

Article

Not peer-reviewed version

Improving Precipitation Forecasting in Jilin Province Using Deep Learning and Gaussian Noise

[Zhang Yibo](#) , Wang Chengcheng , Wang Pengcheng , [Zhang Lu](#) ^{*} , [Yu Qingbo](#) ^{*} , Xu Hui , [Wang Maofa](#)

Posted Date: 5 September 2023

doi: 10.20944/preprints202309.0156.v1

Keywords: deep learning; feature attribution; gaussian noise; LSTM; precipitation prediction; RMSE



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Improving Precipitation Forecasting in Jilin Province Using Deep Learning and Gaussian Noise

Yibo Zhang ^{1,4,†}, Chengcheng Wang ^{1,†}, Pengcheng Wang ^{2,†}, Lu Zhang ^{3,*}, Qingbo Yu ^{1,*}, Hui Xu ¹ and Maofa Wang ³

¹ Jilin Provincial Meteorological Information and Network Center, Changchun 130000, China

² Software Engineering Institute, East China Normal University, Shanghai 200063, China

³ Guangxi Key Laboratory of Trusted Software, Guilin University of Electronic Technology, Guilin 541004, China

⁴ College of Geo-Exploration Science and Technology, Jilin University, Changchun 130026, China

* Correspondence: zhangl@guet.edu.cn (L.Z.); yuqingbo2004@126.com (Q.Y.)

† These authors contributed equally to this work. (Yibo Zhang, Pengcheng Wang, Chengcheng Wang are co-first authors).

Abstract: This paper explores the use of different deep learning models for predicting precipitation in 56 meteorological stations in Jilin Province, China. The models used include Stacked-LSTM, Transformer, and SVR, and Gaussian noise is added to the data to improve their robustness. Results show that the Stacked-LSTM model performs the best, achieving high prediction accuracy and stability. The study also conducts variable attribution analysis using LightGBM and finds that temperature, dew point, precipitation in previous days, and air pressure are the most important factors affecting precipitation prediction, which is consistent with traditional meteorological theory. The paper provides detailed information on the data processing, model training, and parameter settings, which can serve as a reference for future precipitation prediction tasks. The findings suggest that adding Gaussian noise to the dataset can improve the model's generalization ability, especially for predicting days with zero precipitation. Overall, this study provides useful insights into the application of deep learning models in precipitation prediction and can contribute to the development of meteorological forecasting and applications.

Keywords: deep learning; feature attribution; gaussian noise; LSTM; precipitation prediction; RMSE

1. introduction

Precipitation has a significant impact on various aspects such as ecology, environment, and climate, and accurate prediction of precipitation is therefore of great importance for social and economic development. However, due to the high randomness and uncertainty of precipitation, precipitation forecasting has always been a highly challenging issue. In order to improve the accuracy of precipitation prediction, researchers have been exploring various different forecasting methods and techniques.

In earlier years, basic linear models such as autoregressive integrated moving average (ARIMA) and seasonal autoregressive integrated moving average (SARIMA) were used to predict precipitation in the field of meteorology [7,8]. Due to the multivariate and complex nature of precipitation prediction, traditional models of this kind have difficulty in exploring the deep-level relationships between precipitation and variables compared to emerging deep models. Therefore, the use of traditional linear models has limited effectiveness in predicting long-term precipitation. Machine learning models and deep learning models have shown great potential in solving the prediction problem in this field.

In the past few decades, machine learning methods have been applied to the field of meteorology for precipitation prediction, with the most widely used being the support vector regression (SVR) model [2–4]. For example, Shen Chiang et al. [11] used the SVR model to predict Taiwan's Rainfall-Runoff, and compared the results with the traditional Hydrologic Modeling System (HEC-HMS). They concluded that even with low training sets, the SVR model's predictions were still better than

those of the traditional HEC-HMS model for different time steps, and that the HEC-HMS model would require additional effort to collect field data to determine geographically related parameters such as land use and soil type. Additionally, in the study conducted by Ping-Feng Pai et al. [10], the idea of recurrent neural networks was introduced, and the recurrent support vector regression (RSVR) model was used for precipitation prediction. The study showed that RSVR is an effective means of predicting precipitation.

In recent years, deep learning methods have made significant progress in the field of precipitation prediction. Poornima Unnikrishnan et al. [13] used a hybrid SSA-ARIMA-ANN model to forecast daily rainfall on the West Coast of Maharashtra, India, and achieved good prediction results, demonstrating the promising application of ANN models in precipitation forecasting. Radhikesh Kumar et al. [5,6] used a deep neural network (DNN) to predict monthly precipitation in India, injecting new vitality and research methods into the traditional field of precipitation prediction. Suting Chen et al. [15] used a long short-term memory (LSTM) model based on recurrent neural networks (RNNs) to perform precipitation nowcasting, and their study showed that the two-stream convolutional LSTM achieved state-of-the-art prediction performance on a real-world large-scale dataset, and it is a more flexible framework that can be conveniently applied to other similar time series.

Many researchers use time series models, neural network models, and linear models to predict precipitation. However, these methods usually only consider the characteristics of the time series itself and ignore the processing of precipitation data. Techniques such as wavelet analysis, fast Fourier transform, and differencing can add noise to precipitation data, optimize the many zero-value problems in precipitation data, and enhance the robustness of models, thus better capturing periodic and trend changes in time series. Methods such as support vector machines and deep learning can model and predict nonlinear and non-stationary time series with high prediction accuracy. Therefore, in recent years, some new prediction methods and technologies have emerged. For example, Vahid Nourani and others combined wavelet analysis with the ANN model [1,12], transformed the time series model with wavelet transform, and used the ANN model for prediction [36,37], resulting in a significant improvement in precipitation prediction performance compared to experiments without wavelet transform. In addition, Weiwei Xiao et al. [14] used the Wavelet Neural Networks (WNN) model based on wavelet transform to predict precipitation in Taiwan, China. The experimental results show the potential of neural networks based on wavelet transform technology in precipitation prediction.

Therefore, researchers can explore the use of different methods and techniques to improve the accuracy of precipitation forecasts and contribute to the development of the ecological environment and social economy.

2. Exploratory Data Analysis (EDA)

2.1. General Description of the Study Region

The research area of this paper is located in Jilin Province in northeastern China, and precipitation information from 56 different meteorological bureaus in cities and counties across Jilin Province was obtained. Jilin Province, located in the central part of Northeast China, is an important commodity grain base in China. 80% of annual rainfall is concentrated in summer, and the regional difference of precipitation is very obvious, which leads to natural disasters such as drought, floods and so on. The losses caused by meteorological disasters are obviously increased, threatening the lives and property of the broad masses of the people, and causing great damage to economic, social development and ecological environment. In recent years, scholars have done a lot of research on the characteristics of rainfall in Jilin Province [16–21]. The geographical locations of the meteorological stations obtained in this paper are shown in Figure 1 below.

