Received 3 October 2022, accepted 17 October 2022, date of publication 21 October 2022, date of current version 27 October 2022.
Digital Object Identifier 10.1109/ACCESS.2022.3216309

RESEARCH ARTICLE

Research on Sand-Dust Storm Forecasting Based on Deep Neural Network With Stacking Ensemble Learning

QING-DAO-ER-JI REN [1], NA LI [2], AND WEIMING ZHANG [1]
[1]School of Information Engineering, Inner Mongolia University of Technology, Hohhot 010051, China
[2]School of Science, Inner Mongolia University of Technology, Hohhot 010051, China

Corresponding author: Na Li (linarq@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 62141603, in part by the Natural Science Foundation of Inner Mongolia under Grant 2020MS07006 and Grant 2022MS06013, and in part by the Research Start-up Foundation of Inner Mongolia University of Technology under Grant DC2200000934.

ABSTRACT Weather forecasting is an important factor affecting production and life. With the development of technology, weather forecasting methods such as weather map forecasting, numerical weather forecasting, and quantitative forecasting methods have emerged. However, these traditional data analysis methods have shortcomings such as incomplete analysis, insufficient objectivity, inability to quantitatively predict the weather, and low prediction accuracy. The import of neural networks into the field of weather forecasting helps to alleviate the above shortcomings and improve the accuracy of weather forecasting. In this paper, the LSTM network and the CNN network are used to predict the sand-dust storm weather. In order to improve the prediction performance, we use the Stacking integration algorithm to fuse the LSTM and CNN models. To improve the experimental scientific and comprehensive, using fully connected network and support vector machine as meta-classifiers, two LSTM-CNN integrated sand-dust storm prediction models are established. At last, the above integrated model is used in the prediction of sand-dust storms in Inner Mongolia. The experimental results show that compared with a single LSTM or CNN model, the Stacking ensemble model has different degrees of improvement in model evaluation indicators such as accuracy, precision, recall, and f1-score. The Stacking ensemble model uses the fully connected network model as the meta-classifier is even better. These prove that the Stacking ensemble algorithm improves the sand-dust storm classification effect and generalization ability of a single neural network to a certain extent.

INDEX TERMS Sand-dust storm, deep neural network, ensemble learning, convolutional neural network, long short-term memory.

I. INTRODUCTION

Since modern times, the number of sand-dust storms in northern of China has increased significantly, and the impact on people's production and life has become greater. Further reveal the characteristics of sand-dust storms, and In-depth study of the relationship between sand-dust storms, arid climates and the related ecological problems such as desertification has scientific and practical significance [1].

Machine learning is one of the core technologies of artificial intelligence. In the previous sand-dust storm research, machine learning has achieved remarkable results. In 2015, Zhang et al. established a sand-dust storm prediction model based on SMOTE algorithm and Decision Tree algorithm. The research results show that the model can better solve the problem of classification and prediction of unbalanced samples, can be used for actual sand-dust storm warning [2]. In 2020, Gholami et al. used eight machine learning algorithms, including Random Forest, Support Vector Machine, BART, Radial Basis Function, XGBoost, RTA, BRT, and

The associate editor coordinating the review of this manuscript and approving it for publication was Yiqi Liu.

EM algorithms to predict the source of sand-dust storm in Khuzestan Province in southwestern Iran. The data shows that the EM algorithm has the highest prediction accuracy, and its AUC index reaches 99.8% [3]. In the same year, Shi et al. proposed a new method based on support vector machines to automatically detect sandstorms using remote sensing data. The experimental results show that the supervised classification method based on SVM has good performance in SDS detection [4]. In 2022, Wang Wei et al. proposed a new method for mixed detection of sandstorms based on MODIS data of GEE platform to help automatically label training samples, aiming at the low efficiency of manually annotated samples. This method can effectively reduce the false positive rate, and the accuracy rate is more than 98% in the sandstorm detection task [5].

Deep learning is a deep machine learning model. The commonly used deep learning model is a multi-layer neural network. Each layer of the neural network will input non-linear mapping. Through the stacking of multi-layer non-linear mapping, very abstract features can be calculated to help classification [6]. Recurrent Neural Network (RNN) is a type of deep learning model commonly used to process sequence data, which is widely used in speech processing, natural language processing, financial data prediction and other fields [7], [8]. In 2018, Tokgöz et al. used RNN-based variant networks, LSTM and GRU to conduct Turkey electric load time forecasting experiments, and explored the application research of RNN in the electric load field. The experimental results show that compared with the existing power load forecasting methods based on ARIMA and artificial neural networks, the forecasting success rate of this method is increased by 2.6% and 1.8%, respectively [9]. In 2019, Huang et al. proposed a neural network prediction model based on LSTM for the complexity and long-term dependence of financial time series prediction. The model uses the stacked denoising self-encoding mechanism to extract features from the basic market data and technical indicators of financial time series. The experimental results show that compared with traditional neural networks, the prediction model based on LSTM neural network has higher prediction accuracy [10]. In 2022, Feng et al. applied the deep convolutional neural network model to the solar energy forecasting research of long-term time series. Experimental results show that CNNS also have consistently superior performance compared to shallow machine learning models with weather forecasters, with an average improvement rate of about 7% [11].

