IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 1 August 2024, accepted 8 August 2024, date of publication 13 August 2024, date of current version 21 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3443023

RESEARCH ARTICLE

A Prediction Model for Soybean Meal Futures Price Based on CNN-ILSTM-SA

HAITAO LIU, YANG PAN, WEI LIU, AND JINGYANG WANG
Hebei University of Science and Technology, Shijiazhuang, Hebei 050018, China

Corresponding author: Jingyang Wang (ever211@163.com)

This work was supported in part by the Defense Industrial Technology Development Program under Grant JCKYS2022DC10, and in part by the Foundation of Science and Technology Research Project of Colleges and Universities in Hebei Province under Grant QN2023140.

ABSTRACT Predicting the price of soybean meal futures has always been a hot topic in the financial market. It is influenced by many factors and has complex nonlinear characteristics, which makes it difficult to predict accurately. This paper proposes a prediction model for soybean meal futures price based on CNN-ILSTM-SA, which consists of CNN (Convolutional Neural Network), ILSTM (Improved Long Short-Term Memory), and SA (Self-Attention mechanism). CNN can well perform feature extraction on the input time series data. The ILSTM model presented in this paper removes the output gate of LSTM (Long Short-Term Memory) and improves the forget gate and input gate, thus reducing ILSTM's complexity, shortening training time, and improving learning ability. SA can get the correlation between different input terms and predicted values, and assign different weight coefficients according to the correlation level to improve prediction accuracy. To verify the CNN-ILSTM-SA's effectiveness, we compare it with eighteen baseline models, and the experimental results show that it is optimal. The MAE (Mean Absolute Error) of CNN-ILSTM-SA is 39.0441, the smallest of all models. As the data of the soybean meal futures price fluctuates in the range of 3500-4500 in experiments, the proportion of the MAE value of 39.0441 is 0.87%-1.15%, which is at an extremely low level; R^2 (R-square) is 0.97487, which is closest to 1; The training time is 157.64s.

INDEX TERMS CNN, ILSTM, self-attention mechanism, soybean meal futures, price prediction.

I. INTRODUCTION

Soybean meal plays an important role in China's agriculture and animal husbandry. Soybean meal futures trading groups are huge, active, and have strong seasonal characteristics. Several factors (including weather conditions, supply and demand, international trade policy, and monetary policy) affect the soybean meal futures price. The interaction of these factors makes the fluctuation of soybean meal futures price more complicated, and it is difficult to accurately predict the price trend by simply analyzing one aspect [1]. In addition, the trend of soybean meal futures price has a complex nonlinear relationship that is difficult to capture by traditional linear models. Therefore, it is significant to establish an efficient prediction model for soybean meal futures price combined with deep learning models.

Accurately predicting soybean meal futures price is achievable with the ongoing advancements in machine learning. CNN's proposal has achieved great breakthroughs in time analysis and other aspects. However, when the amount of data is large, the problems of ''gradient disappearance'' and ''gradient explosion'' often occur in model training. The proposals of GRU (Gated Recurrent Unit) and LSTM can alleviate the above problems [2], [3].

The analyses of LSTM and GRU find that they decided to forget and retain historical and input data by introducing gate control technology. However, due to the large number of LSTM weight parameters and the complexity of model calculation, the training time will increase when the soybean meal futures data shows large volatility in the time dimension. Based on this problem, this paper presents an improved model of ILSTM based on LSTM by studying LSTM, GRU, and time series data. This model removes LSTM's output gate and improves the forget gate and input gate, so the training time is shorter, and the learning ability is stronger. This paper proposes a prediction model for soybean meal futures price based on CNN-ILSTM-SA. In this model, CNN can well extract feature values, and SA can analyze and calculate the correlation between various data and the predicted value. Eighteen models (including CNN, LSTM, ILSTM,

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

CNN-LSTM, GRU, LSTM-GRU, CNN-GRU, DT (Decision Tree), RF (Random Forest), MLP (MultiLayer Perceptron), IRF-LSTM (Improved Regularization Function-LSTM), ARFIMA-LSTM (Autoregressive Fractional Integral Moving Average-LSTM), LSTM-LightGBM (LSTM-Light Gradient Boosting Machine), Bi-LSTM (bidirectional-LSTM), CEEMDAN-LSTM (Complete Ensemble Empirical Mode Decomposition with Adaptive Noise Analysis-LSTM), CNN-ILSTM, CNN-LSTM-SA, and CNN-GRU-SA) are introduced as baseline models to verify the effectiveness of CNN-ILSTM-SA. MAE, R^2, and training time are used to evaluate the prediction results. Experimental results show that CNN-ILSTM-SA performs better than other models.

The following are this paper's main contributions:

(1) In terms of algorithms, ILSTM improves the algorithm of LSTM's forget gate and input gate. The 1-tanh function is introduced after the forget gate, which improves the model's ability to retain information. A new ALSG (Avoid Learning-Saturation Gate) is proposed, and c_{t-1} and ALSG are introduced into the input gate algorithm, which solves the model training oversaturation problem and improves the sensitivity of the model to input data.

(2) In terms of network structure, compared with LSTM, the ILSTM presented in this paper removes the output gate, reduces the weight parameters to four, and bias parameters to two under the premise of guaranteeing the accuracy of the prediction for soybean meal futures price. ILSTM has reduced computational complexity and has shorter training time.

(3) This paper presents a prediction model for soybean meal futures price based on CNN-ILSTM-SA. CNN performs feature extraction on the input time series data. SA analyzes the correlation between different input terms and the predicted value, assigning different weight coefficients according to the correlation level. Compared with the other eighteen forecasting models, the CNN-ILSTM-SA composite model performs best and can improve the prediction accuracy of soybean meal futures price.