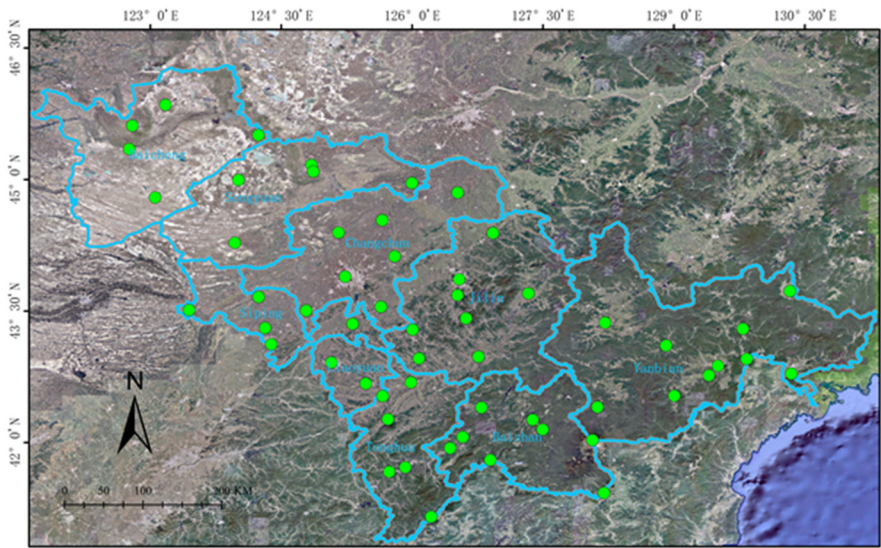


Figure 1. Distribution map of meteorological bureaus in Jilin province.

The geographical location of Jilin City is between 125°40' to 127°56' east longitude and 42°31' to 44°40' north latitude. It borders Yanbian Korean Autonomous Prefecture to the east, and is adjacent to Changchun and Siping to the west, Harbin in Heilongjiang Province to the north, and Baishan, Tonghua, and Liaoyuan to the south. The total area of the city is 27,120 km². Jilin City is located in the transition zone from the Changbai Mountains to the Songnen Plain in the central and eastern part of Jilin Province. The terrain gradually decreases from southeast to northwest. Jilin City has a temperate continental monsoon climate with distinct seasons. The study of its precipitation not only has an important impact on the activities of residents in Jilin region, but also provides important reference for global precipitation prediction work. This paper used daily precipitation data and 16 daily variables affecting precipitation, including daily average precipitation, daily wind speed, daily maximum wind speed, daily maximum wind direction, daily maximum wind direction degree, daily average dew point temperature, daily average temperature, air pressure, and relative humidity, from 56 different meteorological stations in Jilin Province provided by Jilin Meteorological Bureau for prediction. The maximum time span of the variables is the data from 1960 to 2022. Table 1 shows the average values of each variable at each station in 2022.

Table 1. The average values of each variable at each station in 2022.

Station	Pre	DAWS	D10WS	DMWS	DMWSD	DMWSDD	DEWS	DEWSD	DEWSDD	TDMWS	TDEWS	DADPT	DAAT	DAAP	DARH
1	1.1	2.69	2.68	5.55	8.44	234.75	9.64	8.47	232.2	1336.55	1357.33	-2.07	6.37	995.41	60.61
2	1.35	3.04	3.02	6	8.38	201.39	9.92	8.19	203.56	1344.01	1349.68	-2.01	6.35	996.14	60.71
3	1.2	3.24	3.22	6.45	8.55	208.26	10.21	8.57	214.2	1304.41	1338.58	-1.99	5.86	997.09	62.35
4	2.19	3.1	3.1	6.1	9.27	227.92	9.48	9	227.62	1254.33	1269.05	0.41	5.54	997.46	73.94
5	1.75	2.28	2.29	4.65	9.74	205.28	8.8	9.85	214.76	1321.98	1329.42	-0.27	6.43	997.43	66.22
6	1.83	2.75	2.73	5.31	9.35	216.67	9.19	9.27	221.42	1300.28	1300.85	-1.03	6.44	996.47	63.37
7	1.63	2.56	2.57	5.21	6.58	149.27	9.07	6.49	155.05	1326.13	1313.74	-0.8	5.93	997.64	66.3
8	0.77	3.2	3.16	6.14	8.87	214.68	10	8.7	215.13	1322.12	1341.15	-1.31	6.42	996.51	63.23
9	1.91	2.85	2.88	5.62	4.84	159.95	9.64	5.23	169.03	1308.81	1356.47	-0.44	6.69	992.07	64.67
10	1.61	3.42	3.42	6.83	10.2	217.56	10.21	10.08	223.21	1298.57	1298.48	-0.82	4.87	990.53	71.04
11	1.82	3.03	3.02	6.33	10.67	217.21	9.8	10.19	214.53	1292.86	1296.03	-0.11	5.53	994.24	71.45
12	1.8	2.98	2.98	6.09	10.26	221.09	9.6	9.98	216.56	1337.54	1329.78	-0.62	6.08	994.24	66.47
13	2.42	2.83	2.82	6.12	9.9	218.51	9.88	10	218.94	1355.51	1330.57	0.06	6.13	994.08	69.29
14	1.74	3.02	3.03	5.82	10.25	214.58	9.49	10.25	214.24	1290.64	1310.26	-0.83	5.39	990.84	68.51
15	2.55	2.83	2.81	5.71	10.62	218.05	9.56	10.73	221.56	1254.15	1252.03	-0.58	5.41	982.67	69.76
16	2.09	3.24	3.23	6.56	9.8	236.21	10.45	9.45	233.62	1359.67	1350.7	0.3	7.24	1001.12	65.92
17	2.4	2.81	2.81	5.82	10.9	227.77	9.96	10.68	222.11	1295.7	1329.68	0.77	7.52	996.2	67.21
18	2.13	2.79	2.78	5.5	10.29	225.35	9.68	10.2	221.5	1285.27	1346.75	0.35	7.11	997.74	66.76
19	2.2	3.29	3.29	6.3	10.67	235.25	10.29	10.27	233.44	1268.46	1266.52	-0.29	7.09	991.37	64.39
20	2.8	3.04	3.03	6.6	11.19	225.41	10.46	10.96	222.84	1296.7	1308.71	0.53	6.67	994.01	69.92
21	2.01	2.88	2.86	5.8	12	226.58	9.98	11.64	230.65	1323.49	1328.11	-0.24	6.96	986.54	64.36
22	2.73	2.41	2.42	5.24	11.02	219.86	9.16	10.99	221.29	1307.31	1304.11	0.4	6.39	985.55	70.68
23	2.19	2.99	3.01	6.36	11.43	220.93	9.86	11.06	219.83	1299.59	1318.08	0.12	6.66	988.95	67.57
24	2.94	2.68	2.69	6.38	9.65	242.3	9.77	9.79	241.38	1353.35	1363.22	-0.22	5.84	982.95	69.57
25	2.78	1.94	1.94	4.62	10.59	232.97	8.71	10.62	237.38	1345.81	1324.11	-0.28	6.1	987.75	68.49
26	2.16	3.2	3.21	6.4	6.98	168.45	10.48	6.67	166.09	1424.58	1429.96	0.11	6.76	980.2	67.22
27	3.51	2.35	2.36	5.16	9.33	229.1	10.01	9.09	229.82	1244.2	1331.55	-1.45	4.93	951.7	68.31
28	2.25	2.58	2.58	6.08	9.85	220	9.22	9.68	220.77	1386.59	1343.64	-0.25	4.86	979.83	74.06
29	1.99	2.4	2.37	5.33	8.78	230.97	9.42	8.76	224.94	1332.33	1363.42	-1.54	4.69	952.59	68.11
30	1.81	2.36	2.35	5.8	10.64	233.27	10.33	10.48	233.35	1362.89	1393.69	-1.14	4.99	971.4	69.65
31	1.68	1.68	1.68	4.75	9.04	241.8	8.35	9.31	244.52	1395.24	1375.6	-2.06	3.81	968.63	70.36