As mentioned above, deep learning has been well applied in various fields, but there are also some shortcomings, such as the generalization of results and the accuracy of predictions. The main idea of Ensemble Learning is to use multiple models to solve the same problem, which can better improve the generalization ability of the model. It is one of the key research directions in the field of machine learning and deep learning [12], [13]. In 2018, Lu et al. built a stacking learning framework based on five base classifiers of naive Bayes, logistic regression, nearest neighbor, decision tree and rule learning for the classification ensemble problem, and compared with the methods such as AdaBoost, Bagging, Random Forest, Voting and Cross-Validation. The Experimental results show that the Stacking algorithm has the strongest generalization ability and is more suitable for situations with a large number of samples [14]. In 2020, Xiao et al. combined deep learning and ensemble learning to propose an automatic visual classification algorithm. The algorithm adds the Swish activation function to the LSTM network and applies the Bagging algorithm to improve the generalization ability of the model [15]. In 2022, He et al. proposed a combined model for short-term wind power forecasting based on numerical weather prediction analysis. In this model, CNN and LSTM networks were used to predict wind power under different weather conditions, and IOWA operator was used to combine the prediction results of the two models. Experimental results show that compared with Radial Basis Function (RBF), Extreme Learning Machine (ELM) and Support Vector Machine (SVM) methods, the proposed method can effectively improve the accuracy of wind power prediction under different weather [16]. At present, with the in-depth research of ensemble learning, its broad definition is gradually accepted by scholars. It refers to the way of learning multiple sets of learners without distinguishing the nature of learners [17].

In summary, in view of the shortcomings of current sandstorm prediction methods, such as incomplete analysis, insufficient objectivity, inability to quantitatively predict weather conditions, and low prediction accuracy, this paper proposes a sandstorm prediction method based on deep neural network with superimposed ensemble learning. The main contributions of this paper are as follows: The data augmentation technique is studied, and the upsampling (SMOTE algorithm) method is used to perform data balancing on sandstorm data. In the model training, the LSTM network and CNN network are used to predict the sandstorm weather, so as to fully mine the time series features and local attribute related features of sandstorm data. Then, the stacked ensemble algorithm is used to fuse the LSTM model and CNN model to improve the prediction performance of the model.

II. SAND-DUST STORM PREDICTION MODELS BASED ON DEEP NEURAL NETWORK

A. LSTM SAND-DUST STORM PREDICTION MODEL

We first established a sand-dust storm prediction model based on LSTM. In the model, a many-to-many expansion form is adopted inside the LSTM unit, that is, after the time series data flows to the hidden layer, each time step will output an updated state; Overall, it is a many-to-one expansion form, that is, After the time series data flow to the hidden layer, the output is the state of the last time step. The stacking of these two unfolding forms can better learn the timing characteristics, and is suitable for sand-dust storm forecasting. The specific structure of the LSTM sand-dust storm prediction model is shown in Fig.1.

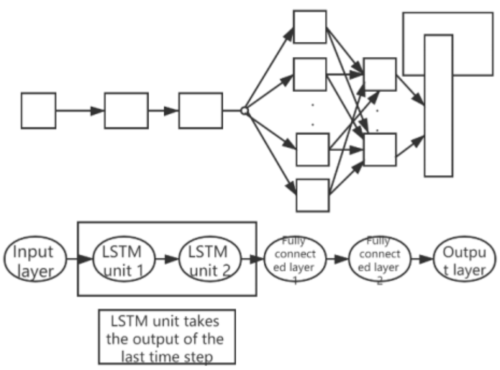

FIGURE 1. LSTM model structure diagram.

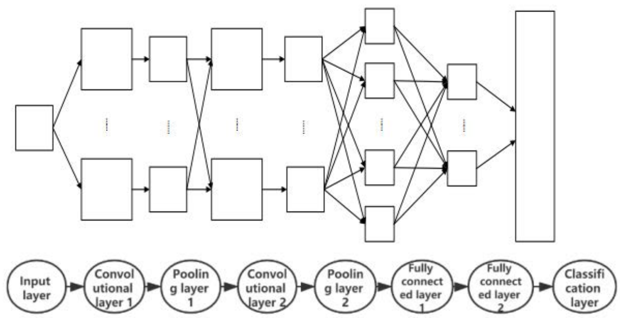

FIGURE 2. CNN model structure diagram.

As shown in Fig.1, this model connects two fully connected layers after the LSTM unit, and the final output layer of the model uses the Softmax function.

B. CNN SAND-DUST STORM PREDICTION MODEL

The CNN-based sand-dust storm prediction model designed in this paper is as follows. The model uses a two-layer convolutional CNN model, and the specific network structure is shown in Fig. 2.

As shown in Fig. 2, the model connects the pooling layer after each convolutional layer, and then passes through two fully connected layers, and the final output layer uses the Softmax function. In the convolution module and fully connected unit of the model, after the pooling operation or the fully connected operation, the Batch-Normalization layer is added before the ReLU activation function to obtain a better generalization effect.

III. STACKING ENSEMBLE LEARNING SAND-DUST STORM PREDICTION MODEL

Ensemble Learning technology uses multiple versions of the base learner to solve the same problem, which can significantly improve the generalization ability of the learning system [18].

Stacking ensemble algorithm refers to training a meta-learner to combine other base learners. The main idea is to train multiple different models (base learners), and then use the output of these models as input to train a new model (meta learner) to get a final output. The specific process is as follows.

