Research Article

Meteorological Satellite Operation Prediction Using a BiLSTM Deep Learning Model

Yi Peng ¹, Qi Han ¹, Fei Su ², Xingwei He ¹ and Xiaohu Feng ¹

¹National Satellite Meteorological Center, Beijing 100081, China

²China Unicom Smart Connection Technology Co., Ltd, Beijing 100037, China

Correspondence should be addressed to Xiaohu Feng; fengxh@cma.gov.cn

Received 13 March 2021; Accepted 31 May 2021; Published 21 June 2021

Academic Editor: Liguang Zhang

Copyright © 2021 Yi Peng et al. 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.

The current satellite management system mainly relies on manual work. If small faults cannot be found in time, it may cause systematic fault problems and then affect the accuracy of satellite data and the service quality of meteorological satellite. If the operation trend of satellite will be predicted, the fault can be avoided. However, the satellite system is complex, and the telemetry signal is unstable, nonlinear, and time-related. It is difficult to predict through a certain model. Based on these, this paper proposes a bidirectional long short-term memory (BiLSTM) deep learning model to predict the operation trend of meteorological satellite. In the method, the layer number of the model is designed to be two, and the prediction results, which are forecasted by LSTM network as the future trend data and historical data, are both taken as the input of BiLSTM model. The dataset for the research is generated and transmitted from Advanced Geostationary Radiation Imager (AGRI), which is the load of FY4A meteorological satellite. In order to demonstrate the superiority of the BiLSTM prediction model, it is compared with LSTM based on the same dataset in the experiment. The result shows that the BiLSTM method reports a state-of-the-art performance on satellite telemetry data.

1. Introduction

Meteorological satellite in a complex space environment that contains all kinds of loads is a complicated system, whose components have coupling relationship. Abnormality and faults easily occur. Small faults may cause systematic fault problems, then affect the accuracy of satellite data and service accuracy, and even cause significant economic losses and serious disasters to people's livelihood and the country. If the operation trend of satellite will be predicted, the fault can be avoided. It is of great significance for the intelligent health management of meteorological satellite and has reference value for the health management of other satellites.

At present, the trend prediction methods mainly include statistical prediction, mathematical prediction, intelligent prediction, and information fusion prediction. Among them, Auto Regression Moving Average (ARMA) prediction model, Support Vector Regression (SVR) prediction model, Back Propagation Network (BP) prediction model, and Long

Short-Term Memory (LSTM) prediction model are widely studied and applied. ARMA prediction model is a linear model with finite parameters. For short-term prediction, the model has high fitting accuracy [1, 2]. But it is not suitable for nonlinear and nonstationary sequences. SVR prediction model based on structural risk minimization criterion has better prediction generalization ability for small sample training set, higher prediction accuracy, and better robustness than ARMA model [3]. But it can only be used as short-term prediction algorithm. In recent years, SVR is widely used to solve practical problems in various fields by combining with other algorithms [4–6]. BP network has become one of the most widely used neural network models because of its strong nonlinear mapping ability and self-learning ability [7–9]. However, it has some problems, such as slow convergence speed, lack of scientific theoretical basis for the determination of hidden layer nodes, and easily falling into local minimum points [10]. Deep learning is a branch of neural networks. The recurrent neural network

(RNN) with depth and time series has been widely used as a prediction model for sequence data [11–13]. Although RNN can deal with time series problem, it has serious gradient dispersion problem. In order to weaken the adverse effects of gradient disappearance and long-distance dependence on neural network, the long-term memory ability of RNN is improved by replacing RNN chain unit with LSTM chain unit [14]. LSTM that uses additional memory cell to store states shows better ability for time series prediction than RNN. In recent years, LSTM has achieved a significant success in many fields [15–17].