32	1.78	1.92	1.93	4.9	7.45	216.15	8.46	7.36	214.41	1401.61	1407.38	-0.95	5.04	985.25	69.91
33	2.83	2.19	2.19	5.25	10.16	243.55	9.27	10.41	239.23	1318.23	1316.88	0.25	6.37	985.35	70.99
34	2.82	2.69	2.69	6.7	10.84	221.17	9.98	10.44	221.06	1360.42	1365.24	-0.08	4.99	974.54	74.89
35	2.6	2.37	2.38	6	10.01	232.46	9.62	10.11	232.87	1399.21	1377.09	-0.32	5.17	975.44	72.27
36	3.19	2.17	2.15	4.84	9.66	230.74	8.41	9.76	224.36	1343.46	1351.79	-0.17	6.32	974.88	67.84
37	3.17	2.34	2.33	5.39	6.54	257.53	8.96	6.93	238.71	1312.6	1360.21	0.1	6.15	972.33	70.16
38	2.55	1.85	1.85	4.96	11.2	215.7	8.81	11.36	219.6	1330.82	1350.26	0.21	5.39	984.26	74.37
39	2.8	2.46	2.45	6.15	10.28	215.6	9.47	10.3	221.06	1351.18	1363.72	0.12	5.33	979.06	73.33
40	3.18	1.7	1.7	4.53	10.55	240.4	8.55	11.15	242.74	1355.57	1364.39	-1.36	3.99	948.44	72.93
41	3.46	1.85	1.86	5.61	9.9	216.92	8.92	9.74	214.93	1359.03	1349.73	-1.09	4.02	947.65	74
42	3.23	2.45	2.44	5.38	10.48	206.3	9.36	10.88	214.81	1315.28	1330.69	-1.98	4.36	924.81	67.85
43	2.64	2.56	2.57	5.64	10.78	254.66	9.85	10.87	245.29	1267.1	1297.85	-2.3	3.68	929.31	70.71
44	1.77	2.06	2.08	5.08	9.38	224.68	9.05	9.7	247.56	1366.9	1345.5	-1.56	5.99	958.23	63.75
45	10	7.67	7.64	14.03	11.74	228.95	20.52	11.24	222.34	1372.44	1399.55	3.06	6.77	738.99	82.3
46	3.02	1.88	1.89	5.12	10.83	234.18	8.99	10.92	239.51	1359.45	1387.96	-1.88	3.75	929.81	71.08
47	1.78	2.24	2.24	5.43	8.91	220.24	9.23	9.2	223.2	1394.16	1370.45	-0.29	6.57	986.05	66.51
48	2.1	3.65	3.64	8.09	7.73	214.05	11.57	7.73	214.19	1406.64	1394.05	1.24	6.75	1010.12	71.95
49	1.64	3.09	3.05	7.41	9.3	204.13	11.18	9.33	207.43	1457.41	1444.04	-0.79	6.52	984.2	64.72
50	1.95	2.74	2.74	6.53	8.32	214.5	10.61	7.96	213.9	1400.31	1405.3	-0.18	6.76	998.1	65.58
51	2.66	1.71	1.7	3.99	9.33	181.06	7.49	9.02	187.51	1374.46	1355.33	0.31	5.4	971.18	75.23
52	2.47	1.8	1.81	4.69	9.06	197.42	8.16	8.99	204.79	1354.55	1359.28	0.04	6.32	968.35	69.18
53	3.11	1.68	1.67	4.96	11.69	220.01	8.55	11.24	227.46	1390.85	1414.32	-0.72	4.8	954.38	72.64
54	2.9	1.72	1.73	4.31	8.86	181.37	7.27	8.69	174.79	1410.86	1387.27	0.41	5.42	971.32	75.27
55	3.19	1.87	1.87	4.66	6.71	164.65	7.83	7.41	191.28	1443.22	1459.31	1.69	7.53	989.55	71
56	2.59	2.05	2.05	5.28	5.51	242.9	9.23	5.73	219.55	1371.58	1394.56	-2.12	3.4	924.45	71.88

The daily maximum wind direction and the daily average wind direction are non-numeric values. In the table above, the label-encoding algorithm is used to label and map each different wind direction value to a unique positive integer value, and all N positive integer values are continuous.

Table 2. Statistics of rainfall datasets.

index	count	Mean(mm)	Standard deviation(mm)	Precipitation(mm)			
				25th percentile	50th percentile	75th percentile	MAX
NWJ	1960-2022	412.82	112.60	334.12	409.43	483.48	797.82
NCJ	1960-2022	566.16	149.33	452.99	563.54	673.47	1057.14
SWJ	1960-2022	510.89	201.78	434.53	536.64	632.89	1023.72
CJ	1960-2022	571.41	304.27	483.98	621.21	756.13	1485.33
EJ	1960-2022	578.53	168.36	486.75	587.86	680.25	1065.65
SCJ	1960-2022	675.46	155.43	568.55	644.55	769.25	1086.71
SEJ	1960-2022	686.27	349.66	585.69	729.22	878.96	1895.42

2.2. Data Cleaning

The dataset used in this paper contains 15 variables related to precipitation in Jilin Province from 1960 to 2022, including (i) precipitation (Pre), (ii) daily average wind speed (DAWS), (iii) daily average 10-meter wind speed (DA10WS), (iv) daily maximum wind speed (DMWS), (v) daily maximum wind speed direction (DMWSD), (vi) daily maximum wind speed direction in degrees (DMWSDD), (vii) daily extreme wind speed (DEWS), (viii) daily extreme wind speed direction (DEWSD), (ix) daily extreme wind speed direction in degrees (DEWSDD), (x) time of daily maximum wind speed (TDMWS), (xi) time of daily extreme wind speed (TDEWS), (xii) daily average dew point temperature (DADPT), (xiii) daily average air temperature (DAAT), (xiv) daily average air pressure (DAAP), and (xv) daily average relative humidity (DARH). The daily maximum wind speed direction and daily extreme wind speed direction are represented by characters indicating the direction, such as east, south, west, and north. In the data processing process, LabelEncoder is used to encode this data. The LabelEncoder method can encode multiple discrete data points and map n different discrete data points to a positive integer dataset in the range [0, n-1]. It is commonly used in deep learning for processing non-numeric variables [22,23].

After non-numeric encoding, missing values in the data are interpolated using cubic interpolation method [24]. However, this method cannot effectively predict continuously missing data, causing errors in the interpolation process that will be compounded by errors from subsequent deep learning. Therefore, for each variable in the training set, continuous missing data segments are discarded if their length is more than 10% larger than the length of continuous data before and after the missing segment. For example, for the 13th meteorological station in Jilin Province, data from January 2 to April 20, 1979 is missing, and to ensure the completeness of the seasonal characteristics of precipitation, the data from this station for that year is removed from the training data. After the above processing, a total of 960,081 rows and 17 columns of complete variable data are obtained in this paper.

The obtained dataset is normalized for all dependent variables during the training process, while the independent variable of precipitation is not normalized. The normalization of dependent variables speeds up the computation of the model during operation, while not normalizing the precipitation variable allows the model to better learn the changing patterns of precipitation.