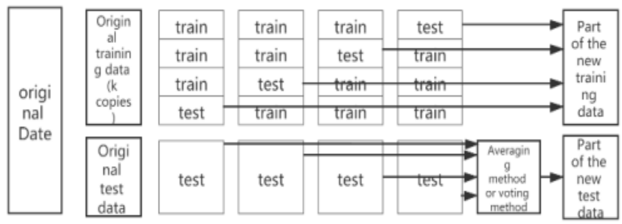

FIGURE 3. Flow of Stacking ensemble policy.

1) Divide the original training set and the original test set into k parts, each time using k-1 original training set as the base learner training set, and 1 original training set as the base learner verification set.Use them to train the base learner model 1.
2) Train a total of k times, and use a copy of the original test set to test the model at the end of each training.
3) The output results of the model validation set will be used as part of the training set of the meta-learner model, and after being averaged or voted the output results of the test set will become part of the test set of the meta-learner model.
4) For base learner model 2, base learner model 3…, repeat the above steps to obtain the training set and test set of the entire meta-learner model.
5) According to the obtained training set and test set, train and predict the meta-learner model.

The first three steps of the above specific process are shown in Fig. 3. We use a 4-fold base classifier, that is, k is equal to 4.

In addition to the base classifier, the Stacking ensemble algorithm also needs a meta-classifier to fuse the base classifiers. A fully connected network model and a support vector machine (SVM) model are used as meta-classifiers to establish two Stacking ensemble models.

In this paper, the test set of the meta-classifier is obtained by averaging, and if there is a decimal, it is rounded up. That is, in the process of constructing the input attributes of the meta-classifier test set, if the priority of two adjacent sandstorm levels is the same, the level with the lighter sandstorm severity is selected, which is more in line with the level distribution of the original sample.

A. STACKING ENSEMBLE MODEL BASED ON FULLY CONNECTED NETWORK META-CLASSIFIER

Fully Connected Network (FCN) is a simple feedforward neural network in which neurons are arranged in layers. Each neuron is only connected to the neuron of the previous layer, receives the output of the previous layer, and outputs it to the next layer. There is a unidirectional connection structure between all nodes of two adjacent layers. It defines the mapping $Output = f(I; \theta)$ from the input to the output. During

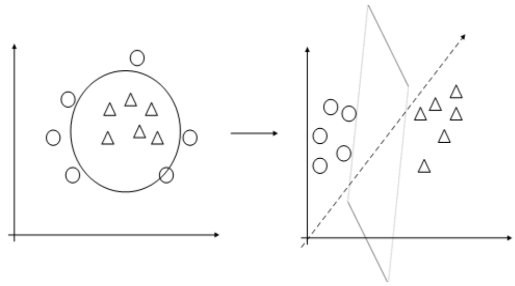

FIGURE 4. Nonlinear mapping.

the network training process, the value of the parameter vector θ will be adjusted to make f approach a certain desired function f* [19]. In theory, as long as the hidden layer of FCN has a sufficient number of neurons, it can fit functions of any complexity [20], [21], [22].

The meta-classifier fully connected network model adopts a three-layer fully connected layer stack structure, and the output layer uses the Softmax function to obtain multi-class prediction results.

The Stacking ensemble model based on the fully connected network meta-classifier uses the LSTM sand-dust storm prediction model and the CNN sand-dust storm prediction model as the two base classifiers of the stacking ensemble strategy, and the fully connected network model is used as the meta-classifier to establish the Stacking ensemble model.

B. STACKING ENSEMBLE MODEL BASED ON SVM META-CLASSIFIER

Support Vector Machine (SVM) is a classification algorithm, which belongs to the category of supervised learning. Because it introduces the concept of the maximum target hyperplane, it has good classification performance [23]. SVM is essentially a two-classification algorithm, but it can be extended to a multi-classification algorithm by modifying the objective function or combining multiple two-classifiers, which is suitable for the prediction target of this topic.

For non-linear problems such as sandstorm level prediction, linear separable support vector machines cannot be effectively solved, and a non-linear model can be used for better classification. Therefore, this paper adopts non-linear support vector machine as the meta-classifier of Stacking ensemble strategy.

As shown in Fig.4, non-linear mapping refers to mapping the training samples from the original space to a higher-dimensional space so that the samples are linearly separable in this space. If the original space has finite dimensions, the attributes are limited. Then there must be a high-dimensional feature space to make the sample separable.

Let $\phi(x)$ denote the feature vector after mapping x, W^T and b denote the corresponding weights and bias values, respectively, so in the feature space, the model corresponding to the divided hyperplane can be expressed as Eq. 1.

$$y_i\left(W^T\phi(x)+b\right)\geq 1 \tag{1}$$

where (x_i, y_i) is the training sample. Then the target formula becomes Eq. 2, where m is the number of training samples, ξ_i is the slack variable introduced by the sample points, which takes a value greater than zero.

$$\max \frac{2}{\|W\|}+C\sum_{i=1}^{m}\xi_i \\ s.t.\ y_i\left(W^T\phi(x)+b\right)\geq 1 \tag{2}$$

The above describes that SVM can map the input space to the high-dimensional feature space through a certain nonlinear transformation $\phi(x)$. Since the dimensionality of the feature space may be very high, there is a problem of difficulty in calculation.