Although LSTM model has good prediction ability of nonlinear time series [18], it takes historical time series data as the data input of the model, neglects the availability of future time series data, and lacks deep data mining. Moreover, the telemetry signal generated and transmitted from the satellite that reflects the operation of the satellite is unstable, nonlinear, and time-related. It is difficult to predict the satellite operation through a certain model. Bidirectional Long Short-Term Memory (BiLSTM) can process the sequence data in both forward and reverse directions and provide the past and future sequence information for each time in the sequence [19]. Therefore, BiLSTM mode is proposed to predict the operation trend of satellite in this paper. Increasing the number of depth layers of neural network can help enrich the feature set of network learning, increase the processing capacity of neural network, and improve the accuracy of the model. In addition, the paper presents using LSTM to predict the future trend and take the prediction results and historical data as the input of BiLSTM in order to provide the future time series data.

The remainder of this paper is organized as follows. In the next section, a brief description of problem definitions is presented, followed by the introduction of the proposed BiLSTM network architecture. Then, the application to meteorological satellite based on BiLSTM deep learning model and our experimental results are shown. Finally, conclusions are drawn.

2. Problem Definitions

Satellite management plays an important role for satellite safe. However, the current satellite management mainly relies on manual work, which takes time and effort. And these can lead to the fact that fault cannot be found in time. This paper intends using artificial intelligence algorithm to solve the current problems, by deeply mining and analyzing telemetry data, study the automatic prediction method for the operation trend of meteorological satellite, and solve the problems that restrict the stable operation of satellite and abnormalities, which cannot be found in time.

Based on the safety and reliability requirements of satellite, when it is on orbit, sensors are designed in the main functional modules of each key subsystem during satellite development. These sensors provide telemetry parameters data of satellite from the launch to in-orbit operation to the retirement. The satellite telemetry system collects the working conditions and values of various subsystems on the satellite according to a certain sampling

period and forms telemetry data after A/D transformation and coding. And then, it transmits them to the ground through modulation and amplification, and these telemetry data are gotten through the inverse process after receiving them on the ground. Satellite telemetry data are divided into two categories: digital quantity and analog quantity. The digital quantity reflects the functional state of the measured unit on the satellite. Analog is the numerical measurement value of the measured unit, which usually reflects the performance state of the measured unit. Satellite telemetry parameters include thermodynamic parameters, power system parameters, and dynamic parameters. It is no doubt that a large amount of data will be generated. These telemetry data reflect the state of satellite payload and the operation situation of satellite platform, which are stored in time series. Therefore, the operation trend for meteorological satellite is related to an amount of spatial-temporal data.

The telemetry data contain a lot of objective laws and knowledge that can be used for trend prediction. The operation trend of satellite platform and load can be predicted based on these heterogeneous, coupled, and large-scale telemetry data (Figure 1 shows some parameters). However, the research faces the following challenges:

- (1) Feature representation is difficult: there are thousands of satellite telemetry data variables; telemetry data have problems such as outliers and uneven time intervals in telemetry data; due to the coupling and correlation of satellites, it is difficult for a single parameter to describe the comprehensive performance, and to determine which parameters can accurately describe a certain performance.
- (2) It is difficult to predict the trend: the satellite system is complex, the telemetry signal is nonstationary and nonlinear, and the telemetry parameters have three different variation patterns: stationary, abrupt, and periodic.

Facing the difficult of feature representations, this paper selects key features according to the contribution of the operation trends of satellite load and the relationship of the features. Through the analysis of the traditional machine learning algorithm for time series learning prediction, the traditional algorithms are no longer suitable for the operation trend prediction of complex meteorological satellite. Therefore, this paper will build a new prediction model for satellite telemetry data.

The central problem in the model of predicting satellite operation can be represented in the following terms. The problem of satellite operation can be defined using the spatiotemporal variable sequence of X_{t-T} for prediction. The model can be denoted by

$$Y_{t+T} = f(X_t, X_{t-T}), \quad (1)$$

where Y_{t+T} is the predicted object in next T hours, f represents the final model learnt by the historical data, X_t denotes the datasets at the predicting moments, and X_{t-T} are the datasets in T hours before the predicting moments.

FIGURE 1: The sketch of complex satellite system.