According to the Markov hypothesis, the current state only depends on the previous few states of the time series. This can be formalized as:

$$p(x_1, \dots, x_N) = p(x_1) \prod_{n=2}^N p(x_n | x_1, \dots, x_{n-1}) \quad (1)$$

The joint probability distribution of an N-observation sequence in a first-order Markov chain is given by:

$$p(x_1, \dots, x_N) = p(x_1) \sum_{n=2}^N p(x_n | x_1, \dots, x_{n-1}) \quad (2)$$

In this study, a joint dependent variable is generated by combining the previous 30 days of observations for the upcoming predicted days, which is used as the unit for precipitation prediction and training.

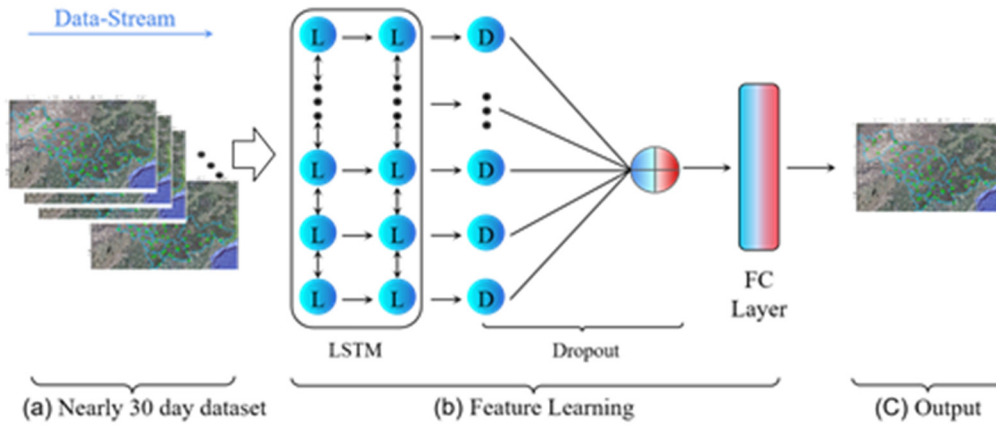


Figure 2. 30-day consolidated data forecasting process.

3. Research Method

In this paper, LSTM, Transformer, and SVR models were used for precipitation prediction, and methods such as wavelet transform, Gaussian noise, and Fourier transform were introduced to preprocess precipitation data and optimize the prediction of zero precipitation.

3.1. SVR Method

SVR is a regression model based on Support Vector Machines (SVM) that can be used to solve prediction problems with continuous target variables. Unlike traditional regression models, SVR maps input data to a high-dimensional space using a kernel function, and finds the optimal hyperplane in that space to fit the data.

The core idea of SVR is to transform the dataset into a high-dimensional space, making the data linearly separable or approximately linearly separable in the new space. By using a kernel function, SVR effectively avoids computation problems in high-dimensional space, enabling efficient regression analysis.

Specifically, the SVR model determines the optimal regression function by minimizing the error between the training data and the model. During training, the SVR model first maps the training data to a high-dimensional space, and then uses support vectors to determine the optimal hyperplane. This hyperplane is the optimal solution that minimizes the prediction error of the model while maintaining low complexity.

$$y = f(x) + \epsilon \quad (3)$$

where y is the target variable, x is the independent variable, $f(x)$ is the regression function, and ϵ is the error term.

SVR typically uses the least squares method to solve the regression equation in order to simplify the problem. Assuming there are N training samples (x_i, y_i) , the goal of the least squares method is to minimize the sum of squared residuals.

$$\min_{f \in H} \frac{1}{2} \|f\|_H^2 + C \sum_{i=1}^N l_\varepsilon(y_i - f(x_i)) \quad (4)$$

where H is the Hilbert space with the kernel function, and C is a regularization parameter.

$$l_\varepsilon(y_i - f(x_i)) = \begin{cases} 0, & \text{if } |y_i - f(x_i)| \leq \varepsilon \\ |y_i - f(x_i)| - \varepsilon, & \text{otherwise} \end{cases} \quad (5)$$

To transform nonlinear problems into linear ones, SVR employs kernel functions to map input data to a high-dimensional space. The kernel function typically takes the form of:

$$K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j) \quad (6)$$

3.2. Transformer Method

The Transformer is a sequence-to-sequence (seq2seq) model based on attention mechanism, proposed by Google in 2017 [30], aimed at addressing the shortcomings of traditional recurrent neural network (RNN) and LSTM models. The Transformer model is widely used in natural language processing tasks such as machine translation, text generation, and speech recognition.

The Transformer model adopts an Encoder-Decoder architecture, where the input sequence is processed by multiple layers of encoders, and the output sequence is generated through the decoder. Each encoder and decoder contains two modules: multi-head attention mechanism and feed-forward neural network. The attention mechanism is used to extract key information from the input sequence, while the feed-forward neural network is used to process and transform the extracted information.

The attention mechanism in the Transformer model is based on self-attention, which can automatically adjust the weights of different positions according to the context relationships in the input sequence, so as to extract key information. The formula for computing self-attention is as follows:

$$Attention(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (7)$$

The Q , K , and V in this formula are the query, key, and value vectors of the input sequence, respectively, with d_k representing the dimensionality of the vectors. This formula can be seen as calculating the similarity between the query vector Q and all key vectors K , normalizing the results, and then multiplying the normalized result with the value vector V to obtain the output.

In the Transformer model, there are not only single attention mechanisms but also multi-head attention mechanisms. Multi-head attention mechanisms can simultaneously extract different aspects of the input sequence, thereby improving the expressive power and generalization ability of the model.

In summary, the Transformer model effectively solves the problems of traditional recurrent neural network models by introducing innovations such as self-attention and multi-head attention mechanisms, and has become one of the popular models in the field of time-series.

