasters, and T_x is the duration of category x meteorological disasters.

Due to varying climate characteristics in different regions, the impact of meteorological disasters differs across seasons. Therefore, it is necessary to conduct an associated assessment between the risk of power outage faults in different transmission lines and meteorological disasters. This aims to clarify the sensitivity of different transmission lines to various meteorological factors. Line meteorological sensitivity refers to the degree of sensitivity of a transmission line to a specific meteorological factor, expressed as the ratio of power outage faults caused by a specific meteorological factor to the total number of line faults.

$$\rho_x = \frac{n_{kx}}{n_x} \quad (15)$$

In the formula, n_{kx} represents the number of power outage failures of line k under meteorological condition x .

Obtain the meteorological disaster factors that cause the highest number of power outages and faults on the line:

$$MFW = \arg \max \rho(x) \quad (16)$$

In the equation, $\arg \max \rho(x)$ represents the set of all independent variables x that maximize the function ρ_x , that is,

the set of meteorological disaster factors that maximize the meteorological sensitivity of the region.

Calculate the difference between the number of power outages on a certain line under certain meteorological conditions and the average number of power outages of the same voltage level in the area:

$$E_{kx} = n_{kx} - \bar{n}_x \quad (17)$$

In the formula, \bar{n}_x represents the average number of power outage failures of a certain voltage level line in the region under meteorological condition x .

D. PREDICTION OF POWER OUTAGE AREAS BASED ON CNN-LSTM

CNN has excellent feature extraction capabilities that allow it to capture local features in images or data. This is crucial for feature extraction in power system data analysis and enables a more accurate identification of the factors leading to power outages. LSTM, on the other hand, is a type of recurrent neural network suitable for sequential data and capable of modeling temporal dependencies. In the context of power systems, time is a critical factor as various factors such as weather, load, and grid state change over time. By introducing an LSTM, it becomes possible to better model time-series data and predict outage regions. Therefore, a CNN-LSTM hybrid model is used in this study to improve the prediction of outage areas. By utilizing CNN for spatial feature extraction and LSTM for modeling temporal dependencies, it becomes possible to more accurately capture the factors that cause power outages and enhance the modeling capabilities for time series data. In addition, the parallel computing capability of CNN-LSTM facilitates faster training and inference processes, leading to improved efficiency in outage area prediction.

The CNN neural network uses local connections and shared weights to extract effective features from the original meteorological data through alternating stacking of convolutional and pooling layers. Finally, the extracted features are sorted and output by the fully connected layer. The specific process is as follows:

- (1) Convolutional layer

The convolutional layer [16], also known as the CNN meteorological data feature extraction layer, utilizes convolution operations to extract the original data features, which further highlights the original information features. For example, the convolutional layer slides a 3×3 convolution kernel horizontally over the input data to extract corresponding feature maps from the raw data information. The formula for this process is:

$$m = \sigma \left(\sum_{k=1}^3 \sum_{l=1}^3 w_{k,l} x_{k,l} + b \right) \quad (18)$$

where, m is the feature map of meteorological data, σ is the nonlinear activation function, w is the weight coefficient, k and l respectively represent the rows and columns that the convolution kernel slides on the original.