II. RELATED WORK

Neural network models have been widely used at home and abroad in recent years. In 2015, Sundermeyer et al. used artificial neural networks to predict word language. The prediction accuracy was high, but the prediction time was long [4]. In 2019, Moon and Kim utilized LSTM to forecast the stock market index and its volatility, but they still needed to solve the problem of long predicting time [5]. Wang et al. used the network model composed of LSTM and GRU to forecast the PM2.5 of four cities with good prediction results [6]. Bukhari et al. used ARFIMA and LSTM to predict financial market, overcoming the overfitting problem of neural networks [7]. Nabipour et al. used DT, bagging, RNN, LSTM for comparative experiments, and the experiments showed that LSTM produced more accurate results [8]. Sun et al. proposed a RF (Random Forest) to predict the air pollution index. The air pollution index changes greatly with the season, and the prediction time is increased [9]. Tong et al. used the DT combination model to predict hormone receptors, and the experiment results proved that the combined model was more accurate than a single model [10]. Yang et al. utilized CNN-LSTM to forecast the concentration of suspended particulate matter in Seoul, and the experiment results showed that CNN-LSTM had a superior prediction effect [11]. Chang et al. utilized LSTM to forecast changes in air pollution, and it was challenging to obtain high-precision prediction due to the single prediction model [12]. Zhang et al. utilized LSTM to forecast the futures price of new energy. They used diversified data to improve the forecast accuracy, which would increase the model training time [13]. Wen et al. presented an air pollution forecast model based on CNN and LSTM to improve the forecast accuracy [14]. Zhu et al. utilized an attention-based CNN-LSTM model in the PM2.5 prediction process [15]. Wen and Zhu et al. employed CNN to address LSTM's feature extraction limitations. However, due to its numerous parameters, LSTM often requires extensive training time. Shu et al. used LSTM to identify human horizontal movements, but there was still the problem of long training time due to complex data features [16]. In 2021, Suebsombut et al. utilized LSTM to forecast soil moisture, and the results showed that the LSTM model performed well. However, when the dataset was large, the training time also increased [17]. Bhimavarapu used IRF-LSTM to enhance the performance of rainfall prediction [18]. In 2022, Tian et al. presented a hybrid model of LSTM and LightGBM to forecast the rise and fall of stocks. Although the accuracy has improved, the prediction time is still long [19]. Li et al. utilized Bi-LSTM to forecast soil cracks in pneumatic subsoiling. The effect was obvious, but the prediction time was relatively long [20]. Zhou et al. utilized the CEEMDAN-LSTM for carbon price prediction, and the prediction results were stable and reliable [21]. Yu et al. applied LSTM to effectively forecast the competitiveness trend of China's machinery equipment export [22]. Liu et al. used LSTM to predict vehicle lane changes, significantly improving prediction accuracy. The training time will increase significantly when the data set is too large [23]. Kervanci et al. utilized a hybrid model of LSTM-GRU to forecast Bitcoin's price, which can effectively reduce the prediction error. However, it still did not solve the problems of long training time and many parameters [24]. As mentioned above, this paper will be based on the existing research results, strive to improve the above deficiencies, and join innovative research.

III. MODELS

A. SA

SA is a commonly used attention mechanism for time series data processing, which allows the model to assign different weights to elements in the input series, capturing their

dependencies [25]. The principle of the SA mechanism is shown in Fig. 1.

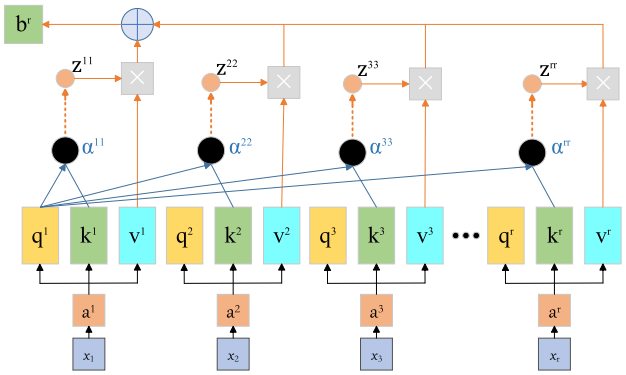

FIGURE 1. SA mechanism.

The SA mechanism is carried out through the following calculation steps: For each element in the input sequence, SA calculates k^i (key vector), q^i (query vector), v^i (value vector) of the i-th eigenvalue, which are obtained by multiplying the element a^i with the trained weight matrices w^q, w^k, w^v. The calculation formulas are shown in Formulas (1), (2), and (3).

$$k^i = w^k \cdot a^i \tag{1}$$

$$q^i = w^q \cdot a^i \tag{2}$$

$$v^i = w^v \cdot a^i \tag{3}$$

α^{ij} and z^{ij} denote the similarity and weight factor of eigenvalue i and j, respectively. α^{ij} is obtained by dividing the dot product of q^i and k^i by a scaling factor $\sqrt{d}$ (usually the square root of the dimension of k^i). $\sum_{j=1}^{n} \exp(\alpha^{ij})$ denotes the exponential operation of e on α^{ij}. The calculation formulas are shown in Formulas (4) and (5).

$$\alpha^{ij} = \frac{q^i \cdot k^j}{\sqrt{d}} \tag{4}$$

$$z^{ij} = \frac{exp(\alpha^{ij})}{\sum_{j=1}^{n} exp(\alpha^{ij})} \tag{5}$$

b^i denotes the SA layer's output, which is the eigenvector obtained by the sum of the v^i vector of the i-th eigenvalue multiplied by the weight factor z^{ij}. As shown in Formula (6).

$$b^i = \sum_{i=0}^{n} z^{ij} \cdot v^i \tag{6}$$

B. ILSTM

LSTM contains three types of gates (input gate, forget gate, and output gate) through which it stores and updates information to mine long-term dependencies in time series data [26]. GRU only contains two gates (update gate and reset gate), where the update gate combines the input and forget gates of LSTM [27]. This paper proposes a new ILSTM model based on gate control technology. The ILSTM's structure is different from that of LSTM and GRU. The structure of ILSTM is shown in Figure 2. The ILSTM first proposed in this paper consists of a forget gate and an input gate. ILSTM is based on the LSTM and has been improved and optimized in algorithm and structure. It is a brand-new model with improvement, innovation, high prediction, and high accuracy.

ILSTM improves the input gate and forget gate in terms of algorithm. The first is the forget gate, which can control the forgetting and retention of information. The $\sigma(x)$ is the sigmoid function. The input data at time step t is denoted by x_t. The previous time step t−1's hidden layer is denoted by h_{t-1}. W_{fh} represents the h_{t-1}'s weight of forget gate, W_{fx} represents x_t's weight, b_f represents the forget gate's bias, the forget gate f_t is shown in Formula (7).

$$f_t = \sigma\left(W_{fh} \cdot h_{t-1} + W_{fx} \cdot x_t + b_f\right) \tag{7}$$

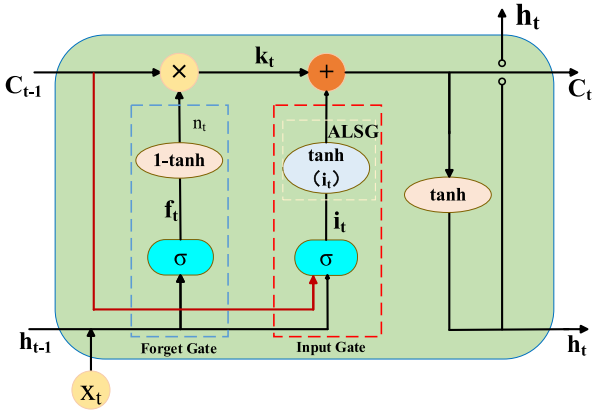

FIGURE 2. ILSTM structure.

The output range of the forget gate in LSTM after passing through the sigmoid function is (0, 1). Through the study of functions, we can know that the range of values of tanh functions between (0, 1) is (0, 0.76), and the range of values of 1-tanh functions between (0, 1) is (0.24, 1).