If the solution of the support vector machine only uses the inner product operation, and there is a certain function $K\left(x, x'\right)$ in the low-dimensional input space, it is equal to the inner product in the high-dimensional space, that is, $K\left(x, x'\right) = \left\langle\varphi(x), \varphi'(x)\right\rangle$.Then the support vector machine does not need to calculate the complex nonlinear transformation, and the inner product of the nonlinear transformation is directly obtained from this function $K\left(x, x'\right)$, which greatly simplifies the calculation. Such a function $K\left(x, x'\right)$ is called a kernel function.

The Stacking ensemble model based on SVM meta-classifier uses LSTM sand-dust storm prediction model and CNN as the two base classifiers of the stacking ensemble strategy, and uses the SVM model as the meta-classifier to establish the Stacking ensemble mode.

As shown in Figure 5, the sandstorm dataset is first input into the model. During the training of the neural network, the model will expand the read sample data in a grid, that is, the time is arranged from top to bottom, and the meteorological attributes are arranged from left to right. In the experiment of this paper, we set the time series length to 15 and select 15 key meteorological attributes, so the meteorological data after gridding is 15 × 15 matrix data. Secondly, the LSTM model and the CNN model are used for training. Here, we make full use of the time series memory capability of the LTSM model to focus on extracting the sequence features of the sandstorm data; we use the local feature extraction capability of the CNN model to focus on extracting key attribute features of the sandstorm data. Finally, the Stacking algorithm is used to further integrate the training results of the LSTM model and the CNN model to improve the sandstorm prediction performance of the entire model.

IV. EXPERIMENTAL DESIGN

A. EXPERIMENTAL DATA

In this paper, the sand–dust storm data of Inner Mongolia in the past 54 years is used to verify the validity of the proposed model. We uses two data sets, namely "China's Severe Dust Storm Sequence and Its Supporting Data Set" and "China's Surface Climate Data Daily Value Data Set".

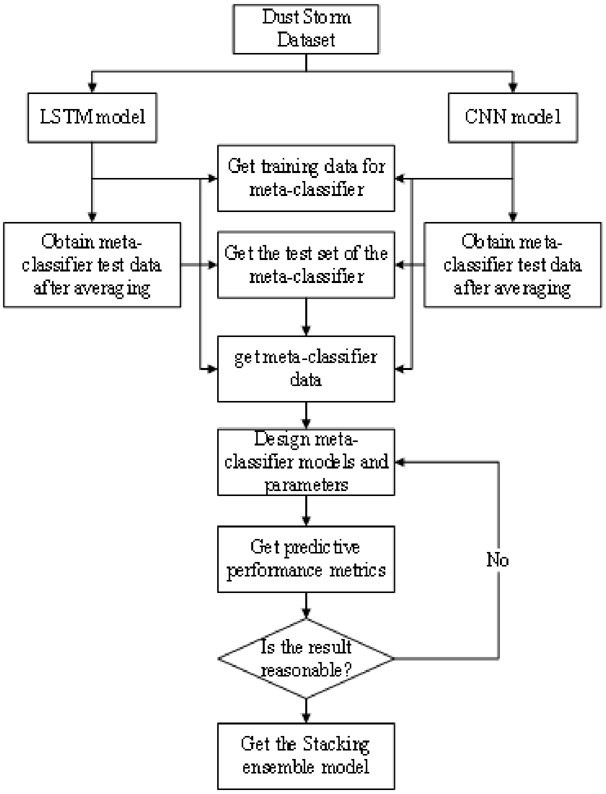

FIGURE 5. Flow chart of the experiment.

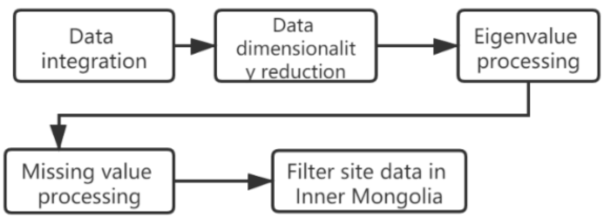

FIGURE 6. Flow chart of data preprocessing.

1) DATA PREPROCESSING

The preprocessing process is shown in Fig.6.

2) LABEL ATTRIBUTES AND INPUT ATTRIBUTES

According to the national standards, the grades of sand and dust storms of the label attributes in this article are divided into: super strong sand storm, strong sand storm, sand storm, blowing sand, floating dust, and no sand storm. Due to the fact that there are too few occurrences of extremely strong sand-dust storms, there are only 76 sample data in Inner Mongolia since 54 years, so it is combined with strong sand and dust storms to form a grade, so this paper finally uses five sand-dust storm grade labels, namely 0, 1, 2, 3, 4, the intensity decreases with this.

The input attribute of this paper is the weather factor after preprocessing, and the time span is 15 days, that is, the weather factor of the previous 15 days predicts the sand and dust storm level on the 16th day. Among them, the key attributes of the model input sequence are shown in Table 1.

TABLE 1. Key properties of the model input sequence.