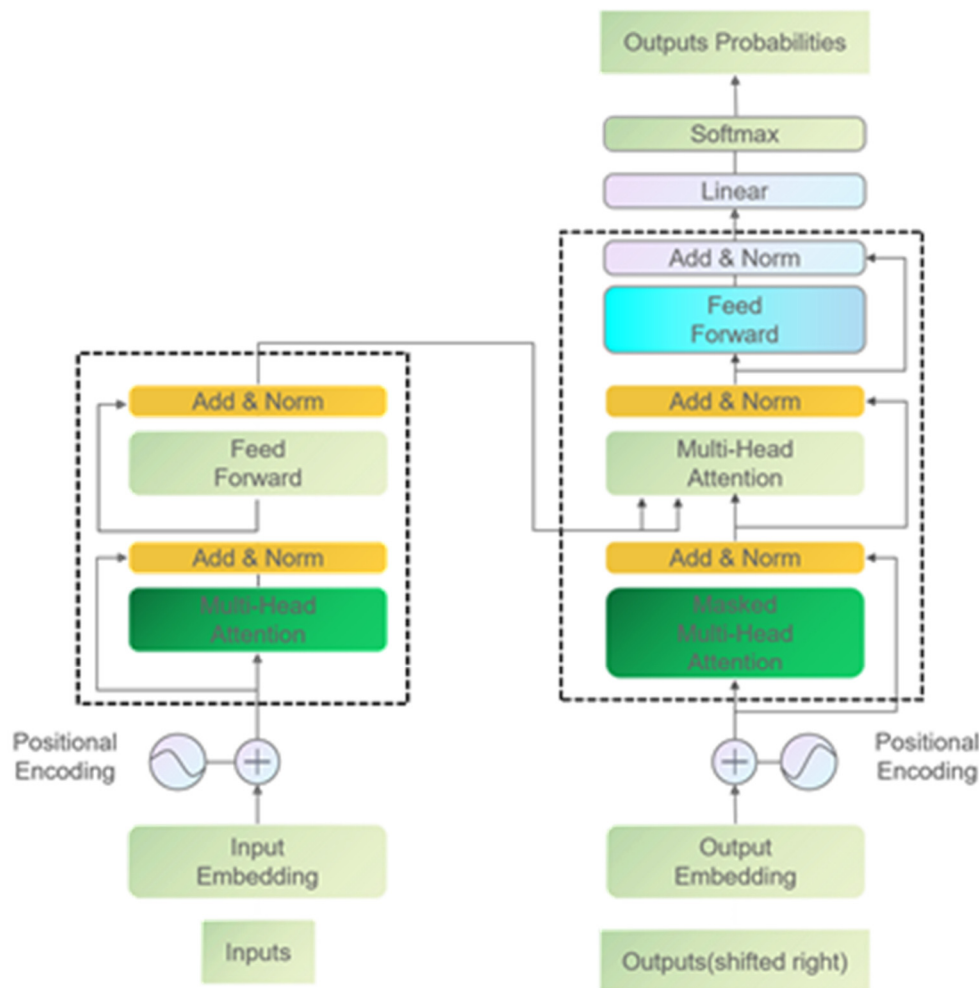


Figure 3. LSTM architecture.

3.3. LSTM Method

The LSTM model [25] is a type of deep learning model used to process data with a time-series structure, such as weather forecasting, speech recognition, and natural language processing. LSTM is a type of Recurrent Neural Network (RNN) designed to solve the problems of vanishing and exploding gradients in traditional RNN models.

Unlike traditional RNNs, each neuron in the LSTM model has three gates (a type of neural network structure used to control the flow of information in and out), namely the input gate, forget gate, and output gate. These gates allow the network to selectively remember and forget information, enabling the network to better handle long sequential data.

Specifically, the LSTM model consists of an input layer, an output layer, and one or more LSTM layers. In the input layer, the LSTM model receives time-series data as input and feeds it into the LSTM layer. In the LSTM layer, LSTM units generate the current hidden state based on the current input and the previous hidden state, and pass it to the next time step. This way, the LSTM model can capture long-term dependencies in the data and pass this information to subsequent time steps.

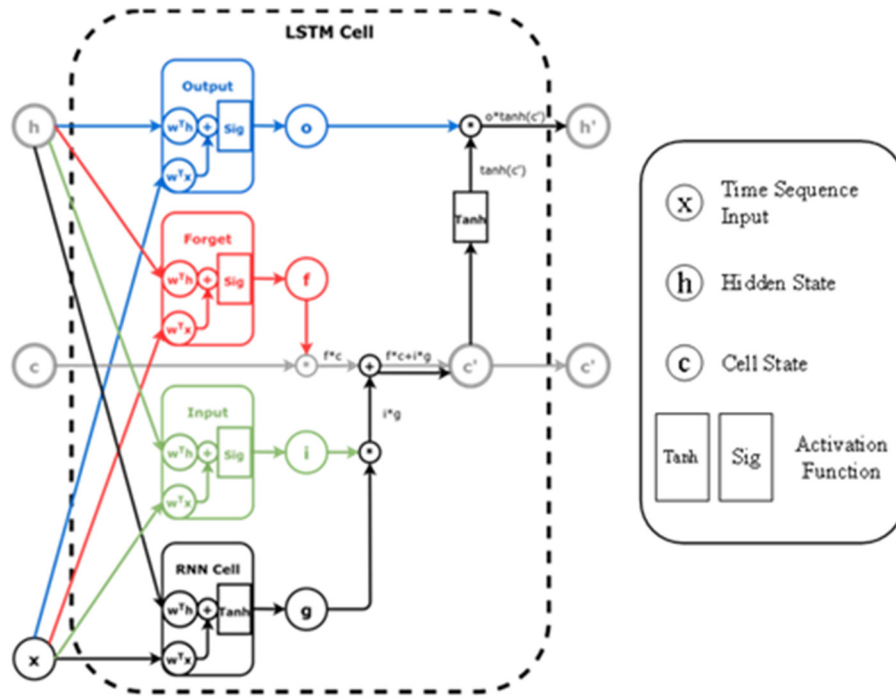


Figure 4. LSTM architecture.

In the above figure, sig represents the Sigmoid function:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (8)$$

The implementation of the forget gate can be represented as:

$$\begin{aligned} f_t &= \sigma(W_f \cdot [H_{t-1}, X_t] + b_f) \\ C_{f_t} &= f_t * C_{t-1} \end{aligned} \quad (9)$$

The forget gate uses Sigmoid activation to decide what information should be forgotten or retained. It calculates the value of C based on H_{t-1} and X_t , where C ranges from 0 to 1. A value of 0 indicates complete forgetfulness, while a value of 1 means complete retention of information.

The implementation of the input gate can be represented as:

$$\begin{aligned} i_t &= \sigma(W_i \cdot [H_{t-1}, X_t] + b_i) \\ \hat{C} &= \tanh(W_C \cdot [H_{t-1}, X_t] + b_C) \\ C_{i_t} &= i_t * \hat{C} \end{aligned} \quad (10)$$

At this step, again, a Sigmoid is used to determine whether the data should be updated. A tanh layer is then used to create a vector that will be combined with the Sigmoid output to update the cell state.

The updating process of the cell state is as follows:

$$C_t = C_{f_t} + C_{i_t} = f_t * C_{t-1} + i_t * \hat{C}_t \quad (11)$$

The implementation of the output gate can be represented as:

$$\begin{aligned} o_t &= \sigma(W_o \cdot [H_{t-1}, X_t] + b_o) \\ C_{o_t} &= \tanh(C_t) \\ H_t &= o_t * C_{o_t} \end{aligned} \quad (12)$$

The content output by the final output layer is based on the Cell State. First, the content of the Cell State is compressed to the range of $[-1, 1]$ by the tanh function, and then multiplied with the output of the Sigmoid function to obtain the final output. During the training process, the LSTM model uses backpropagation algorithm to update parameters to minimize the loss function. Through continuous training, the LSTM model can gradually learn the patterns in the data and predict future results given input data.

In the Jilin region precipitation prediction task, the LSTM model can receive historical precipitation data as input and predict the precipitation in the future. With appropriate tuning and training, the LSTM model can achieve good prediction performance.

In this paper employs two commonly used types of LSTM structures, namely Vanilla LSTM [26,27] and Stacked LSTM [34,35]. Vanilla LSTM is the traditional LSTM model structure, while Stacked LSTM has additional hidden layers, making the model deeper and more accurately described as a deep learning technique. The depth of neural networks enables their success in challenging prediction problems. Additional hidden layers can be added to a multilayer perceptron neural network, making it deeper. The additional hidden layers are understood as recombining learned representations from previous layers, and creating new representations at higher levels of abstraction [28]. For example, from linear to shape to object. A large enough single hidden layer multilayer perceptron can be used to approximate most functions. Increasing the depth of the network provides an alternative solution that requires fewer neurons and faster training. Finally, increasing depth is a representative optimization technique.