3. The Proposed Scheme

3.1. The Introduction of LSTM Model. Compared with the traditional RNN, LSTM not only has a hidden state, but also adds a cell state. At the same time, it adds input gate, output gate, and forgetting gate in each layer to control the degree of adding or deleting information. The model can learn the long-term dependence information and avoid the problem of gradient disappearance [20, 21], with smaller error and higher prediction accuracy. At present, the most widely used LSTM network is to use the LSTM unit to replace the neural nodes in the hidden layer of RNN [22]. Its structure is shown in Figure 2, where c_{t-1} denotes the cell state of the previous moment, h_{t-1} is the output of the former LSTM, x_t and h_t represent the input and state output for the current moment respectively, and f_t , i_t and o_t represent the output value of the forgetting gate, the input gate, and the output gate, respectively. The LSTM unit update process is as follows:

$$\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c), \quad (2)$$

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \quad (3)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \quad (4)$$

$$c_t = f_t c_{t-1} + i_t \tilde{c}_t, \quad (5)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \quad (6)$$

$$h_t = o_t \tanh(c_t), \quad (7)$$

where $\tilde{c}_t$ and c_t represent the candidate cell status and the current cell state, respectively. W_c , W_f , W_i and W_o represent the weight of the candidate input gate, the forget gate, the input gate, and the output gate, respectively. b_c , b_f , b_i and b_o represent the bias of the candidate input gate, the forget gate, the input gate, and the output gate, respectively. σ and $\tanh$ represent sigmoid activation function and hyperbolic tangent activation function, respectively.

The training process of LSTM model adopts BPTT algorithm, which is similar to the classic back propagation (BP) algorithm. It can be roughly divided into four steps:

- (1) Calculating output value of LSTM cells according to forward calculation method (formulas (2)–(7))
- (2) Reversely computing error terms for each LSTM cell, which include two reverse propagation directions by time and network level
- (3) Calculating the gradient of each weight according to the corresponding error term
- (4) Updating weights based on the Optimization Algorithm of Gradient

There are many kinds of gradient-based optimization algorithm, such as Stochastic Gradient Descent (SGD) [23], Adaptive Gradient (AdaGrad) [24], Root Mean Square Prop (RMSProp) [25], and Adaptive Moment Estimation (Adam) [26]. Under the same number of iterations, the convergence speed of Adam algorithm is better than that of other algorithms. Furthermore, the loss function of Adam algorithm has the lowest value, and it costs the least time in all these stochastic optimization methods. Therefore, this paper chooses Adam algorithm as the optimizer, which performs

FIGURE 2: The basic of LSTM memory cell.

better in practice than other stochastic optimization methods based on gradient.

4. The Proposed BiLSTM Deep Learning Model

The situation of satellite operation is characterized by telemetry data transmitted from satellite, which has a certain spatial and temporal correlation. BiLSTM can not only solve the problem of long-term dependence between current moment and past moment, but also pay attention to the correlation between current moment and future moment. Based on this, the paper introduces BiLSTM to predict the operation trend of satellite. But, in practical application, we cannot get the future time series data, so that the paper presents using LSTM model to predict the future trend and takes the prediction results and historical data as the input of BiLSTM. Although BiLSTM is a deep neural network in time sequence, its network structure is extremely shallow in the number of layers. Increasing the number of depth layers of neural network can increase the processing capacity of neural network. The paper proposes stacking multiple BiLSTM and designing a deep bidirectional long short-term memory neural network. By stacking multiple hidden states, the deep BiLSTM is expanded to predict the operation trend of meteorological satellite. Trend prediction framework is shown in Figure 3.

The Deep BiLSTM network model contains multiple hidden layers, and the output of the former layer is used as the input of the latter layer, which helps mine the relationship between the front and back time series data, enrich the feature set of network learning, and improve the accuracy of the model. In the model, the transmission of network parameters in the input layer is the same as BiLSTM. But, in the out layer, Deep BiLSTM transmits the state to the hidden layer as input, and the state of multiple hidden layers is transmitted upward layer by layer, and finally to the output layer, as shown in Figure 4.

The historical time series data is input as the forward data of Deep BiLSTM network model, and the future time

series feature data predicted by LSTM is input as the backward data of Deep BiLSTM network model.