The property name	Unit	The property name	Unit
Small evaporation	0.1 mm	Mean station pressure	0.1 hPa
Mean surface air temperature	0.1℃	Average relative humidity	0.1 hPa
Daily maximum surface temperature	0.1℃	Sunshine hours	0.1 h
Daily minimum Surface air temperature	0.1℃	Average temperature	0.1℃
Precipitation at 20-8	0.1 mm	Daily maximum temperature	0.1℃
Precipitation from 8 to 20 hours	0.1 mm	Daily minimum temperature	0.1℃
24-hour cumulative precipitation	0.1 mm	Average wind speed	0.1 m/s

TABLE 2. The number of samples after dividing train set and test set.

Sandstorm level	Number of training sets	Proportion	Number of test sets	Proportion	total
class #0	1050	0.803	257	0.197	1307
class #1	1997	0.795	516	0.205	2513
class #2	3048	0.790	808	0.210	3856
class #3	2747	0.797	700	0.203	3447
class #4	80000	0.800	20000	0.200	100000

3) DATA PARTITION

The original sand-dust storm data set in Inner Mongolia obtained by preprocessing was used as the original sample, and the samples that did not emit sand -dust storms were down-sampled, and the samples were sampled to 100,000 according to the time interval. A total of 110853 pieces of sample data were obtained. The ratio of training set to test set is about 8: 2, and the data distribution is shown in Table 2.

4) DATA BALANCE

There is a serious imbalance in the number of sand - dust storm samples between each level, which may lead to over-fitting of the prediction model. In order to alleviate the occurrence of such problems, The SMOTE (Synthetic Minority Over Sampling Technique) is used to balance the training data, and the test data remains unchanged. Then, the sand-dust storm samples with grades 0, 1, 2, and 3 in the training set are respectively enhanced to 100,000. The basic formula of the SMOTE algorithm is Eq. 3, where x represents a minority sample, and x_n represents a random neighbor of x.

$$x_{new} = x + rand\,(0,1) \times (x_n - x) \tag{3}$$

B. BASE CLASSIFIER MODEL ESTABLISHMENT

1) LSTM SAND-DUST STORM PREDICTION MODEL PARAMETERS

The choice of parameters has an important impact on the prediction performance of the neural network. The main

TABLE 3. Main parameters of LSTM model structure.

Network structure	Meaning description	Input specifications	Output specifications
LSTM_CELL_1	LSTM unit 1	15 × 15 × 1	15 × 1 × 192
LSTM_CELL_2	LSTM unit 2	15 × 1 × 192	1 × 1 × 192
Fc1	Fully connected layer 1	1 × 1 × 192	1 × 1 × 64
Fc2	Fully connected layer 2	1 × 1 × 64	1 × 1 × 32
Cls	Classification layer	1 × 1 × 32	5

TABLE 4. Main parameters of CNN model structure.

Network structure	Meaning description	Input specifications	Output specifications
Conv_strides	Convolution step size = 1	—	—
Pool_padding	Convolution filling method = 'AME'	—	—
Pool_strides	Pooling step =2	—	—
Pool_padding	Pooled filling method = 'VALID'	—	—
Conv_1	Convolutional layer 1, 64 5×5 convolution kernels	15 × 15 × 1	15 × 15 × 64
Pool_1	Pooling layer 1, maximum pooling	15 × 15 × 64	7 × 7 × 64
Conv_2	Convolutional layer 2, 128 3×3 convolution kernels	7 × 7 × 64	7 × 7 × 128
Pool_2	Pooling layer 2, maximum pooling	7 × 7 × 128	3 × 3 × 128
Fc_1	Fully connected layer 1, specification = 64	3 × 3 × 128	1 × 1 × 64
Fc_2	Fully connected layer 2, specification = 32	1 × 1 × 64	1 × 1 × 32
Cls	Classification layer, specification = 5	1 × 1 × 32	5

structural parameters of this LSTM sand-dust storm prediction model are shown in Table 3. The specifications of LSTM unit 1 and unit 2 are 192, the specification of fully connected layer 1 is 64, the specification of fully connected layer 2 is 32, and the final output layer specifications are equal to the number of sandstorm types is the same, 5.

2) CNN SAND-DUST STORM PREDICTION MODEL PARAMETERS

The main structural parameters of this CNN sandstorm prediction model are shown in Table 4. The convolution step size is unified to 1, the convolution filling method is 'SAME', the pooling step size is unified to 2, and the pooling filling method is unified to 'VALID'. The number of convolution kernels in convolution layer 1 is 64, the size of convolution kernel is 5 × 5, the number of convolution kernels in convolution layer 2 is 128, the size of convolution kernel is 3 × 3, and the pooling method is unified as maximum pooling.The specification of fully connected layer 1 is 64, the specification of fully connected layer 2 is 32, and the final output layer specification is the same as the number of sandstorm categories, which is 5.

TABLE 5. Main parameters of fully connected network model structure.

Network structure	Meaning description	Input specifications	Output specifications
Fc_1	Fully connected layer 1 Specification = 64	1 × 2 × 1	1 × 1 × 64
Fc_2	Fully connected layer 2 Specification = 128	1 × 1 × 64	1 × 1 × 128
Fc_3	Fully connected layer 3 Specification = 64	1 × 1 × 128	1 × 1 × 64
Cls	Classification layer Specification = 5	1 × 1 × 64	5

TABLE 6. Main parameters of fully connected network model.