The ILSTM presented in this paper adds the 1-tanh function after the forget gate, so that the output range of the forget gate is changed to (0.24,1), as shown in Formula (8).

$$n_t = 1 - \tanh(f_t) \tag{8}$$

The value range of n_t is (0.24,1).

The mainline forgetting.. k_t [28] is calculated by c_{t-1} and n_t, where c_{t-1} is the cell state information of the previous time step t−1, k_t denotes the effect of information on the current cell state c_t, as shown in Formula (9).

$$k_t = n_t \times c_{t-1} \tag{9}$$

The introduction of the 1-tanh function helps to reduce information loss in the cell state so that the model can better retain important information. This improves the model's learning efficiency and stability and allows it to adapt better to various sequence data processing tasks.

The second is the improvement of the input gate. ILSTM introduces c_{t-1} to the input gate algorithm. The function of the input gate is to control how much of the current input

data x_t is stored in the memory cell, that is, how much can be saved in c_t. The introduction of c_{t-1} can improve the ability of ILSTM's input gate to capture the long-term dependence in the sequence. In terms of information transfer, it also helps ILSTM retain and transfer information about the previous state, better understand the overall context of the sequence and improve the ability to model the data. As shown in Formula (10), where W_{ih} and W_{ix} are the weights of h_{t-1} and x_t of the input gate, respectively, and b_i denotes the bias of the input gate.

$$i_t = \sigma\left(W_{ih} \cdot h_{t-1} + W_{ix} \cdot x_t + c_{t-1} + b_i\right) \tag{10}$$

According to the saturation property of the sigmoid function, when x is less than -6 and greater than 6, the value of the sigmoid activation function will approach 0 and 1, respectively. That is to say, when the input value is large or small, the derivative of the sigmoid function will tend to zero, resulting in the problem of gradient disappearance.

Therefore, ILSTM introduces ALSG in the input gate, which can alleviate this problem to a certain extent. The function of ALSG is to map the sigmoid function's output to the tanh activate function to expand the range of output values. This operation enhances input data's sensitivity and increases the amplitude of gradient propagation, which helps alleviate the problem of gradient disappearance, as shown in Formula (11).

$$ALSG = \tanh(i_t) \tag{11}$$

The introduction of ALSG aims to improve the sensitivity of the input data through nonlinear mapping when dealing with larger or smaller input values, thereby alleviating the sigmoid function's supersaturation problem.

According to the Formula (12), c_t is the information saved from the beginning to the present.

$$c_t = k_t + ALSG \tag{12}$$

h_t denotes the information saved at the current time. c_t determines how much information can be retained through the tanh function, as shown in Formula (13).

$$h_t = \tanh(c_t) \tag{13}$$

In terms of structure, ILSTM removes the output gate compared with LSTM and has a simple structure compared with GRU.

In terms of parameters, the parameters of ILSTM are optimized. Compared with LSTM, ILSTM reduces the number of the weight parameters from eight to four and the number of the bias parameters from four to two. Compared with GRU, ILSTM reduces the number of the weight parameters from six to four and the number of the bias parameters from three to two.

ILSTM enhances the overall learning ability by improving the internal unit structure of LSTM and optimizing weight and bias parameters.

C. CNN-ILSTM-SA

Fig. 3 shows the CNN-ILSTM-SA's structure. The model has five layers. The first layer is the data input and data preprocessing layer, which requires preprocessing operations such as normalization processing and 3D time series construction of the raw data. The second layer is the CNN layer, which inputs the normalized data into the convolution layer of CNN for calculation, selects the important feature data, and then enters the pooling layer for data dimensionality reduction operation to realize the feature extraction of soybean meal futures data. The third layer is the prediction layer, which inputs data into the ILSTM layer for calculation. The fourth layer is the SA layer, which uses SA to calculate the attention value of the data at different moments for the predicted value and obtain the output data. The last layer is the output layer, which de-normalizes and outputs the calculation results. CNN-ILSTM-SA has fully realized the prediction of soybean meal futures price.

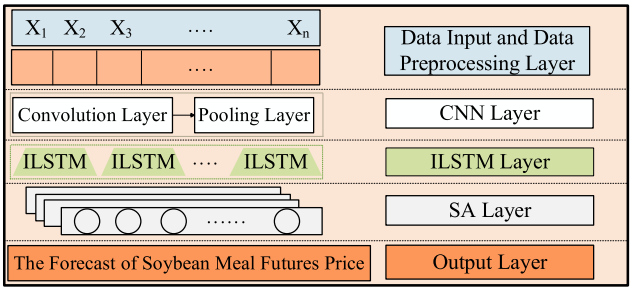

FIGURE 3. Prediction model for soybean meal futures price based on CNN-ILSTM-SA.

IV. EXPERIMENTS

A. EXPERIMENTAL ENVIRONMENT

In this paper, all the training and prediction experiments of soybean meal prediction models are carried out in the same operating environment of the same computer, and Table 1 displays the specific hardware and software environments.

TABLE 1. Experimental hardware and software environments table.

	Hardware/software	environment
Hardware environment	CPU	AMD Ryzen 9 5900HX
	Memory	32GB
	Hard Disk	1TB
	GPU	NVIDIA GeForce GTX
Software environment	PyCharm	2020.1.3 x64
	Python	3.7.0
	TensorFlow	1.14.0
	keras	2.1.0

B. DATA COLLECTION AND PREPROCESSING

Through the research and analysis of the existing market, the economic factors affecting soybean meal futures price mainly come from two aspects. On the one hand, corn is a substitute for soybean meal, and its market fluctuations will affect soybean meal futures price. On the other hand, fluctuations in

TABLE 2. Part of the raw data.

trade_date	open	high	low	close	vol	Corn	SPX	US30	NAS100	SH	USDCNH
2022/11/21	4119	4157	4108	4139	74868500	2841	3612.39	31990.8	10696.1	3085.044	7.1919
2022/11/22	4140	4187	4139	4168	67697100	2853	3588.84	32390.3	10845.4	3088.943	7.2275
2022/11/23											

TABLE 3. Experimental data.

trade_date	open	high	low	close	vol	Corn	SPX	US30	NAS100	SH	USDCNH
2022/11/21	4119	4157	4108	4139	74868500	2841	3612.39	31990.8	10696.1	3085.044	7.1919
2022/11/22	4140	4187	4139	4168	67697100	2853	3588.84	32390.3	10845.4	3088.943	7.2275
2022/11/23	4129.5	4172	4123.5	4153.5	71282800	2782	3363.55	32505.6	10852.3	3091.247	7.1773

the soybean meal futures market will fluctuate with changes in the domestic and foreign economies. The influencing factors selected in this paper include Corn futures price (Corn), Standard & Poor's 500 Index (SPX), Dow Jones Industrial Average (US30), Nasdaq Composite Index (NAS100), Shanghai Composite Index (SH), USD/CNY Exchange rate (USDCNH).