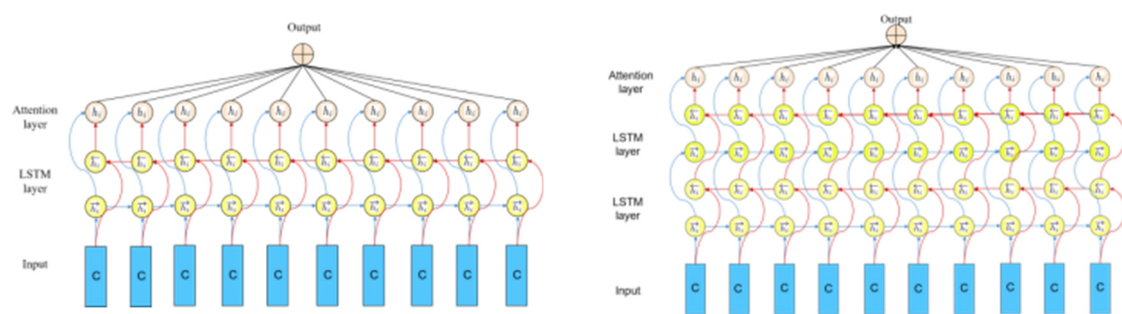


Figure 5. Vanilla LSTM and Stacked LSTM.

3. Experiments and Results

In this paper, the dataset of each station was divided, with precipitation data from 1960 to 2020 used as the training set and precipitation data from 2021 used as the test set. To avoid specific optimizations for the test set during model optimization, which may reduce the robustness of the model, the final result was evaluated using the latest precipitation data from 2022 as the measure. Additionally, based on the precipitation data from 56 meteorological stations in 2020, there were a total of 20,252 days with valid daily precipitation, among which 12,072 days had zero precipitation, accounting for approximately 59.61%. A large amount of identical zero precipitation data leads to poor performance of the deep learning model in predicting zero precipitation. Therefore, this article used the method of adding Gaussian noise to the data to add noise to the data. Rainfall with and without added noise is shown in the figure below:

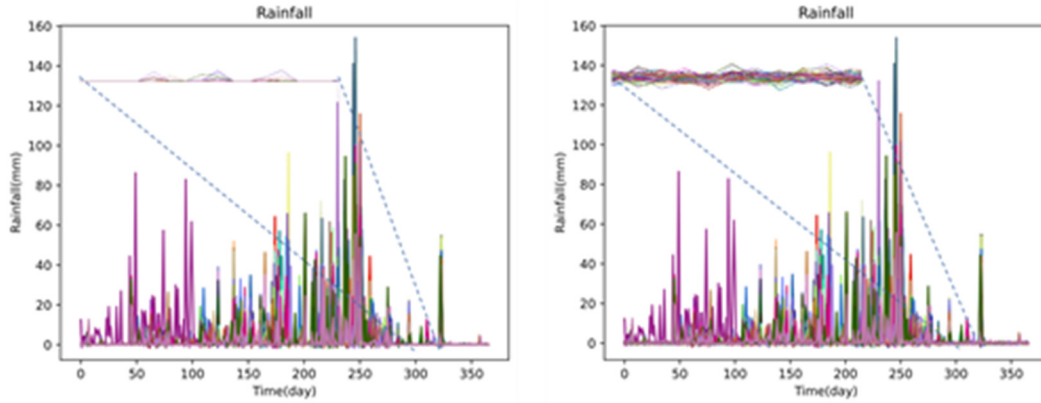


Figure 6. Rainfall with and without added Gaussian noise.

Gaussian noise, also known as normal distribution noise, is a commonly used random noise model. According to the central limit theorem, when a random variable is the sum of many independent random variables with the same distribution, the random variable will approximate a normal distribution. Therefore, adding random noise that follows a Gaussian distribution to data can simulate many noise sources in the real world. Adding Gaussian noise to data is a commonly used data augmentation technique in deep learning. It can effectively help the model learn more data features, increase the model's generalization ability, and thus improve the model's accuracy. Specifically, the formula for adding Gaussian noise to data x is as follows:

$$\hat{x} = x + \varepsilon \quad (13)$$

The epsilon (ε) is a random number that follows a Gaussian distribution with a mean of 0 and a standard deviation of σ . Typically, its value is chosen based on the specific situation, ranging from 0.01 to 0.1. In this article, it is set to 0.1. Its probability distribution function can be represented as:

$$p(\varepsilon) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\varepsilon^2}{2\sigma^2}} \quad (14)$$

After comparing the three models mentioned above, this article found that using the Stacked-LSTM model for predicting precipitation had the best results. Its parameters included a loss function of MSE (Mean Square Error), an optimizer of Adam (adaptive moment estimation), a batch size of 1, and 250 epochs. The specific predicted values and actual values are shown in the following figure:

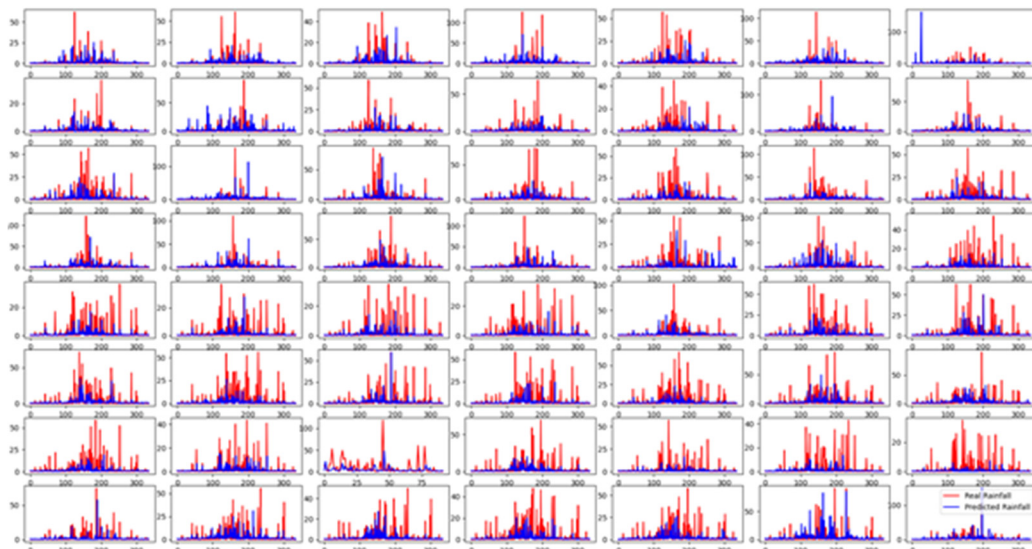


Figure 7. The precipitation prediction of 56 meteorological stations for 2022.

The root mean square error (RMSE) is a commonly used metric for measuring the difference between the predicted values and the true values in regression models. It helps evaluate the accuracy of the model. The RMSE is calculated by taking the average of the squared differences between the predicted and true values, and then taking the square root. The mathematical formula for RMSE is:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (15)$$

The RMSE (Root Mean Square Error) is a common metric to measure the difference between the predicted and true values in regression models, and it can help us evaluate the accuracy of the model. A smaller RMSE indicates a smaller gap between the predicted and true values, and a better predictive capability of the model. In this article, the RMSE for the best model's prediction on 56 meteorological stations is shown in the figure below:

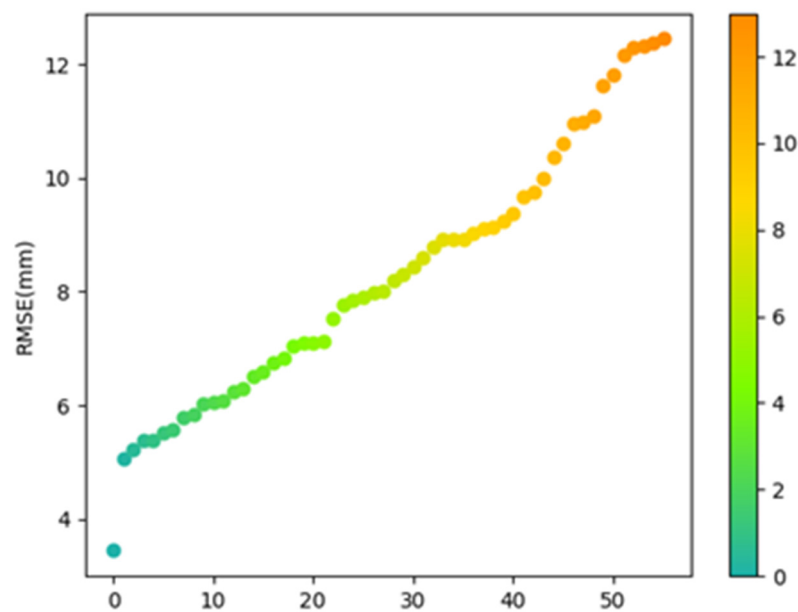


Figure 8. The RMSE of 56 meteorological stations.