In BiLSTM network, the historical time series data $X = [x_1, x_2, \dots, x_T]$ is input to the forward network unit of BiLSTM, which gets the forward hidden layer state $\vec{h}_t$. Historical time series data $X = [x_1, x_2, \dots, x_T]$ as the input of LSTM network, is used to predict the future time series data $X = [x'_1, x'_2, \dots, x'_T]$. The predicted time series data $X = [x'_1, x'_2, \dots, x'_T]$ is input to the backward network unit of BiLSTM, which gets the backward hidden layer state $\overleftarrow{h}_t$. The output p_t is obtained by $\vec{h}_t$ and $\overleftarrow{h}_t$, which are expressed as follows:

$$\vec{h}_t = \sigma(W_{x_h} x_t + W_{h_h} \vec{h}_{t-1} + b_h), \quad (8)$$

$$\overleftarrow{h}_t = \sigma(W_{x_h} x_t + W_{h_h} \overleftarrow{h}_{t-1} + b_h), \quad (9)$$

where σ is the nonlinear activation function of the hidden layer, W_{x_h} is the weight from the input x of the current neuron to the hidden layer $\vec{h}_t$ at this moment, W_{h_h} is the weight from the state quantity at the previous moment to the current state quantity, $\vec{h}_{t-1}$ is the output value of the hidden layer at the previous moment, and b_h is the offset term.

The BiLSTM model is iterated in two directions at the same time and calculates the predicted value p_t by weighting the hidden layer state, which is expressed as

$$p_t = W_{h_p} \vec{h}_t + W_{h_p} \overleftarrow{h}_t + b_p. \quad (10)$$

Here, W_{h_p} is the weight from the hidden layer $\vec{h}_t$ to the output, W_{h_p} is the weight from the hidden layer $\overleftarrow{h}_t$ to the output, and b_p is the offset term.

FIGURE 3: Trend prediction framework for meteorological satellite.

FIGURE 4: The network structure of BiLSTM.

5. Experiments

5.1. Datasets. FY4A on orbit is the second generation of geostationary meteorological satellite. It has four loads, that

is, Advanced Geostationary Radiation Imager (AGRI), Geostationary Interferometric Infrared Sounder (GIIR), Lightning Mapping Imager (LMI), and Space Environment Monitoring Instrument Package (SEP). Among these, the operation trend of AGRI is used to evaluate the performance of the approach proposed in the paper, whose main function is to realize the quantitative observation of Earth's surface and clouds. The telemetry data of AGRI, which are generated and transmitted on certain days, are put into use for identifying the algorithm.

Among telemetry data, 26 observation variables most likely to be relevant to the operation trend, that is, 26 features, are selected. Therefore, each sample data for trend prediction consists of 26 variables, that is, $X = [TMC1, TMC2, \dots, TMC26]$, which are shown in Table 1. The dataset is generated at a sampling interval of 2 second. A total of 43189 data samples are collected as training samples and test samples. Figure 5 shows that the variables change with time from UTC 01:00 to UTC 01:59 on February 23, 2019. Figures 5(a)-5(d) are, respectively, the changing situation of variable $TMC3$, $TMC6$, $TMC11$, and $TMC19$ with time. From Figure 5, we can see that different variables can have different operation trends with time.

6. Evaluation Metric

In order to evaluate the prediction effect of BiLSTM deep learning model proposed by this paper, the Mean Absolute

TABLE 1: Variables of satellite telemetry.