In this experiment, 1,508 DCE (Dalian Commodity Exchange) soybean meal futures and related influencing factors data from November 1, 2017, to January 11, 2024, are used as experimental data. The data are sourced from 'https://tushare.pro/'. Part of the raw data of soybean meal futures data is shown in Table 2.

The following is the processing of the raw data used in this experiment:

(1) Data alignment. Soybean meal futures belongs to the Chinese market, while some influencing data comes from the international market. Due to different legal holidays at home and abroad, the closing date of the futures market varies, resulting in the inability of some data dates at home and abroad to correspond. Therefore, aligning the influencing factor data with soybean meal futures data is necessary.

(2) Data filling. Data may be lost during data collection due to storage failures and other reasons, such as the data of November 23, 2022. Since futures trading data changes smoothly over time in most cases and the value does not change suddenly, this experiment uses the average data value from the previous two trading days to fill the missing part. Experimental data after data filling are shown in Table 3.

In this experiment, Input data's time is taken as a series, and the two-dimensional segmentation and three-dimensional construction are carried out. For example, suppose that there are X pieces of soybean meal price data and influencing factor data, and the three-dimensional construction is carried out according to step=1 and sequence length=5. The soybean meal price data and influencing factor data from the first to the fifth trading day constitute the Y_1 layer. The soybean meal price data and influencing factor data from the second to the sixth trading day constitute the Y_2 layer, and so on to obtain the X-4 layer (Y_1, $Y_2 \ldots Y_{X-4}$) data. In each layer, the prediction model takes the soybean price data and influencing factor data of the first four trading days as input and takes the closing price of soybean meal price data of the fifth trading day as output for training, verification, and evaluation. The construction method of the three-dimensional time series is shown in Fig.4.

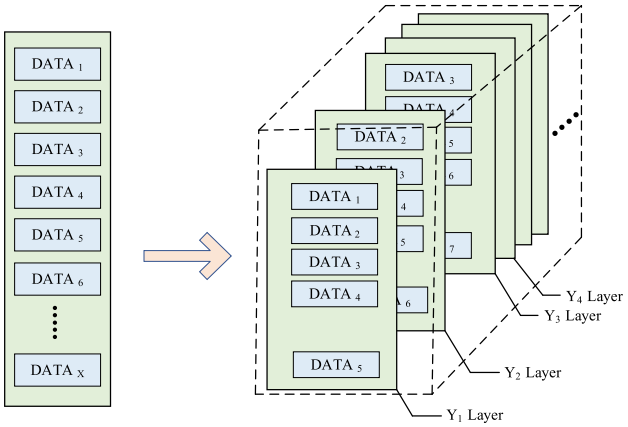

FIGURE 4. Construction method of the three-dimensional time series.

C. DATASET SEGMENTATION AND EVALUATION INDEX

This experiment's training, validation, and test sets' data ratios are 6:2:2. This division avoids model overfitting on a specific dataset.

This experiment uses MAE, R^2, and training time as comprehensive evaluation indexes of the model. The MAE is used to quantify the average absolute error between the predicted value and the true value, as shown in Formula (14).

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{14}$$

where n is the number of samples, $\hat{y}_i$ is the model predicted value of the i-th sample, and y_i is the true value of the i-th sample. The smaller the MAE value, the lower the deviation between the model-predicted value and true value, and the higher the prediction accuracy.

The R^2 describes the degree of similarity between the true value and the predicted value [21], as shown in Formula (15).

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{n} (\bar{y} - y_i)^2} \tag{15}$$

where n is the number of samples, $\hat{y}_i$ is the model predicted value of the i-th sample, y_i is the true value of the i-th sample, and $\bar{y}$ is the average of the true values. The closer the value of R^2 is to 1, the better the model fits the data.

Training time is the time to train the model using the training and validation set data. The smaller the training time, the better the model performs regarding computational resource utilization and algorithm efficiency.

D. ADJUSTMENT OF MODEL PARAMETERS

In order to adjust the parameters, this experiment uses a combination of the hyperparameter optimization technique (grid search optimization algorithm) and empirical mode. Parameter optimization aims to improve the model's expressiveness by finding a set of suitable parameters. Table 4 displays the details of the parameters for the final experimental setting.

TABLE 4. The details of the parameters.

Model	Layer	Parameters
MLP	MLP	hidden_layer_sizes=(4,4),activation='relu', solver='adam',random_state=150
RF	RF	n_estimators=1,max_depth=8, random_state=38,min_samples_leaf=0.2
DT	DT	max_depth=7.5,random_state=150
CNN	CNN	filters=16,kernel_size=3, strides=1,padding='valid'
GRU	GRU	Units=64,kernel_initializer='he_normal', activation='tanh',epochs=50
LSTM	LSTM	Units=64,kernel_initializer='he_normal', activation='tanh',epochs=50
IRF-LSTM	IRF	filters=16,kernel_size=1,activation='relu', input_shape=input_shape,pool_size=2
	LSTM	Units=64,kernel_initializer='he_normal', activation='tanh',epochs=50
LSTM-GRU	LSTM	Units=64,kernel_initializer='he_normal', activation='tanh',epochs=50
	GRU	Units=64,kernel_initializer='he_normal', activation='tanh',epochs=50
LSTM-LightGBM	LSTM	Units=64,kernel_initializer='he_normal', activation='tanh',epochs=50
	Light GBM	num_leaves=25,max_depth=5,learning_rate=0.06,min_split_gain=0.07,min_child_weight'=20,mi
ARFIMA-LSTM	ARFIMA	squared=126.6,df=12,p-value=0.003,t-Ratio=14.3
	LSTM	Units=64,kernel_initializer='he_normal', activation='tanh',epochs=50
Bi-LSTM	Bi-LSTM	Units=64,kernel_initializer='he_normal', activation='tanh',epochs=50
CEEMDAN-LSTM	CEEMDAN	n_samples=1000,noise_level=0.5,max_imf=10,ensemble_size=250,noise_seed=42
	LSTM	Units=64,kernel_initializer='he_normal', activation='tanh',epochs=50
CNN-LSTM-SA	CNN	filters=16,kernel_size=3,strides=1, padding='valid'
	LSTM	Units=64,kernel_initializer='he_normal', activation='tanh',epochs=50
	SA	output_dim=' 64'
CNN-GRU-SA	CNN	filters=16,kernel_size=3,strides=1, padding='valid'
	GRU	Units=64,kernel_initializer='he_normal', activation='tanh',epochs=50
	SA	output_dim=' 64'
ILSTM	ILSTM	Units=64,kernel_initializer='he_normal', activation='tanh',epochs=50
CNN-ILSTM-SA	CNN	filters=16,kernel_size=3,strides=1, padding='valid'
	ILSTM	Units=64,kernel_initializer='he_normal', activation='tanh',epochs=50
	SA	output_dim='64'

E. EXPERIMENTAL ANALYSIS

In this paper, eighteen models are used as the baseline models. The comparative experiments are carried out in the same experimental environment and using the same experimental data. Table 5 displays the experiment's results. The experimental results show that CNN-ILSTM-SA has the best prediction effect. The introduction of CNN and SA slightly increases the composite model's training time, but the prediction accuracy is improved. Fig.5 illustrates the MAE, R^2, and training time of each model.