4. Conclusions

This paper used three different models (Stacked-LSTM, Transformer, SVR) to predict precipitation in 56 meteorological stations in Jilin Province, and improved the robustness of the models by adding Gaussian noise to the data. The experimental results showed that the Stacked-LSTM model performed the best in this task, achieving high prediction accuracy and stability. For the 16 different variables used in this article, the article conducted attribution analysis on the variables using the LightGBM [31–33] algorithm, and the conclusions are as follows:

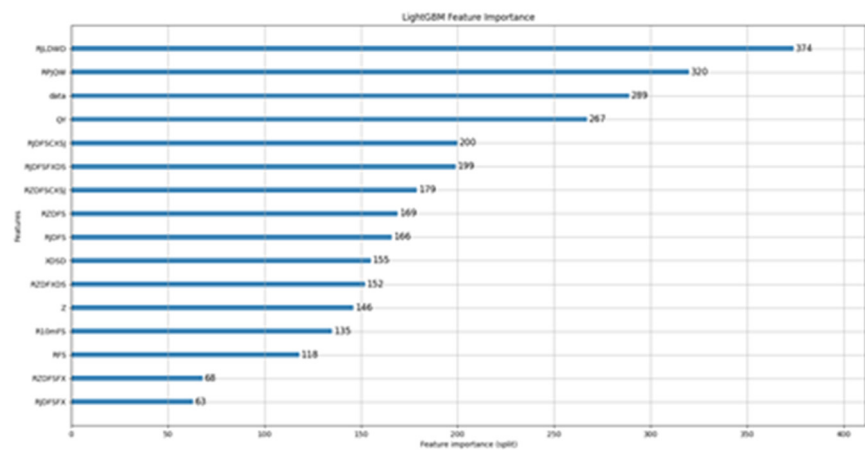


Figure 9. Variable attribution analysis.

This paper explores the use of deep learning models, specifically Stacked-LSTM, Transformer, and SVR, for predicting precipitation in 56 weather stations in Jilin province. Gaussian noise is added to the data to improve the robustness of the models. The results show that Stacked-LSTM performs the best, achieving high prediction accuracy and stability.

The importance of different variables in the prediction process was analyzed using the LightGBM algorithm for variable attribution analysis. The findings show that the importance of different variables is consistent with traditional meteorological experience and theory. The most influential factors include daily dew point temperature, daily air temperature, previous precipitation, and air pressure. Dew point temperature and air temperature ensure the generation of rainwater in the atmosphere and are crucial factors for predicting the likelihood of precipitation. Previous precipitation provides important trends and directions for predicting precipitation, while air pressure affects whether water vapor in the air will rise to a sufficient height to generate precipitation.

Furthermore, the article provides detailed information on the training process, including data preprocessing and model parameter settings, which can be useful for future precipitation prediction tasks. Additionally, the article finds that adding Gaussian noise can improve the model's generalization ability for datasets with many zero precipitation days, leading to better prediction results.

In conclusion, this article verifies the performance of different models in precipitation prediction tasks and provides a reference for related research fields. The use of more advanced data preprocessing techniques and model optimization methods can further improve model performance and applicability, promoting the development of meteorological prediction and applications in the future.

Acknowledgments: This work was supported by Guangxi science and technology base, and talent project (2020AC19134).

References

1. Vahid Nourani; Mohammad T. Alami; Mohammad H. Aminfar (2009). A combined neural-wavelet model for prediction of Ligvanchai watershedprecipitation. , 22(3), 466–472. DOI:10.1016/j.engappai.2008.09.003
2. Xu, D., Zhang, Q., Ding, Y. et al. Application of a hybrid ARIMA-LSTM model based on the SPEI for drought forecasting. Environ Sci Pollut Res 29, 4128–4144 (2022). DOI:s11356-021-15325-z
3. Dehe Xu , Qi Zhang ,et alApplication of a Hybrid ARIMA–SVR Model Based on the SPI for the Forecast of Drought—A Case Study in Henan Province, China,Journal of Applied Meteorology and Climatology,vol. 59,pp 1239–1259,2,2022,DOI:JAMC-D-19-0270.1
4. Yu Xiang, Ling Gou, Lihua He, Shoulu Xia, Wenying Wang,A SVR–ANN combined model based on ensemble EMD for rainfall prediction,Applied Soft Computing,Vol. 73,pp 874-883,2018,DOI: j.asoc.2018.09.018.