Variable name	Variable symbol
AGRI scanning mechanism status	<i>TMC1</i>
AGRI EW angle	<i>TMC2</i>
AGRI NS angle	<i>TMC3</i>
AGRI 0.55 preamplifier +5 V	<i>TMC4</i>
AGRI 0.45 preamplifier +5 V	<i>TMC5</i>
AGRI 1.36 preamplifier +5 V	<i>TMC6</i>
AGRI 0.55 preamplifier bias voltage	<i>TMC7</i>
AGRI 0.45 preamplifier bias voltage	<i>TMC8</i>
AGRI 1.36 preamplifier bias voltage	<i>TMC9</i>
AGRI visible/near-infrared main amplifier +5 V	<i>TMC10</i>
AGRI visible/near-infrared main amplifier −5 V	<i>TMC11</i>
AGRI infrared main amplifier +12 V	<i>TMC12</i>
AGRI infrared main amplifier −12 V	<i>TMC13</i>
AGRI medium-long wave photoconduction preamplifier +12 V	<i>TMC14</i>
AGRI medium-long wave photoconduction preamplifier −12 V	<i>TMC15</i>
AGRI medium wave photovoltaic preamplifier +1.5 V	<i>TMC16</i>
AGRI medium wave photovoltaic preamplifier −1.5 V	<i>TMC17</i>
AGRI motor control +5 V	<i>TMC18</i>
AGRI motor drive +8 V	<i>TMC19</i>
AGRI temperature control +5 V	<i>TMC20</i>
AGRI excitation +5 V	<i>TMC21</i>
AGRI excitation −5 V	<i>TMC22</i>
AGRI induct synchronizer +15 V	<i>TMC23</i>
AGRI induct synchronizer −15 V	<i>TMC24</i>
AGRI simulation conditioning +5 V	<i>TMC25</i>
AGRI simulation conditioning −5 V	<i>TMC26</i>

(a)

(b)

FIGURE 5: Continued.

FIGURE 5: The variables changing with time. (a) TMC3 variable. (b) TMC6 variable. (c) TMC11 variable. (d) TMC19 variable.

Error (MAE) and Root Mean Square Error (RMSE) are selected as the criteria.

The MAE calculation formula is as follows:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|, \quad (11)$$

where $\hat{y}_i$ and y_i represent the predicted value and the actual value, respectively, and n is the number of samples. The smaller the MAE is, the more accurate the prediction is.

The RMSE calculation formula is as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}, \quad (12)$$

where n is the number of samples, $\hat{y}_i$ is the predicted value, and y_i is the actual value. The smaller the RMSE is, the more accurate the prediction is.

MAE and RMSE both can measure the error of the predictions. Their difference is that RMSE can penalize huge errors, but MAE cannot.

7. Results and Discussion

In order to prove the effectiveness of the BiLSTM deep learning model proposed by the paper, this method is compared with LSTM using the same training set and test set data under the same operating environment. All the experiments are carried out under the running environment of Inter(R) Core(TM) i7- 1.99 GHz, 8.00 GB RAM, and Windows 10. The development environment is Python 3.6 along with TensorFlow 2.0.0.

For the sake of reducing the dimension difference of characteristics and its possible impact, all data are normalized. After data preprocessing, roulette method is used to divide data sets, and 70% of the data set is used as training set, 20% as verification set, and the remaining 10% as test set.

Because the setting of experimental parameters has a great influence on the experimental results, the method of fixed parameters is used in the experiment. Multiple experiments are done when the dropout is 0.1, 0.2, ..., 0.9, respectively. Among them, when dropout value is 0.5, the RMSE is the smallest; that is, the prediction accuracy is the highest, as shown in Figure 6. By comparing the results of many experiments, it is found that the prediction effect of BiLSTM deep learning model proposed is the best when the parameters in Table 2 are taken. The epoch is 100, the optimizer chooses Adam, batch size is 64, training steps are 200, and learning rate is 0.0001.

Considering that the depth of neural network can improve the model representation, the number of BiLSTM layers is set to two. The BiLSTM layer number is determined by evaluating the loss value and the prediction of different BiLSTM layers, which are, respectively, shown in Figure 7 and Table 3. Figure 7 illustrates that the loss value of two layers is smaller than that of one layer. In Table 3, the RMSE and MAE of two layers are 0.0231 and 0.0188, respectively, which are smaller than those of one layer. These mean that the two layers are superior to one layer. Thus, the layer number of the BiLSTM model proposed by the paper is chosen as two.

In the experiment, infrared main amplifier -12 V, medium wave photovoltaic preamplifier +1.5 V, medium wave photovoltaic preamplifier -1.5 V, motor control +5 V, motor drive +8 V, temperature control +5 V, excitation +5 V, excitation -5 V, induct synchronizer +15 V, induct synchronizer -15 V, simulation conditioning +5 V, and simulation conditioning -5 V that sign the operation situation of meteorological satellite are predicted, which are labeled as R1, R2... R12, respectively. The training set data is used to train BiLSTM deep learning model, LSTM model, and RNN model. The verification set data is applied to adjust parameter of the model. And the model achieved by training is used to predict the test set data. The evaluation error indexes of each method can be calculated on the base of the predicted value of each method and the real value. The comparison results of the three

FIGURE 6: Dropout curve.