TABLE 5. Experimental results.

Model	MAE	R^2	Training time(s)
MLP	216.0508	0.74653	287.16
RF	210.9263	0.76821	136.23
DT	190.3311	0.67481	102.01
CNN	89.9485	0.88571	147.50
GRU	55.6206	0.94479	152.45
LSTM	51.9843	0.94427	188.07
IRF-LSTM	51.4532	0.94166	209.46
LSTM-GRU	50.4113	0.94618	237.36
LSTM-LightGBM	49.4492	0.94824	211.78
ARFIMA-LSTM	48.7565	0.95116	238.37
CNN-LSTM	48.4028	0.95331	196.19
CNN-GRU	47.3433	0.95639	161.46
Bi-LSTM	46.3448	0.95729	251.84
CEEMDAN-LSTM	45.1044	0.95252	225.69
CNN-GRU-SA	45.1807	0.96692	172.26
CNN-LSTM-SA	44.2623	0.96528	215.41
ILSTM	43.5825	0.95884	139.51
CNN-ILSTM	42.9087	0.96556	149.44
CNN-ILSTM-SA	39.0441	0.97487	157.64

Through the evaluation indexes of the prediction results of each model, it can be found that:

(1) LSTM, GRU, and ILSTM are all variants of RNN in essence, and ILSTM has the best fitting effect after improving the internal control unit. Compared with LSTM, the R^2 of the ILSTM is 0.015 higher and the MAE is 8.40 lower. Compared with GRU, the R^2 of ILSTM is 0.014 higher and MAE is 12.04 lower. Fig.6 illustrates the comparison between the true value and the predicted value of the three models.

(2) ILSTM is an improved model based on LSTM, and its prediction effect is better than simple composite LSTM models. The MAE of ILSTM is 7.8707 lower than that of IRF-LSTM, 6.8288 lower than that of LSTM-GRU, 5.8667 lower than that of LSTM-LightGBM, and 5.1740 lower than that of ARFIMA-LSTM, 2.7623 lower than that of Bi-LSTM, and 1.5219 lower than that of CEEMDAN-LSTM. The R^2 of ILSTM is 0.01718 higher than that of IRF-LSTM, 0.01266 higher than that of LSTM-GRU, 0.01060 higher than that of LSTM-LightGBM, 0.00768 higher than that of ARFIMA-LSTM, 0.00155 higher than that of

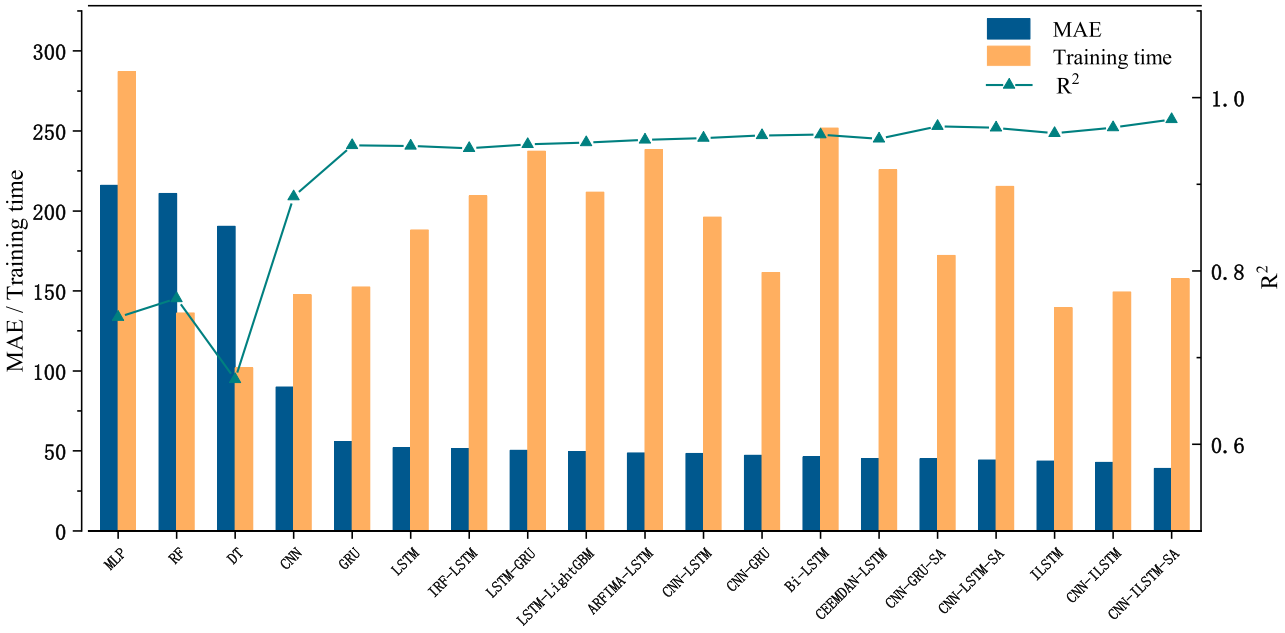

FIGURE 5. Comparison of MAE, R^2, training time.

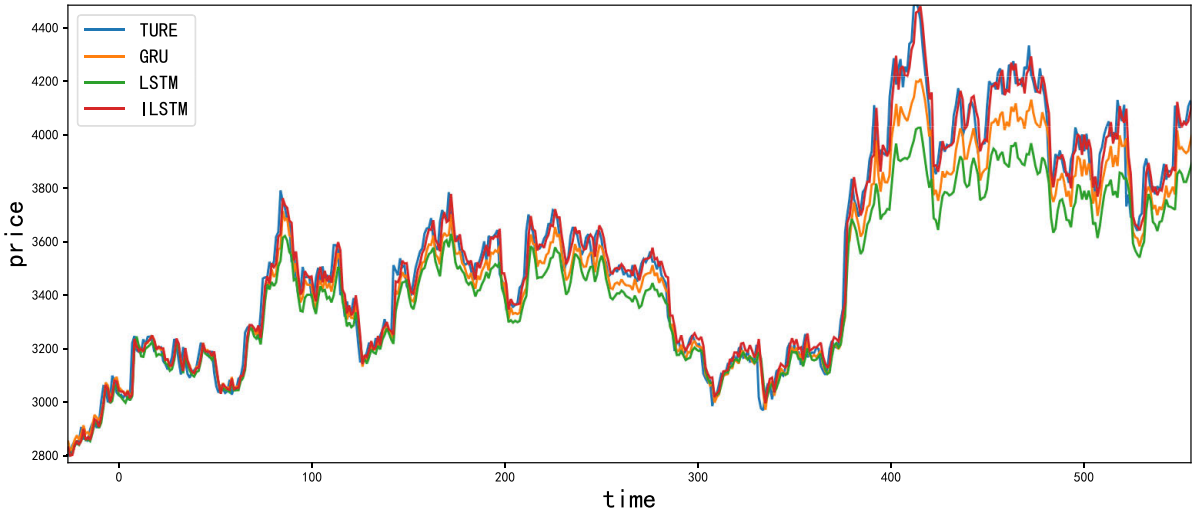

FIGURE 6. Comparison of true and predicted values of ILSTM, GRU, LSTM.