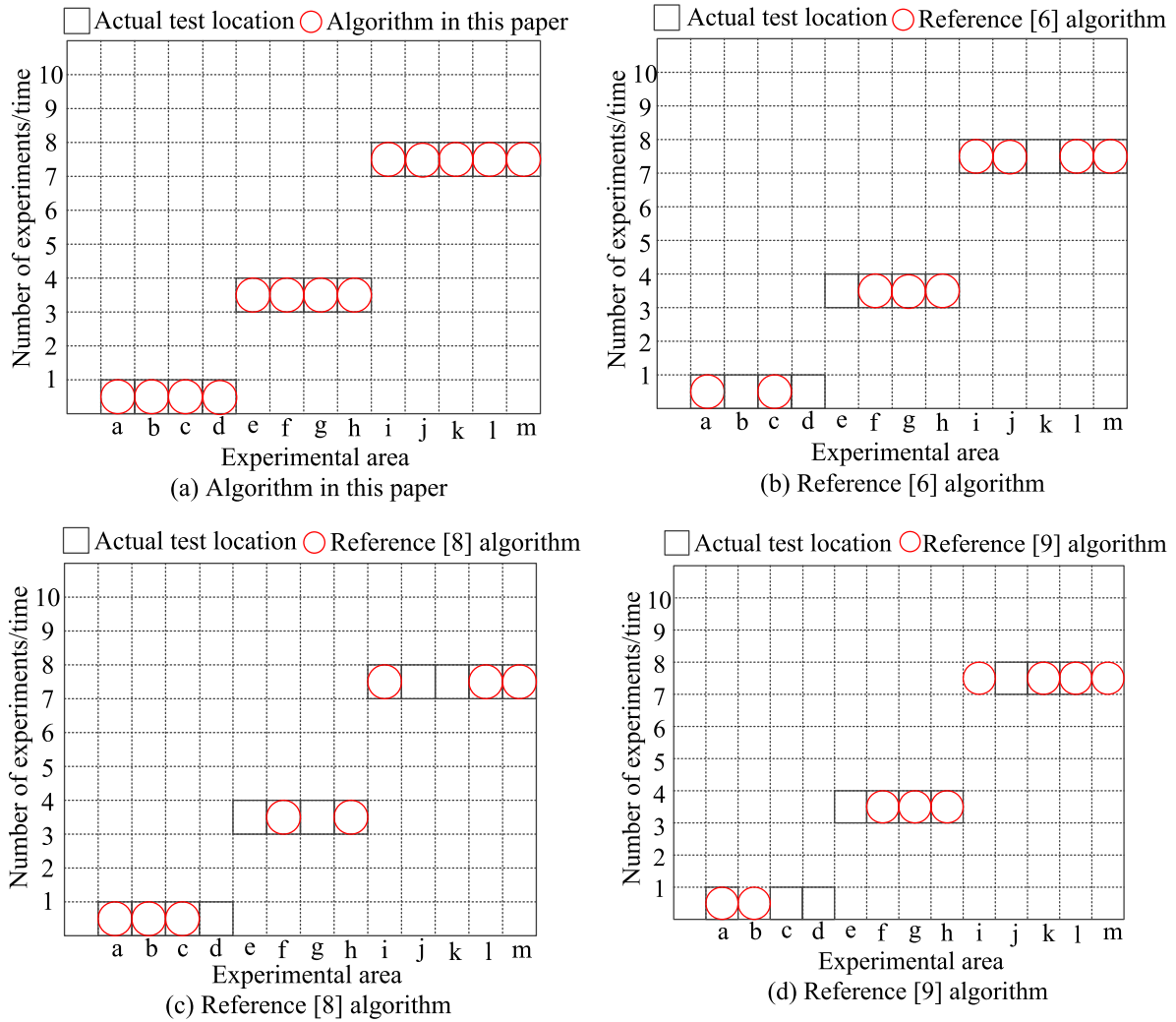


FIGURE 1. Prediction results of power outage areas for affected customers using four algorithms.

Meteorological data map each time, x is the data at the corresponding position, and x is the offset.

(2) Pooling layer

The pooling process, also known as the downsampling layer, is generally connected after the convolutional layer and is a process of secondary extraction of meteorological data features. Convolutional layers can use the combined action of multiple convolutional kernels to learn various features of meteorological data. The expression for maximum pooling is:

$$s_i = \max \{m_1, m_2, \dots, m_n\} \quad (19)$$

where, m is the value of different meteorological data characteristics in the pooling window area. Assuming that it is a n dimension column vector, the maximum pooling is to extract the maximum value in m .

After maximum pooling, the original $n \times n$ meteorological data features are compressed into s_i and the redundant features of the data are reduced, so the input data of the next layer of neural network becomes less, thus reducing the calculation

amount of data in the transmission process and reducing the risk of overfitting of the training network. In CNN networks [17], pooling layers are generally connected behind convolutional layers, and the two are alternately stacked. Construct a feature extraction layer.

(3) Fully connected layer

The fully connected layer generally appears behind several convolutional and pooling layers. Its function is to integrate the local meteorological data features extracted from convolutional and pooling layers into a complete and dense feature vector. The fully connected layer process is:

$$y = \sigma(ws + b) \quad (20)$$

where, σ is the nonlinear activation function, w is the weight coefficient, b is the offset, and y is the output data result of the full connection layer.

Based on the meteorological data feature extraction results of the CNN neural network, the LSTM network is used to predict the power outage area of affected customers.

The core components of LSTM are three gate unit structures, namely forgetting gate f_t , input gate i_t , and output gate o_t .