TABLE 2: Parameter settings of BiLSTM model.

Parameter	Value	Parameter	Value
Epoch	100	Batch size	64
Optimizer	Adam	Learning rate	0.0001
Training steps	200	Dropout	0.5

FIGURE 7: Loss value for different number of BiLSTM layers.

TABLE 3: Prediction error for different number of BiLSTM layers.

	One BiLSTM layer	Two BiLSTM layers
RMSE	0.3276	0.0231
MAE	0.3259	0.0188

methods are shown in Table 4. We can see that the RMSE of BiLSTM is smaller than LSTM and RNN from Table 4. This

TABLE 4: Various error indicators for prediction using different model.

	RMSE (BiLSTM)	RMSE (LSTM)	RMSE (RNN)
R1	0.02213	0.25671	0.69161
R2	1.14932	3.65821	2.07490
R3	1.13257	2.98576	2.22344
R4	0.03683	0.67384	1.19673
R5	0.03285	0.43784	2.14687
R6	0.06222	0.98673	0.63541
R7	0.39643	1.25473	0.39750
R8	0.02880	0.36528	1.69788
R9	0.01315	0.19327	0.21264
R10	1.13424	3.06785	1.99287
R11	0.03279	0.42356	1.07320
R12	0.02310	0.27856	0.65139

means that the BiLSTM deep learning model proposed in this paper can better predict the operation trend of meteorological satellite than LSTM and RNN models.

8. Conclusions

In order to avoid the occurrence of major satellite fault and provide the complete and accuracy satellite data, this paper proposes meteorological satellite operation prediction method. Considering the unstable, nonlinear, and temporal characteristics of satellite telemetry data, a BiLSTM deep learning model for satellite operation trend prediction is put forward. The method uses multidimensional telemetry variables as the input, and LSTM is used to predict the future trend, and then, we take the prediction results and historical data as the input of BiLSTM deep learning model. Increasing the number of depth layers of neural network can enrich the feature set of network learning, increase the processing capacity of neural network, and improve the accuracy of the model. The research on the data sets transmitted from AGRI load of FY4A satellite shows that the BiLSTM deep learning model proposed by the paper has highest prediction accuracy and better performance than the classical LSTM model.

Further research work will be undertaken to improve the current method to make the results more accurate based on more dataset. Further work will also research that the method can be applied with real-time prediction for satellite, rather than just focusing on existing historical data set.

Data Availability

The telemetry data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This research was supported by the National Key Research and Development Program of China (No.2018YEC1507803).