Bi-LSTM, and 0.00632 higher than that of CEEMDAN-LSTM.

(3) After the introduction of CNN and SA, the composite model can capture various features of the input data more comprehensively and improve the learning ability of complex relationships. Compared with the single model, the MAE of combined LSTM, GRU, and ILSTM is decreased, and R^2 is increased. Fig.7 illustrates the comparison of the two indexes after combining CNN and SA with a single model. Among all compared models, CNN-ILSTM-SA performs best with MAE of 39.0441, R^2 of 0.97487, and training time of 157.64s. The MAE of CNN-ILSTM-SA reaches around 40. Although this value is large, we should consider the specific case of the experimental dataset and the prediction task. In this paper's experimental data, soybean meal futures' prices fluctuate between 3500 and 4500, and the proportion of MAE around 40 is 1.42%-0.89%, which is very small. Compared with CNN-LSTM-SA and CNN-GRU-SA, the MAE of CNN-ILSTM-SA is decreased by 11.79% and 13.58%, respectively, and the R^2 is increased by 0.99% and 0.82%, respectively. Fig.8 illustrates the comparison between the true value and the predicted value of the models.

(4) When all parameters are the same and epoch=50, ILSTM has the shortest training time and the best performance in a single model. Compared with LSTM and GRU, the training time of ILSTM is increased by 25.8% and 8.49%, respectively. The composite

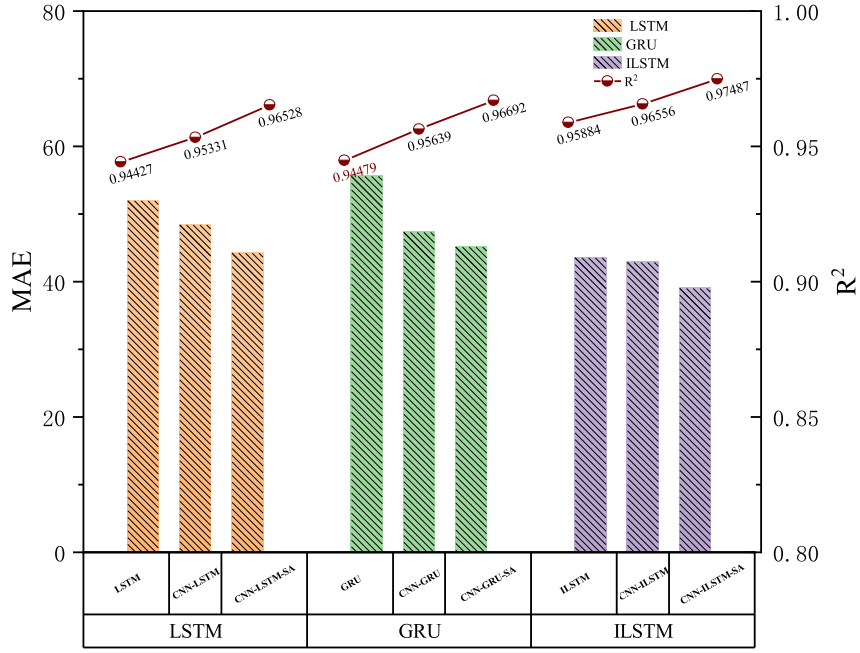

FIGURE 7. The comparison of the two indexes after combining CNN and SA with a single model.

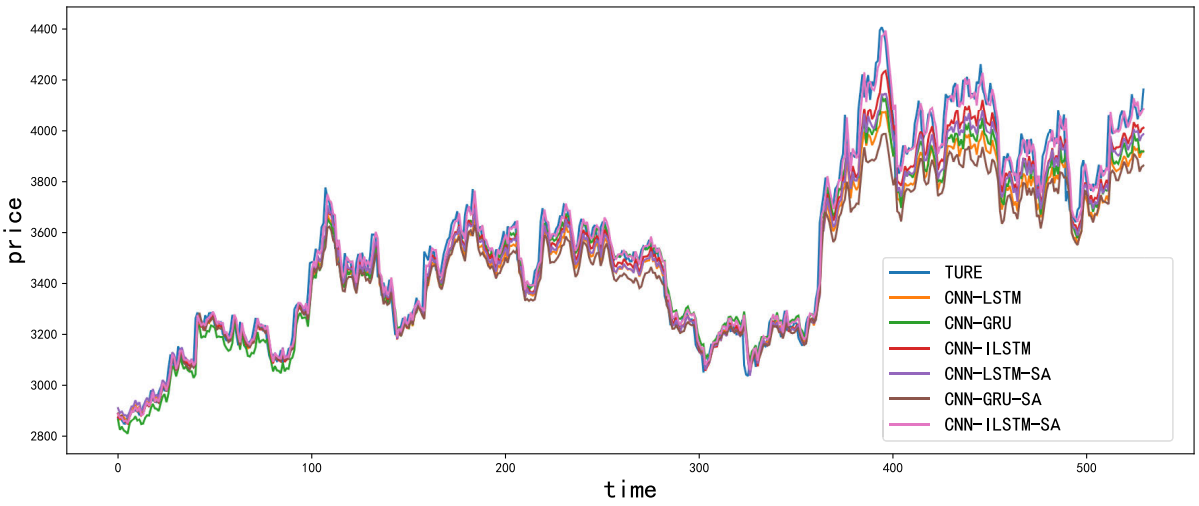

FIGURE 8. Comparison of the true value and predicted value of the composite model.

models of ILSTM performed well. Compared with CNN-LSTM and CNN-GRU, the training time of CNN-ILSTM is increased by 23.83% and 7.44%, respectively. Compared with CNN-LSTM-SA and CNN-GRU-SA, the training time of CNN-ILSTM-SA is increased by 26.82% and 8.49%, respectively. Experiments show that ILSTM removes the output gate and controls the weight parameters and bias parameters less than those of LSTM and GRU is very effective. The training time comparison is shown in Fig.9.

V. DISCUSSION

Through comparative experiments, we can find that the fitting degree of the prediction results of the recurrent neural network model is higher than the traditional regression prediction model. Though the traditional regression model can achieve a high-precision prediction of linear data, because the soybean meal futures data is a nonlinear and nonstationary time series, it is difficult for the traditional prediction model to learn the deep information of the data, and the traditional regression model has poor prediction effect on soybean meal futures price.