At time t , the LSTM memory unit has three inputs: the input value x_t of the current network, the output value h_{t-1} of the previous memory unit, and the unit state C_{t-1} of the previous time; There are two outputs of the LSTM memory unit: the output value h_t of the memory unit at the current time and the unit state C_t at the current time. The prediction process of power outage areas for affected customers in the LSTM network is as follows:

(1) Firstly, a prediction model for power outage areas of affected customers based on LSTM network is constructed:

$$f_t = \delta(W_f \cdot X + b_f) \quad (21)$$

In the formula, W_f is the weight matrix of LSTM, and b_f is the offset of LSTM.

(2) The input gate determines how much information is saved to the cell state at the current moment. The input gate outputs the output value of the sigmoid function between 0 and 1. If the input gate output is 0, it indicates that the corresponding information has not been updated. If it is 1, it indicates that the corresponding information needs to be updated. This paper uses the meteorological data feature extraction results of the CNN neural network as input sample x_t to input into the disaster stricken customer power outage area prediction model, generating a numerical value i_t between 0 and 1 to determine how much information the memory unit needs to retain. Meanwhile, a tanh layer will determine the candidate memory state C'_t through the output h_{t-1} of the previous state and the input x_t of the current state:

$$C'_t = \tanh(W_c [h_{t-1}, x_t] + b_c) \quad (22)$$

(3) Forgetting gate f_t is responsible for controlling whether to continue saving the long-term state C , which is jointly determined by the input value x_t at time t and the output value h_{t-1} of the previous hidden layer [18]:

$$f_t = \sigma(W_f [h_{t-1}, x_t] + b_f) \quad (23)$$

In the formula, W_f is the weight matrix of the forgetting gate f_t at time t , and b_f is the offset.

(4) The output gate determines what information needs to be output from the memory unit. Similar to the input gate, a sigmoid function is used to determine how much information C_t needs to be output from the memory unit. The expression is:

$$o_t = \sigma(W_o [h_{t-1}, x_t] + b_o) \quad (24)$$

The information C_t of the memory unit is multiplied by o_t and activated through a tank layer to obtain the output information of the current LSTM block, which is the predicted result of the affected customer's power outage area:

$$h_t = o_t \cdot \tanh(C_t) \quad (25)$$

III. SIMULATION EXPERIMENT ANALYSIS

In order to verify the effectiveness of the CNN-LSTM based algorithm for predicting the power outage area of affected customers in practical applications, a distribution line in a certain area was selected as the experimental object for experimental testing.

Before the experiment, collect the initial meteorological data using ground automatic weather observation stations. Specifically, collect temperature data of 8 GB, humidity data of 12 GB, wind speed data of 6 GB, and precipitation data of 10 GB. Set the parameters of the CNN-LSTM model as follows:

- Number of convolutional layers: 2
- Filter size: 3×3
- Pooling operation: Maximum pooling
- Number of LSTM layers: 1
- Number of units per LSTM layer: 128

Using the CNN-LSTM based algorithm for predicting the power outage area of affected customers, reference [6] algorithm, reference [8] algorithm, and reference [9] algorithm proposed in this paper, the power outage area of affected customers is predicted, and the predicted results are compared with actual test results. The comparison results are shown in Figure 1.

According to Figure 1, it can be seen that the CNN-LSTM based algorithm for predicting the power outage area of affected customers proposed in this paper is consistent with the actual test results. However, the results of the reference [6] algorithm, reference [8] algorithm and reference [9] algorithm for predicting the power outage area of affected customers are significantly different from the actual test results. The above results indicate that our method has undergone detailed processing in data preprocessing and feature construction, which helps to extract effective features from the data. However, the methods in references [6], [8], and [9] only consider a single influencing factor during prediction, which affects the accuracy of the prediction. This further demonstrates that the algorithm proposed in this paper has high accuracy in predicting the power outage areas of affected customers and good prediction results.

Using the CNN-LSTM based algorithm for predicting the power outage area of affected customers, reference [6] algorithm, reference [8] algorithm, and reference [9] algorithm proposed in this paper, a comparison was made of the time used for predicting the power outage area of affected customers. The comparison results are shown in Figure 2.