References

- [1] A. Al-Smadi, "The estimation of the order of an ARMA process using third-order statistics," *International Journal of Systems Science*, vol. 36, no. 15, pp. 975–982, 2005.
- [2] A. Kizilkaya and A. H. Kayran, "ARMA model parameter estimation based on the equivalent MA approach," *Digital Signal Processing*, vol. 16, no. 6, pp. 670–681, 2006.
- [3] F. Yuan, U. Kumar, and D. Galar, "Reliability prediction using support vector regression," *International Journal of System Assurance Engineering and Management*, vol. 263, 2010.
- [4] S. Safarzadegan Gilan, H. Bahrami Jovein, and A. A. Ramezani-pour, "Hybrid support vector regression - particle swarm optimization for prediction of compressive strength and RCPT of concretes containing metakaolin," *Construction and Building Materials*, vol. 34, pp. 321–329, 2012.
- [5] K. F. Fung, Y. F. Huang, and C. H. Koo, "Coupling fuzzy-SVR and boosting-SVR models with wavelet decomposition for meteorological drought prediction," *Environment Earth Science*, vol. 78, 2019.
- [6] W. Gao and J. Han, "Prediction of destroyed floor depth based on principal component analysis (PCA)-Genetic algorithm (GA)-Support vector regression (SVR)," *Geotechnical and Geological Engineering*, vol. 38, no. 4, pp. 3481–3491, 2020.
- [7] G. Hao, "Study on prediction of urbanization level based on GA-BP neural network," in *Proceedings of the 21st International Conference on Industrial Engineering and Engineering Management*, pp. 1021–1026, Singapore, 2019.
- [8] J. D. Watson and F. H. C. Crick, "BP neural network-based product quality risk prediction," *Advances in Intelligent Systems and Computing*, vol. 1117, pp. 1021–1026, 2019.
- [9] F. He and L. Zhang, "Mold breakout prediction in slab continuous casting based on combined method of GA-BP neural network and logic rules," *The International Journal of Advanced Manufacturing Technology*, vol. 95, pp. 4081–4089, 2018.
- [10] C. Yuan, D. Niu, C. Li, L. Sun, and L. Xu, "Electricity consumption prediction model based on bayesian regularized BP neural network," *Advances in Intelligent Systems and Computing*, vol. 928, pp. 528–535, 2020.
- [11] N. R. Metu and T. Sasikala, "Prediction analysis on web traffic data using time series modeling, RNN and ensembling techniques," *Lecture Notes on Data Engineering and Communications Technologies*, vol. 26, pp. 611–618, 2018.
- [12] C. Lin and M. Chi, "A comparisons of BKT, RNN and LSTM for learning gain prediction," *Lecture Notes in Computer Science*, vol. 10331, pp. 536–539, 2017.
- [13] Q. Wu, K. Ding, and B. Huang, "Approach for fault prognosis using recurrent neural network," *Journal of Intelligent Manufacturing*, vol. 31, no. 7, pp. 1621–1633, 2020.
- [14] N. K. Manaswi, "RNN and LSTM," in *Deep Learning with Applications Using Python*, Apress, Berkeley, CA, USA, 2018.
- [15] S. Poornima and M. Pushpalatha, "Drought prediction based on SPI and SPEI with varying timescales using LSTM recurrent neural network," *Soft Computing*, vol. 23, no. 18, pp. 8399–8412, 2019.
- [16] F. Mtibaa, K.-K. Nguyen, M. Azam, A. Papachristou, J.-S. Venne, and M. Cheriet, "LSTM-based indoor air temperature prediction framework for HVAC systems in smart buildings," *Neural Computing and Applications*, vol. 32, no. 23, pp. 17569–17585, 2020.
- [17] J. Zhao, J. J. Wu, X. W. Guo et al., "Prediction of radar sea clutter based on LSTM," *Journal of Ambient Intelligence and Humanized Computing*, 2019.
- [18] V. Gundu and S. P. Simon, "PSO-LSTM for short term forecast of heterogeneous time series electricity price signals," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 2, pp. 2375–2385, 2021.
- [19] S. N. Sima, T. Neda, and S. N. Akbar, "The performance of LSTM and BiLSTM in forecasting time series," in *Proceedings of the IEEE International Conference on Big Data*, Los Angeles, CA, USA, 2019.
- [20] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [21] W. J. Lu, J. Z. Li, J. Y. Wang et al., "A CNN-BiLSTM-AM method for stock price prediction," *Neural Computing and Applications*, vol. 33, pp. 4741–4753, 2020.
- [22] A. Graves, "Long short-term memory," *Studies in Computational Intelligence*, vol. 385, pp. 37–45, 2012.
- [23] S.-I. Amari, "Backpropagation and stochastic gradient descent method," *Neurocomputing*, vol. 5, no. 4-5, pp. 185–196, 1993.
- [24] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," *Journal of Machine Learning Research*, vol. 12, no. 7, pp. 257–269, 2011.
- [25] S. Yeung, O. Russakovsky, N. Andriluka, G. Mori, and L. Fei-Fei, "Every moment counts: dense detailed labeling of actions in complex videos," *International Journal of Computer Vision*, vol. 126, no. 2-4, pp. 375–389, 2018.
- [26] D. P. Kingma and L. J. Ba, "Adam: a method for stochastic optimization," in *Proceedings of the International Conference On Learning Representations*, (ICLR), San Diego, CA, USA, 2015.