Under the same data set and experiment conditions, the comprehensive evaluation indexes of the CNN-ILSTM-SA are better than those of other models. Compared with LSTM and GRU, ILSTM can significantly shorten the training time due to the reduction of parameters while maintaining higher prediction accuracy. CNN-ILSTM-SA compared with

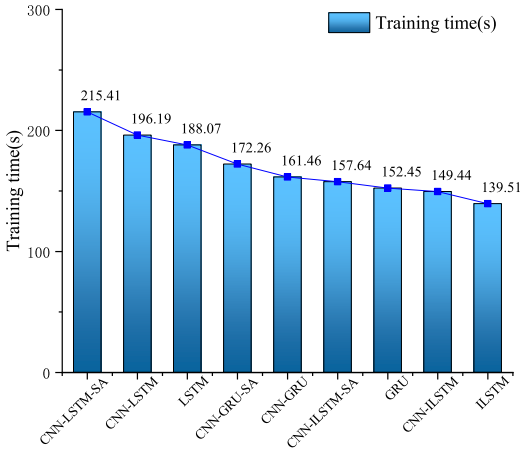

FIGURE 9. Training time comparison.

ILSTM, the introduction of CNN and SA improves prediction ability.

The reasons for improving the prediction accuracy of CNN-ILSTM-SA are as follows:

(1) The construction of a composite model makes up for the shortcomings of a single model in some aspects. CNN can well perform data feature extraction. SA can get the correlation between different input terms and predicted values, and assign different weight coefficients according to the correlation level to improve prediction accuracy.

(2) ILSTM introduces the 1-tanh function in the forget gate, which changes the output range of the forget gate to (0.24,1) so that the forget gate's output value is in a more obvious range. Thus, more input data characteristics are retained, and the model's learning ability is improved.

(3) In the improvement of the LSTM's input gate, c_{t-1} is added to the input gate algorithm, which is the cell state information of the previous step t-1. The introduction of c_{t-1} makes the model's input gate have a higher memory effect on the current data retention, and the introduction of ALSG improves the model's sensitivity to the input data, greatly reduces the degree of oversaturation, and makes the model learn more thoroughly.

ILSTM and CNN-ILSTM-SA have a great improvement in model training time because:

(1) The structure of the ILSTM model is simple. ILSTM consists of the input gate and the forget gate. Compared with LSTM, ILSTM removes the output gate.

(2) ILSTM has fewer parameters. Compared with LSTM, ILSTM reduces the weight parameters from eight to four and the bias parameters from four to two. Compared with GRU, ISLTM decreases the weight parameters from six to four and the bias parameters from three to two.

VI. CONCLUSION

This paper presents a prediction model for soybean meal futures price based on CNN-ILSTM-SA. Compared with eighteen baseline models, CNN-ILSTM-SA has the best comprehensive evaluation of prediction results. In this paper, ILSTM is presented for the first time. Under the premise of guaranteeing high prediction accuracy, ILSTM is improved and optimized, which alleviates the problems of ''gradient explosion'' and ''gradient disappearance'' caused by long-term data dependence in RNN. Compared with LSTM and GRU, ILSTM's training time is 25.8% and 8.49% shorter, respectively, and its prediction accuracy is the highest. In addition, the introduction of CNN compensates for the shortcomings of ILSTM feature extraction. The introduction of SA can obtain the correlation between different input terms and predicted values, and assign different weight coefficients according to the correlation level to improve the prediction accuracy. This paper's conclusions are as follows:

(1) The experiment results show that ILSTM has an excellent prediction performance. ILSTM is an improved model of LSTM, which removes LSTM's output gate and improves the forget gate and input gate. By introducing the 1-tanh function into the forget gate, and adding c_{t-1} and ALSG into the input gate algorithm, the prediction accuracy of ILSTM is improved. The prediction accuracy of ILSTM is higher than that of LSTM and GRU.

(2) The CNN-ILSTM-SA compensates for the shortcomings of the single prediction model, such as insufficient feature data extraction and insufficient learning of historical data. The prediction model for soybean meal futures price based on CNN-ILSTM-SA performs best in this experiment.

(3) The design and parameters of the model are improved and optimized. ILSTM has four weight parameters and two bias parameters, which are smaller than LSTM and GRU, and significantly improve CNN-ILSTM-SA's prediction speed.

CNN-ILSTM-SA is of great significance to the price prediction of China's soybean meal futures. It can help participants grasp the movement regularity of the soybean meal futures market in general, make reasonable decisions, and promote the steady and healthy development of the market. However, because weather and climate factors also impact soybean meal futures price, our next work considers processing weather and climate-related factors data into time series and using it as an input item of the prediction model to further improve the prediction accuracy.

REFERENCES

[1] Y. Wang, C. Lin, and S. Shih, ''The dynamic relationship between agricultural futures and agriculture index in China,'' *China Agricult. Econ. Rev.*, vol. 3, no. 3, pp. 369–382, Sep. 2011, doi: 10.1108/17561371111165798.

[2] B. Kanuparthi, D. Arpit, G. Kerg, N. R. Ke, I. Mitliagkas, and Y. Bengio, ''h-DETACH: Modifying the LSTM gradient towards better optimization,'' presented at the 7th ICLR Conf., New Orleans, LA, USA, May 2019.

[3] Y. Zhang and G. M. Tumibay, "Stock price prediction based on the Bi-GRU-attention model," *J. Comput. Commun.*, vol. 12, no. 4, pp. 72–85, Apr. 2024, doi: 10.4236/jcc.2024.124007.

[4] M. Sundermeyer, H. Ney, and R. Schlüter, "From feedforward to recurrent LSTM neural networks for language modeling," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 23, no. 3, pp. 517–529, Mar. 2015, doi: 10.1109/TASLP.2015.2400218.

[5] K.-S. Moon and H. Kim, "Performance of deep learning in prediction of stock market volatility," *Econ. Comput. Econ. Cybern. Stud. Res.*, vol. 53, no. 2/2019, pp. 77–92, Jun. 2019, doi: 10.24818/18423264/53.2.19.05.

[6] B. Wang, W. Kong, H. Guan, and N. N. Xiong, "Air quality forecasting based on gated recurrent long short term memory model in Internet of Things," *IEEE Access*, vol. 7, pp. 69524–69534, 2019, doi: 10.1109/ACCESS.2019.2917277.

[7] A. H. Bukhari, M. A. Z. Raja, M. Sulaiman, S. Islam, M. Shoaib, and P. Kumam, "Fractional neuro-sequential ARFIMA-LSTM for financial market forecasting," *IEEE Access*, vol. 8, pp. 71326–71338, 2020, doi: 10.1109/ACCESS.2020.2985763.

[8] M. Nabipour, P. Nayyeri, H. Jabani, A. Mosavi, E. Salwana, and S. Shahab, "Deep learning for stock market prediction," *Entropy*, vol. 22, no. 8, p. 840, Jul. 2020, doi: 10.3390/e22080840.