Main parameters of the fully connected model	Parameter meaning	Parameter setting value
Weight_stddev	Weight standard deviation	0.01
Regularizer_scale	Regular term coefficient	0.01
epochs	Number of training rounds	10
batch_size	Batch size	200
train_iter_steps	Training steps per round	2000
display_step	Show interval steps	200
count_step	Statistics interval steps	200
learning_rate	Learning rate	0.01
learning_rate_decay	Learning rate decay coefficient	0.90
learning_rate_step	Learning rate decay step	100/batch_size × 400000

C. META-CLASSIFIER (FULLY CONNECTED NETWORK) PREDICTION RESULTS

1) MODEL PARAMETERS

The main structural parameters of this fully connected neural network model are shown in Table 5. The specification of fully connected layer 1 is 64, the specification of fully connected layer 2 is 128, and the specification of fully connected layer 1 is 64. The final output layer specification is the same as the number of sandstorm categories,5.

2) MODEL TRAINING

The training parameters of the meta-classifier (fully connected network) are shown in Table 6, and the Loss curve during model training is shown in Fig. 7.

3) MODEL PREDICTION RESULTS AND ANALYSIS

Stacking ensemble model (fully connected network) sand-dust storm prediction and evaluation indicators are shown in Table 7.

It can be seen from Table 7 that in the prediction results of the test set, the recall rates of grades 0, 3, and 4 are higher,

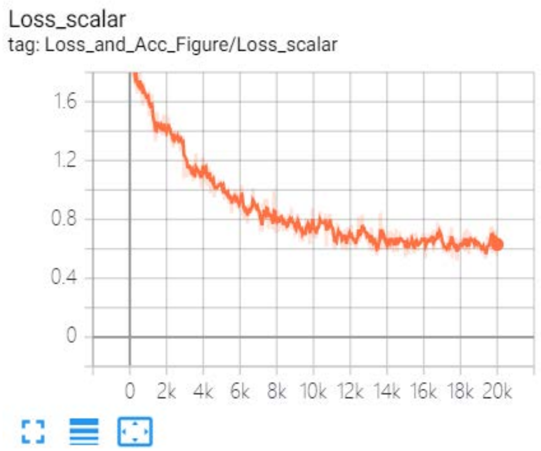

FIGURE 7. Loss curve of fully connected network model.

TABLE 7. Forecast evaluation indicators of Stacking ensemble model(FC).

Sandstorm level	precision	recall	f1-score
class #0	0.9807	0.9883	0.9845
class #1	0.6314	0.6473	0.6392
class #2	0.4378	0.6448	0.5215
class #3	0.1850	0.7529	0.2970
class #4	0.9830	0.8579	0.9162
Weighted average	0.9300	0.8435	0.8768

TABLE 8. Main parameters of SVM model.

SVM main parameters	Parameter meaning	Parameter setting value
cache_size	Batch size	200
class_weight	Classification weight strategy	‘balanced’
C	Soft spacing factor	1
decision_function_shape	Multi-class fusion strategy	‘ovr’
kernel	Kernel function	‘Rbf’, Gaussian Radial Basis Nucleus
tol	Residual convergence condition	0.001

reaching 0.98, 0.75, and 0.85 respectively, and the recall rates of other lower grades are relatively low; Due to the large difference in the number of samples in the original test set, the accuracy of the sandstorm level samples is low, but their weights are averaged to 0.93. Combined with the f1-score indicator, it can be concluded that the overall prediction performance of the Stacking ensemble model (fully connected network) on the original test set is significantly improved compared to the single neural network model.

D. META-CLASSIFIER (SVM) PREDICTION RESULTS

1) MODEL PARAMETERS

The main parameters of the SVM model are shown in Table 8.

Where cache_size is set to 200, its function is to limit the amount of calculation and prevent the decrease in calculation performance due to excessive data volume; class_weight is the classification weight strategy, set to ’balanced’, it means that the weights are automatically assigned according to the proportion of the sample size in each class; C is the slack variable penalty coefficient, set to 1; decision_function_shape is a multi-class fusion strategy, set to’ovr’, that is, one category is divided from other categories; kernel is the kernel function type, set to’rbf’, that is, Gaussian radial basis core; tol is the residual convergence condition, set to 0.001, that is, tolerate an error in 1000 classifications, and stop training when the error term reaches the specified value.

TABLE 9. SVM model support vector number.

Sandstorm level	Number of support vectors
Class #0	3989
Class #1	28761
Class #2	36953
Class #3	17400
Class #4	14827
total	101930

TABLE 10. Forecast evaluation indicators of Stacking ensemble model(SVM).

Sandstorm level	precision	recall	f1-score
class #0	0.9807	0.9883	0.9845
class #1	0.5400	0.6143	0.5748
class #2	0.4178	0.5854	0.4876
class #3	0.1850	0.7529	0.2970
class #4	0.9830	0.8579	0.9162
Weighted average	0.9271	0.8405	0.8741

2) MODEL TRAINING

The number of support vectors of each type of sample obtained by the meta-classifier (SVM) training is shown in Table 9.

3) MODEL PREDICTION RESULTS AND ANALYSIS

Stacking ensemble model (SVM) sand and dust storm prediction and evaluation indicators are shown in Table 10.