5. Kumar, R., Singh, M.P., Roy, B. et al. A Comparative Assessment of Metaheuristic Optimized Extreme Learning Machine and Deep Neural Network in Multi-Step-Ahead Long-term Rainfall Prediction for All-Indian Regions. *Water Resour Manage*, Vol. 35, pp1927–1960, 2021. DOI : s11269-021-02822-6
6. Kumar D, Singh A, Samui P, Jha RK (2019) Forecasting monthly precipitation using sequential modelling. *Hydrol Sci J* 64(6):690–700
7. Xu, DH (Xu, Dehe) ; Ding, Y (Ding, Yan) ; Liu, H (Liu, Hui), et al, Applicability of a CEEMD-ARIMA Combined Model for Drought Forecasting: A Case Study in the Ningxia Hui Autonomous Region, 2022, 13(7):1109.
8. Poornima, S. Pushpalatha, M. Drought prediction based on SPI and SPEI with varying timescales using LSTM recurrent neural network. *Soft Computing*, 2019, 8399-8412.
9. Ziegel E R, Box G E P, Jenkins G M, et al. Time Series Analysis, Forecasting, and Control. *Journal of Time*, 1976, 31(2):238-242.
10. Ping-Feng Pai; Wei-Chiang Hong (2007). A recurrent support vector regression model in rainfall forecasting. , 21(6), 819–827. doi:10.1002/hyp.6323
11. Chiang s, Chang CH, Chen WB. Comparison of Rainfall-Runoff Simulation between Support Vector Regression and HEC-HMS for a Rural Watershed in Taiwan, *WATER*, 2022, 14(2):191
12. Hai-Long Liu; An-Ming Bao; Xi Chen; Ling Wang; Xiang-liang Pan (2011). Response analysis of rainfall-runoff processes using wavelet transform: a case study of the alpine meadow belt. , 25(14), 2179–2187. doi:10.1002/hyp.7969
13. Unnikrishnan, P., Jothiprakash, V. Hybrid SSA-ARIMA-ANN Model for Forecasting Daily Rainfall. *Water Resour Manage* 34, 3609–3623 (2020). DOI: 10.1007/s11269-020-02638-w
14. Xiao W W, Luan W J. The Application of Wavelet Neural Networks on Rainfall Forecast. *Applied Mechanics and Materials*, 2013, 263-266(4):3370-3373.
15. Chen, S., Xu, X., Zhang, Y. et al. Two-stream convolutional LSTM for precipitation nowcasting. *Neural Comput & Applic* 34, 13281–13290 (2022). DOI: 10.1007/s00521-021-06877-9
16. Liu B . Spatial Distribution and Periodicity Analysis of Annual Precipitation in Jilin Province[J]. *Water Resources and Power*, 2018.
17. Li Y F, Xu S Q, Zhang T, et al. Regional characteristics of interdecadal and interannual variations of summer precipitation in Jilin Province. *Meteorol. Disaster Prev*, 2018, 25:38-43
18. Li M, Wu Z F, Meng X J, et al. Temporal-spatial distribution of precipitation over Jilin Province during 1958–2007. *J. Northeast Normal Univ. (Nat. Sci. Ed.)*, 2010, 42:146-51
19. Sun Y, Yang Z C, Yun T, et al. Weather characteristics of short-term heavy rainfall and large-scale environmental conditions in Jilin Province. *Meteorol. Disaster Prev*, 2015, 229-12+33
20. Li M, Wu Z F, Meng X J, et al. Temporal-spatial distribution of precipitation over Jilin Province during 1958–2007. *J. Northeast Normal Univ. (Nat. Sci. Ed.)*, 2010, 42:146-51
21. Lin, M K; Wei, N; Xie, J C; Lu, K M (2019). Study on multi-time scale variation rule of rainfall during flood season in Jilin Province in recent 65 years. *IOP Conference Series: Earth and Environmental Science*, 344, 012143. DOI:10.1088/1755-1315/344/1/012143
22. G, T.R., Bhattacharya, S., Maddikunta, P.K.R. et al. Antlion re-sampling based deep neural network model for classification of imbalanced multimodal stroke dataset. *Multimed Tools Appl* 81, 41429–41453 (2022). DOI: 10.1007/s11042-020-09988-y
23. Senan, Ebrahim Mohammed, Al-Adhaileh, Mosleh Hmoud, et al, Diagnosis of Chronic Kidney Disease Using Effective Classification Algorithms and Recursive Feature Elimination Techniques, *Journal of Healthcare Engineering*, vol. 2021, pp 2040-2295, DOI: 10.1155/2021/1004767.
24. L.A. Shirman; C.H. Séquin (1992). Procedural interpolation with curvature-continuous cubic splines. , 24(5), 278–286. doi:10.1016/0010-4485(92)90081-k
25. Hochreiter S , Schmidhuber J . Long Short-Term Memory. *Neural Computation*, 1997, 9(8):1735-1780.
26. Graves A , Schmidhuber J. Framewise phoneme classification with bidirectional LSTM networks[C] , *IEEE International Joint Conference on Neural Networks. IEEE*, 2005. DOI: 10.1016/j.neunet.2005.06.042.
27. Greff, Klaus; Srivastava, Rupesh K.; Koutnik, Jan; Steunebrink, Bas R.; Schmidhuber, Jurgen (2016). LSTM: A Search Space Odyssey. *IEEE Transactions on Neural Networks and Learning Systems*, (), 1–11. DOI:10.1109/tnnls.2016.2582924.
28. Hermans M , Schrauwen B . Training and analysing deep recurrent neural networks. *Advances in neural information processing systems*, 2013.
29. Graves, Alex; Mohamed, Abdel-rahman; Hinton, Geoffrey (2013)., *Speech and Signal Processing - Speech recognition with deep recurrent neural networks*. PP 6645–6649. DOI:10.1109/ICASSP.2013.6638947
30. Vaswani A , Shazeer N , Parmar N , et al. Attention Is All You Need. *arXiv*, 2017.
31. Liang, Junchao , et al. "Estimation of Stellar Atmospheric Parameters with Light Gradient Boosting Machine Algorithm and Principal Component Analysis." (2022). DOI: 10.3847/1538-3881/ac4d97

32. [1] Cui Z , Qing X , Chai H ,et al.Real-time rainfall-runoff prediction using light gradient boosting machine coupled with singular spectrum analysis.Journal of Hydrology, 2021(Pt.C):603.DOI:10.1016/j.jhydrol.2021.127124.
33. Yunxin, Liang J , Wu W ,et al.Product Marketing Prediction based on XGboost and LightGBM Algorithm[C]//0[2023-07-01].DOI:10.1145/3357254.3357290.
34. Zeng c, Ma Cxm, Wang K, et la. Parking Occupancy Prediction Method Based on Multi Factors and Stacked GRU-LSTM[J]. IEEE ACCESS, 2022, 10: 47361-47370
35. Hermans M , Schrauwen B . Training and analysing deep recurrent neural networks[J]. Advances in neural information processing systems, 2013.
36. Cheng X, Rao Z, Chen Y , et al. Explaining Knowledge Distillation by Quantifying the Knowledge[J] 2020 IEEE CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020.

Short Biography of Authors



Yibo Zhang received the master's degree in Cartography and Geographical Information System from Jilin University, China. He is currently a lecturer with Jilin Provincial Meteorological Information and Network Center, China. He has published more than 10 papers in journals, including the World Geology, Meteorological Disaster Prevention, and others. His current research interests include machine learning, deep learning, GIS, data analysis, and meta-learning and their applications to address geological and Meteorological problems.



Chengcheng Wang received the master's degree in software engineering from Jilin University, China. She is currently an Associate Professor with Jilin Provincial Meteorological Information and Network Center, China. She has published 6 papers in journals, including the Agriculture of Jilin, Meteorological Disaster Prevention, Modern Agricultural Science and Technology, Journal of Sichuan Normal University (Natural Science), and others. Her current research interests include machine learning and deep learning and their applications to address Meteorological problems.



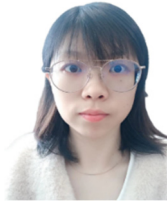
Pengcheng Wang is Graduate student in progress from East China Normal University, China. He has published 5 papers in journals, including the Computer & Telecommunication, and others. His current research interests include machine learning, and deep learning, and their applications to address geological problems.



Lu Zhang received the Master degree in electronic information engineering from Southwest Jiaotong University, China. Her current research interests include machine learning, mathematics, and their application in solving geological problems and the medical field.



Qingbo Yu is currently pursuing a master's degree in agricultural extension and graduated from Chengdu University of Information Technology. He is a senior engineer qualification at the Jilin Provincial Meteorological Information Network Center, mainly engaged in research in meteorological data processing and analysis, deep learning. More than 10 papers have been published in journals, and their data processing and analysis, as well as deep learning research achievements, are mainly applied in the field of meteorology.



Hui Xu graduated from Nanjing University with a master's degree in meteorology, she is an associate senior engineer, and she has had 8 papers accepted by national journals. Her main research interests are climate change, climate data analysis and processing.



Maofa Wang received the Ph.D. degree in geo-information engineering from Jilin University, China. He is currently a Professor with the School of Computer and Information Security, Guilin University of Electronic Technology, China. He has published more than 20 papers in international journals and conference proceedings, including the Journal of Seismology, Computers & Geosciences, and others. His current research interests include machine learning, deep learning, and meta-learning and their applications to address geological problems.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.