[9] H. Sun, D. Gui, B. Yan, Y. Liu, W. Liao, Y. Zhu, C. Lu, and N. Zhao, "Assessing the potential of random forest method for estimating solar radiation using air pollution index," *Energy Convers. Manage.*, vol. 119, pp. 121–129, Jul. 2016, doi: 10.1016/j.enconman.2016.04.051.

[10] W. Tong, H. Hong, H. Fang, Q. Xie, and R. Perkins, "Decision forest: Combining the predictions of multiple independent decision tree models," *J. Chem. Inf. Comput. Sci.*, vol. 43, no. 2, pp. 525–531, Mar. 2003, doi: 10.1021/ci020058s.

[11] G. Yang, H. Lee, and G. Lee, "A hybrid deep learning model to forecast particulate matter concentration levels in seoul, South Korea," *Atmosphere*, vol. 11, no. 4, Mar. 2020, Art. no. 348, doi: 10.3390/atmos11040348.

[12] Y.-S. Chang, H.-T. Chiao, S. Abimannan, Y.-P. Huang, Y.-T. Tsai, and K.-M. Lin, "An LSTM-based aggregated model for air pollution forecasting," *Atmos. Pollut. Res.*, vol. 11, no. 8, pp. 1451–1463, Aug. 2020, doi: 10.1016/j.apr.2020.05.015.

[13] L. Zhang, J. Wang, and B. Wang, "Energy market prediction with novel long short-term memory network: Case study of energy futures index volatility," *Energy*, vol. 211, Nov. 2020, Art. no. 118634, doi: 10.1016/j.energy.2020.118634.

[14] C. Wen, S. Liu, X. Yao, L. Peng, X. Li, Y. Hu, and T. Chi, "A novel spatiotemporal convolutional long short-term neural network for air pollution prediction," *Sci. Total Environ.*, vol. 654, pp. 1091–1099, Mar. 2019, doi: 10.1016/j.scitotenv.2018.11.086.

[15] J. Zhu, F. Deng, J. Zhao, and H. Zheng, "Attention-based parallel networks (APNet) for $PM_{2.5}$ spatiotemporal prediction," *Sci. Total Environ.*, vol. 769, May 2021, Art. no. 145082, doi: 10.1016/j.scitotenv.2021.145082.

[16] X. Shu, L. Zhang, Y. Sun, and J. Tang, "Host–parasite: Graph LSTM-in-LSTM for group activity recognition," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 2, pp. 663–674, Feb. 2021, doi: 10.1109/TNNLS.2020.2978942.

[17] P. Suebsombut, A. Sekhari, P. Sureephong, A. Belhi, and A. Bouras, "Field data forecasting using LSTM and bi-LSTM approaches," *Appl. Sci.*, vol. 11, no. 24, p. 11820, Dec. 2021, doi: 10.3390/app112411820.

[18] U. Bhimavarapu, "IRF-LSTM: Enhanced regularization function in LSTM to predict the rainfall," *Neural Comput. Appl.*, vol. 34, no. 22, pp. 20165–20177, Jul. 2022, doi: 10.1007/s00521-022-07577-8.

[19] L. Tian, L. Feng, L. Yang, and Y. Guo, "Stock price prediction based on LSTM and LightGBM hybrid model," *J. Supercomput.*, vol. 78, no. 9, pp. 11768–11793, Feb. 2022, doi: 10.1007/s11227-022-04326-5.

[20] X. Li, S. Wang, Z. Jiang, X. Li, B. You, and C. Rui, "Study on soil cracks in pneumatic subsoiling based on LSTM," *Soil Use Manage.*, vol. 39, no. 1, pp. 298–315, Jan. 2023, doi: 10.1111/sum.12859.

[21] F. Zhou, Z. Huang, and C. Zhang, "Carbon price forecasting based on CEEMDAN and LSTM," *Appl. Energy*, vol. 311, Apr. 2022, Art. no. 118601, doi: 10.1016/j.apenergy.2022.118601.

[22] M. Yu, T. Liu, Z. Guan, Y. Sun, J. Guo, L. Chen, and Y. He, "SALSTM: An improved LSTM algorithm for predicting the competitiveness of export products," *Int. J. Intell. Syst.*, vol. 37, no. 9, pp. 6185–6200, Feb. 2022, doi: 10.1002/int.22839.

[23] H. Liu, K. Wu, S. Fu, H. Shi, and H. Xu, "Predictive analysis of vehicular lane changes: An integrated LSTM approach," *Appl. Sci.*, vol. 13, no. 18, p. 10157, Sep. 2023, doi: 10.3390/app131810157.

[24] I. S. Kervanci, M. F. Akay, and E. Özceylan, "Bitcoin price prediction using LSTM, GRU and hybrid LSTM-GRU with Bayesian optimization, random search, and grid search for the next days," *J. Ind. Manage. Optim.*, vol. 20, no. 2, pp. 570–588, Jul. 2024, doi: 10.3934/jimo.2023091.

[25] S. Fang, Y.-S. Tan, T. Zhang, and Y. Liu, "Self-attention networks for code search," *Inf. Softw. Technol.*, vol. 134, Jun. 2021, Art. no. 106542, doi: 10.1016/j.infsof.2021.106542.

[26] G. Li, H. Liu, G. Li, S. Shen, and H. Tang, "LSTM-based argument recommendation for non-API methods," *Sci. China Inf. Sci.*, vol. 63, no. 9, Aug. 2020, Art. no. 190101, doi: 10.1007/s11432-019-2830-8.

[27] T. Ergen and S. S. Kozat, "Efficient online learning algorithms based on LSTM neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 8, pp. 3772–3783, Aug. 2018, doi: 10.1109/TNNLS.2017.2741598.

[28] J. Wang, X. Li, L. Jin, J. Li, Q. Sun, and H. Wang, "An air quality index prediction model based on CNN-ILSTM," *Sci. Rep.*, vol. 12, no. 1, May 2022, Art. no. 8373, doi: 10.1038/s41598-022-12355-6.

HAITAO LIU received the bachelor's degree in information management and systems from Hebei University of Science and Technology, in 2003, and the master's degree in management, in 2012. He is currently with the Information Construction and Management Center, Hebei University of Science and Technology. His main research interests include deep learning, management information systems, big data processing, and network security.

YANG PAN is currently pursuing the master's degree with Hebei University of Science and Technology. His research interests include machine learning and deep learning.

WEI LIU is currently pursuing the master's degree with Hebei University of Science and Technology. His research interests include machine learning and deep learning.

JINGYANG WANG received the B.Eng. degree in computer software from Lanzhou University, China, in 1995, and the M.Sc. degree in software engineering from Beijing University of Technology, China, in 2007. He is currently a Professor with Hebei University of Science and Technology, Shijiazhuang, Hebei, China. His research interests include machine learning, deep learning, natural language processing, and big data processing.