According to Figure 2, as the number of experiments increases, the prediction time of affected customer power outage areas for the three algorithms increases. When the number of experiments reaches 10, the proposed CNN-LSTM based algorithm for predicting the power outage area of affected customers takes 4.3 seconds, while the algorithms in reference [6] algorithm, reference [8] and [9] take 10 seconds, 11.1 seconds and 12.3 seconds, respectively. From this, it can be seen that the algorithm in this paper

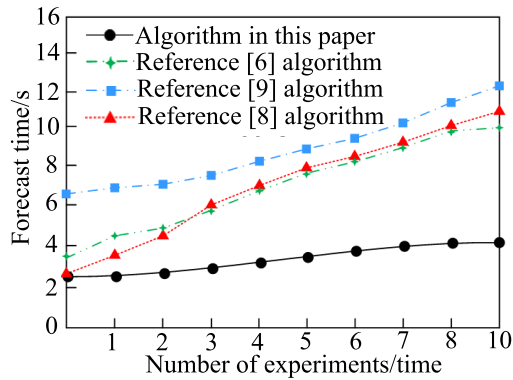


FIGURE 2. Comparison results of four algorithms for predicting the time of power outage areas for affected customers.

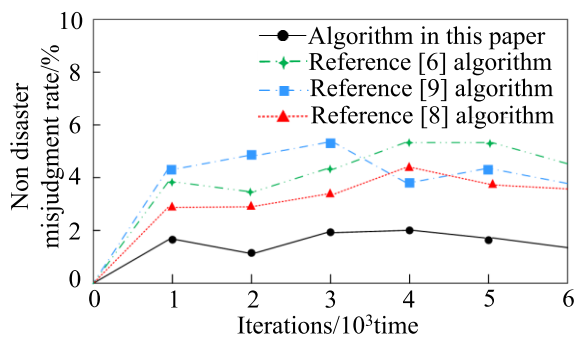


FIGURE 3. Non disaster misjudgment rates using different methods.

takes the shortest time to predict the power outage area of affected customers and has the highest prediction efficiency. It has been proven that compared with the algorithms in references [6], [8], and [9], our method combines the spatial feature extraction ability of CNN for time series data and the modeling ability of LSTM for dynamic changes in the time dimension, fully leveraging the advantages of both, resulting in better prediction efficiency.

In addition to focusing on accurate prediction of power outage areas for disaster affected customers, it is also possible to consider the prediction model mistakenly predicting non disaster customers as power outage areas for disaster affected customers. This indicator can be defined as the non disaster misjudgment rate. By measuring this indicator, the error rate of the model in identifying non disaster areas can be evaluated. The comparison results of the four methods are shown in Figure 3.

Analyzing Figure 3, it can be seen that in multiple iterations, the non disaster misjudgment rate of the CNN-LSTM based disaster affected customer power outage area prediction algorithm proposed in this paper is lower than that of the algorithms in references [6], [8], and [9]. The maximum non disaster misjudgment rate of the method in this paper is 2%. Because the CNN-LSTM based algorithm for predicting the power outage area of affected customers proposed in this paper preprocesses and processes the original data with features. This preprocessing and feature construction

process helps to extract effective information from the data and reduces the misjudgment rate of non disaster predictions.

IV. CONCLUSION

This paper is based on CNN-LSTM and conducts research on the prediction algorithm for power outage areas of affected customers. According to this study, the proposed method is able to accurately predict the outage areas of disaster-affected customers and achieves consistent results in comparison with actual test results. This algorithm can be used as a reliable forecasting tool to help relevant departments and institutions better understand the outage and take appropriate measures to ensure the power supply needs of affected customers. The prediction time of this algorithm is significantly reduced to 4.3 seconds. This means that outage prediction results can be obtained more quickly in emergency situations, enabling timely rescue and restoration measures to minimize the outage time for affected customers and improve disaster response efficiency. After applying the proposed method, the maximum non-disaster misjudgment rate is 2 percent. This means that the algorithm can better distinguish between disaster-hit and non-disaster-hit areas, reduce prediction errors in non-disaster areas, provide more accurate outage prediction results, and enable relevant departments to take more targeted response measures. In summary, the CNN-LSTM based algorithm for predicting outage areas for affected customers has achieved good results in terms of prediction accuracy, prediction time and misjudgment rate, demonstrating its important significance and potential in practical applications. This algorithm can provide reliable and efficient decision support for relevant departments and agencies to respond to disaster situations and ensure the power supply needs of affected customers.