It can be seen from Table 10 that in the prediction results of the sand and dust storm test set, the recall rates of grades 0 and 4 are higher, reaching about 0.98 and 0.85, respectively, and the recall rates of other lower grades are relatively low; Due to the large difference in the number of samples in the original test set, the accuracy of the sandstorm grade samples is low, but their weights averaged around 0.92. Combining the f1-score indicator, it can be concluded that the overall prediction performance of the Stacking ensemble Model (SVM) on the test set has been improved compared to a single neural network model.

E. COMPARISON OF PREDICTION PERFORMANCE OF VARIOUS MODELS

As mentioned above, this paper establishes the Stacking ensemble model with the fully connected network model and the SVM model as meta-classifiers, and uses the test

TABLE 11. Comparison of forecast performance on original test set of various models.

Model	overall accuracy	accuracy of sandstorm occurrence	accuracy rate of no sandstorm	precision	recall	f1-score
LSTM	0.8296	0.677	0.85	0.91	0.83	0.86
CNN	0.8333	0.670	0.85	0.91	0.83	0.87
Stacking-FC	0.8435	0.717	0.86	0.93	0.84	0.88
Stacking-SVM	0.8405	0.689	0.86	0.93	0.84	0.87

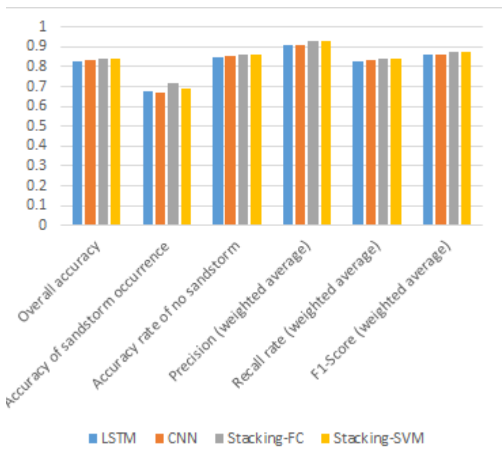

FIGURE 8. Comparison diagram of Forecast performance of various models.

set to evaluate its performance. Table 11 shows the comparison of sand-dust storm prediction performance of each model. Among them, Stacking-FC represents the Stacking integration model with the fully connected network model as the meta-learner, and Stacking-SVM represents the Stacking integration model with the SVM model as the meta-learner, the same below. Visualize Table 11, as shown in Fig. 8.

It can be seen from Table 11 and Fig.8 that for the original test set, the various prediction performance indicators of the two types of Stacking ensemble models are higher than the single RNN and CNN models, and the accuracy of Stacking-FC in predicting the occurrence of sandstorms is significantly higher. In Stacking-SVM, the accuracy of predicting no sandstorms is the same as Stacking-SVM, and other indicators are slightly higher than Stacking-SVM.

Among them, compared to the LSTM model, the overall accuracy of the Stacking-FC model has increased by about 1.4%, the accuracy of sandstorms has increased by about 3.8%, and the accuracy of no sandstorms has increased by about 1.1%. Its weight accuracy, recall rate, F1-Score increased by 1.8%, 1.4%, and 1.7% respectively.

V. CONCLUSION

This paper takes the sand-dust storm data in Inner Mongolia as the research object, uses the SMOTE algorithm to enhance the sand and dust storm data training set, uses the LSTM model and the training set to establish the sand and dust storm weather prediction model, and uses the test set to evaluate its performance. The experimental results show that LSTM The overall accuracy rate of the sandstorm prediction model reached about 0.8295, and the weighted average accuracy, recall rate and F1-Score reached 0.9118, 0.8296, and 0.8592 respectively. In order to improve the prediction performance of the above-mentioned LSTM model, this paper adopts the Stacking integration strategy to improve the performance of the model. The LSTM model and the CNN model based on time series data are used as the base classifiers, and the fully connected network model and the SVM model are respectively used as the meta-classifiers. Two ensemble models, and a comparative analysis of the experimental results of each model. Experiments show that the prediction performance of the two ensemble models has been improved to a certain extent compared to the single RNN and CNN, and the Stacking ensemble model with the fully connected network model as the meta-classifier has a slightly better prediction performance than the SVM model The Stacking ensemble model of the meta classifier has an overall accuracy rate of about 0.8434, and the weight average accuracy, recall rate and F1-Score have reached 0.9300, 0.8435, 0.8768, respectively.

Although the use of data augmentation methods and ensemble learning methods can improve the sandstorm prediction performance of the model to a certain extent, the prediction accuracy of the model still needs to be further improved due to limitations such as the limited number of publicly available datasets. In the future research on sand and dust storm prediction, we will optimize the model from the following aspects: in terms of data, collect more data sets, such as introducing satellite cloud image data, to obtain richer sand and dust storm data information and improve the effect of the model; in terms of methods, Use the deep neural network structure to mine more valuable data information.

REFERENCES

[1] Y.-H. Li and S. Y. Zhang, "Review of the research on the relationship between sand–dust storm and arid in China," *Advance Earth Sci.*, vol. 11, pp. 1169–1177, Aug. 2007.

[2] Z. H. Zhang, J. H. Xu, and L. X. Li, "The short-term forecasting model of sandstorms based on SMOTE and decision tree learning algorithm," *J. Xuzhou Inst. Technol. Natural Sci. Ed.*, vol. 30, no. 38, pp. 40–46, Aug. 2015.