REFERENCES

- [1] H. Hou, Z. Zhang, R. Wei, Y. Huang, Y. Liang, and X. Li, "Review of failure risk and outage prediction in power system under wind hazards," *Electric Power Syst. Res.*, vol. 210, Sep. 2022, Art. no. 108098, doi: 10.1016/j.epr.2022.108098.
- [2] H. Guo, S. Wang, and Y. Zhang, "Supply interruption supply chain network model with uncertain demand: An application of chance-constrained programming with fuzzy parameters," *Discrete Dyn. Nature Soc.*, vol. 2021, pp. 1–8, Feb. 2021, doi: 10.1155/2021/6686992.
- [3] A. A. Gibadullin, V. V. Sorokina, M. I. Sadriddinov, T. V. Petrushevich, T. B. Izzuka, and A. D. Kokurina, "Assessment of factors ensuring industrial safety of enterprises of the electric power complex," *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 862, no. 6, May 2020, Art. no. 062029, doi: 10.1088/1757-899x/862/6/062029.
- [4] N. Bhusal, M. Gautam, M. Abdelmalak, and M. Benidris, "Modeling of natural disasters and extreme events for power system resilience enhancement and evaluation methods," in *Proc. Int. Conf. Probabilistic Methods Appl. Power Syst. (PMAPS)*, Aug. 2020, pp. 1–6, doi: 10.1109/PMAPS47429.2020.9183679.
- [5] G. Joshi and S. Mohagheghi, "Optimal operation of combined energy and water systems for community resilience against natural disasters," *Energies*, vol. 14, no. 19, p. 6132, Sep. 2021, doi: 10.3390/en14196132.
- [6] H. Hou, X. Chen, M. Li, L. Zhu, Y. Huang, and S. Zhu, "A space prediction method for power outage in a typhoon disaster based on a Stacking integrated structure," *Power Syst. Prot. Control*, vol. 50, no. 3, pp. 76–84, Mar. 2022, doi: 10.19783/j.cnki.pspc.210580.

- [7] J. J. Zhao, K. Yan, D. S. Yang, and H. B. Wang, "Research on application of probabilistic model in blackout prediction," *Electr. Meas. Instrum.*, vol. 59, no. 4, pp. 164–168, May 2022, doi: [10.19753/j.issn1001-1390.2022.04.024](https://doi.org/10.19753/j.issn1001-1390.2022.04.024).
- [8] Y. Zhao, K. Wang, A. Zhao, Y. Zhou, and P. Yuan, "Power outage classification prediction method based on bagging ensemble learning," in *Proc. IEEE 4th Int. Electr. Energy Conf. (CIEEC)*, May 2021, pp. 1–5, doi: [10.1109/CIEEC50170.2021.9510882](https://doi.org/10.1109/CIEEC50170.2021.9510882).
- [9] C. Zhai, T. Y.-J. Chen, A. G. White, and S. D. Guikema, "Power outage prediction for natural hazards using synthetic power distribution systems," *Rel. Eng. Syst. Saf.*, vol. 208, Apr. 2021, Art. no. 107348, doi: [10.1016/j.res.2020.107348](https://doi.org/10.1016/j.res.2020.107348).
- [10] G. Giani, L. Tarasova, R. A. Woods, and M. A. Rico-Ramirez, "An objective time-series-analysis method for Rainfall-Runoff event identification," *Water Resour. Res.*, vol. 58, no. 2, Feb. 2022, Art. no. e2021WR031283, doi: [10.1029/2021wr031283](https://doi.org/10.1029/2021wr031283).
- [11] G. Colorado-Ruiz and T. Cavazos, "Trends of daily extreme and non-extreme rainfall indices and intercomparison with different gridded data sets over Mexico and the Southern United States," *Int. J. Climatol.*, vol. 41, no. 11, pp. 5406–5430, Sep. 2021, doi: [10.1002/joc.7225](https://doi.org/10.1002/joc.7225).
- [12] H. Hou, X. Chen, M. Li, L. Zhu, Y. Huang, and S. H. Zhu, "A Stacking integrated structure method for power outage space prediction under typhoon disaster," *Power Syst. Prot. Control*, vol. 50, no. 3, pp. 76–84, Mar. 2022, doi: [10.19783/j.cnki.pspc.210580](https://doi.org/10.19783/j.cnki.pspc.210580).
H. Hou, X. Chen, M. Li, L. Zhu, Y. Huang, and S. H. Zhu, "A space prediction method for power outage in a typhoon disaster based on a Stacking integrated structure," *Power Syst. Prot. Control*, vol. 50, no. 3, pp. 76–84, Mar. 2022, doi: [10.19783/j.cnki.pspc.210580](https://doi.org/10.19783/j.cnki.pspc.210580).
- [13] B. Han, Z. Ming, Y. Zhao, T. Wen, and M. Xie, "Comprehensive risk assessment of transmission lines affected by multi-meteorological disasters based on fuzzy analytic hierarchy process," *Int. J. Electr. Power Energy Syst.*, vol. 133, Dec. 2021, Art. no. 107190, doi: [10.1016/j.ijepes.2021.107190](https://doi.org/10.1016/j.ijepes.2021.107190).
- [14] X. Liu and G. Zhu, "Simulation of heat loss of ventilation windows in residential building's indoor thermal environment," *Comput. Simul.*, vol. 39, no. 7, pp. 441–445, Sep. 2022, doi: [10.3969/j.issn.1006-9348.2022.07.085](https://doi.org/10.3969/j.issn.1006-9348.2022.07.085).
- [15] B. Wang, F. Ma, L. Ge, H. Ma, H. Wang, and M. A. Mohamed, "Icing-EdgeNet: A pruning lightweight edge intelligent method of discriminative driving channel for ice thickness of transmission lines," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–12, 2021, doi: [10.1109/TIM.2020.3018831](https://doi.org/10.1109/TIM.2020.3018831).
- [16] J. Liang, T. Jing, H. Niu, and J. Wang, "Two-terminal fault location method of distribution network based on adaptive convolution neural network," *IEEE Access*, vol. 8, pp. 54035–54043, 2020, doi: [10.1109/ACCESS.2020.2980573](https://doi.org/10.1109/ACCESS.2020.2980573).
- [17] A. A. Zhuk and V. M. Bulochyik, "Neural network method of the decision of the nonlinear problem of optimum distribution of the non-uniform resource," *Syst. Anal. Appl. Inf. Sci.*, no. 1, pp. 45–52, Apr. 2021, doi: [10.21122/2309-4923-2021-1-45-52](https://doi.org/10.21122/2309-4923-2021-1-45-52).
- [18] M. Dashtdar and M. Dashtdar, "Fault location in radial distribution network based on fault current profile and the artificial neural network," *Sci. Bull. Electr. Eng. Fac.*, vol. 20, no. 1, pp. 14–21, Apr. 2020, doi: [10.2478/sbeef-2020-0103](https://doi.org/10.2478/sbeef-2020-0103).