[3] H. Gholami, A. Mohamadifar, and A. L. Collins, "Spatial mapping of the provenance of storm dust: Application of data mining and ensemble modelling," *Atmos. Res.*, vol. 233, Mar. 2020, Art. no. 104716.

[4] L. Shi, J. Zhang, D. Zhang, T. Igbawua, and Y. Liu, "Developing a dust storm detection method combining support vector machine and satellite data in typical dust regions of Asia," *Adv. Space Res.*, vol. 65, no. 4, pp. 1263–1278, Feb. 2020.

[5] W. Wang, A. Samat, J. Abuduwaili, Y. Ge, P. De Maeyer, and T. Van De Voorde, "A novel hybrid sand and dust storm detection method using MODIS data on GEE platform," *Eur. J. Remote Sens.*, vol. 55, no. 1, pp. 420–428, Dec. 2022.

[6] Y. Hu, D. Y. Luo, K. Hua, and H. M. Lu, "Overview on deep learning," *CAAI Trans. Intell. Syst.*, vol. 14, no. 1, pp. 1–19, Aug. 2019.

[7] R. Zhang, W. P. Li, and T. Mo, "Review of Deep Learning," *Inf. Control*, vol. 47, no. 4, pp. 385–395, Aug. 2018.

[8] A. Onan, "A fuzzy-rough nearest neighbor classifier combined with consistency-based subset evaluation and instance selection for automated diagnosis of breast cancer," *Expert Syst. Appl.*, vol. 42, no. 20, pp. 6844–6852, Aug. 2015.
[9] A. Tokgoz and G. Unal, "A RNN based time series approach for forecasting Turkish electricity load," in *Proc. 26th Signal Process. Commun. Appl. Conf. (SIU)*, Izmir, Turkey, May 2018.
[10] T. G. Huang and L. Yu, "Application of SDAE-LSTM model on financial time series forecasting," *Comput. Eng. Appl.*, vol. 55, no. 1, pp. 142–148, Aug. 2019.
[11] C. Feng, J. Zhang, W. Zhang, and B.-M. Hodge, "Convolutional neural networks for intra-hour solar forecasting based on sky image sequences," *Appl. Energy*, vol. 310, Mar. 2022, Art. no. 118438.
[12] A. Onan and M. A. Tocoglu, "A term weighted neural language model and stacked bidirectional LSTM based framework for sarcasm identification," *IEEE Access*, vol. 9, pp. 7701–7722, 2021.
[13] A. Onan, S. Korukoğlu, and H. Bulut, "Ensemble of keyword extraction methods and classifiers in text classification," *Expert Syst. Appl.*, vol. 57, pp. 232–247, Sep. 2016.
[14] Y. Lu and S. Z. Zheng, "A comparative study learning and general ensemble method," *Chin. Sci.*, vol. 11, no. 4, pp. 372–379, Feb. 2018.
[15] X. Zheng, W. Chen, Y. You, Y. Jiang, M. Li, and T. Zhang, "Ensemble deep learning for automated visual classification using EEG signals," *Pattern Recognit.*, vol. 102, Jun. 2020, Art. no. 107147.
[16] B. He, L. Ye, M. Pei, P. Lu, B. Dai, Z. Li, and K. Wang, "A combined model for short-term wind power forecasting based on the analysis of numerical weather prediction data," *Energy Rep.*, vol. 8, pp. 929–939, Nov. 2022.
[17] Y. Caii, X. F. Zhu, and Z. G. Sun, "Semi-supervised and ensemble learning: A review," *Comput. Sci.*, vol. 77, pp. 7–13, Aug. 2017.
[18] W. Tang and Z. H. Zhou, "Bagging-based selective clusterer ensermble," *J. Softw.*, vol. 4, pp. 496–502, Aug. 2005.
[19] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016, pp. 326–366.
[20] C. L. Wu, Q. Li, and J. X. Hou, "PM 2.5 concentration prediction using convolutional netural networks," *Sci. Surv. Mapping*, vol. 43, no. 8, pp. 68–75, Aug. 2018.
[21] L. Moshe, L. V. Ya, and P. Allan, "Multilayer feedforward networks with a nonpolynomial activation function can approximate any function," *Pergamon*, vol. 6, no. 6, pp. 861–867, Aug. 1993.
[22] N. J. Guliyev and V. E. Ismailov, "A single hidden layer feedforward network with only one neuron in the hidden layer can approximate any univariate function," *Neural Comput.*, vol. 28, no. 7, pp. 1289–1304, Aug. 2016.
[23] V. Vapnik, "Statistical learning theory," *Ann. Inst. Stat. Math.*, vol. 55, no. 2, pp. 371–389, Aug. 2003.

QING-DAO-ER-JI REN received the Ph.D. degree from Xidian University, China. He is currently a Professor with the School of Information Engineering, Inner Mongolia University of Technology. His research interests include artificial intelligence and data mining.

NA LI received the Ph.D. degree from Inner Mongolia University, in 2021. She is currently an Associate Professor with the School of Science, Inner Mongolia University of Technology. Her research interests include optimization and artificial intelligence.

WEIMING ZHANG is currently pursuing the master's degree with the School of Information Engineering, Inner Mongolia University of Technology. His research interests include artificial intelligence and data mining.