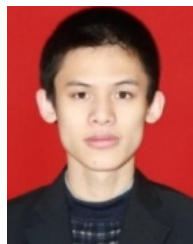
WEI ZHANG was born in Zongyang, Anhui, Han Nationality, in August 1983. He received the bachelor's degree from Guizhou University, in 2005, and the master's degree from North China Electric Power University, in 2014.

From 2005 to 2011, he was with Guangxi Pingban Hydropower Development Company Ltd. Since 2011, he has been with the Electric Power Research Institute, Guangxi Power Grid Company Ltd. He has published five academic articles. His research interests include power disaster prevention and reduction and monitoring of power equipment status.



QIANYI CHEN was born in Zhuang Ethnic Group, Hechi, Guangxi, in September 1991. He received the bachelor's and master's degrees from Guangxi University, in 2014, and 2017, respectively.

Since 2017, he has been with the Electric Power Research Institute, Guangxi Power Grid Company Ltd. He has published three academic articles. His research interests include distribution automation and intelligent operation, maintenance, and inspection of distribution networks.



BO FENG was born in Yao Ethnic Group, Guilin, Guangxi, in May 1986. He received the master's degree in electrical engineering from Chongqing University, in 2012.

Since 2012, he has been with the Electric Power Science Research Institute, Guangxi Power Grid Company Ltd.



WEIXIANG HUANG was born in Zhuang Ethnic Group, Nanning, Guangxi, in April 1993. He received the bachelor's and master's degrees from Shanghai Jiao Tong University, in 2015 and 2018, respectively.

Since 2018, he has been with the Electric Power Research Institute, Guangxi Power Grid Company Ltd. He has published five academic articles. His research interests include distribution network disaster prevention and mitigation and intelligent distribution networks.



XIN LI was born in Nanning, Guangxi, Han Nationality, in October 1987. He received the master's degree in signal and information processing from the Guilin University of Electronic Science and Technology, in 2012.

Since 2012, he has been the Network Security Specialist with the Electric Power Science Research Institute, Guangxi Power Grid Company Ltd.

...