IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 10 April 2024, accepted 18 June 2024, date of publication 21 June 2024, date of current version 28 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3417533

RESEARCH ARTICLE

Temperature Forecasting of Grain in Storage: An Improved Approach Based on Broad Learning Network

QIFU WANG[1,2], MINGLEI HOU[1,2,3], YAO QIN[1,3], AND FEIYU LIAN[1,3]

[1]College of Information Science and Engineering, Henan University of Technology, Zhengzhou 450001, China
[2]Institute of Applied Physics, Henan Academy of Sciences, Zhengzhou 450001, China
[3]Key Laboratory of Grain Information Processing and Control, Ministry of Education, Henan University of Technology, Zhengzhou 450001, China

Corresponding author: Feiyu Lian (lfywork@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 62373136, in part by the Special Project for Key Research and Development of Henan Province under Grant 221111230300, in part by the Special Project for Scientific Research and Development of Henan Academy of Sciences under Grant 230607048, in part by the Science and Technology Opening Cooperation Project of Henan Academy of Sciences under Grant 220907016, in part by the Open Fund of the Key Laboratory of Grain Information Processing and Control under Grant KFJJ-2021-103, in part by the Key Scientific Research Project Program of Universities of Henan Province under Grant 22A510014, and in part by the Innovative Funds Plan of Henan University of Technology under Grant 2022ZKCJ02.

ABSTRACT Temperature forecasting of grain in storage is crucial for timely granary temperature control, mitigating adverse effects of extreme temperatures on grain quality. Traditional machine learning methods struggle with stability and high error rates in grain storage temperature forecasting, while deep learning models are more accurate but time-consuming and have heavy parameters. To address these problems, an improved model with light weight and good accuracy is proposed in this paper, which broad learning network is combined with one-dimensional convolution module and multi-head self-attention mechanism (BLN-1DCNN-MHSA). Firstly, we employ a one-dimensional convolution module at the feature nodes of the model to extract local temporal correlations, compensating for temporal sequence learning limitations of the BLN. Secondly, a multi-head self-attention mechanism at the enhancement nodes to captures important features dependencies and global temporal correlations. Lastly, our model achieves better prediction through enhanced representation ability of model nodes. The results with real grain storage temperature data demonstrate that the RMSE, MAPE, and MAE of the proposed model are 0.341, 0.54%, 0.28, respectively, which represent more than 2 times improvement in accuracy compared to the BLN, and it also reduces training time by more than 90% compared with LSTM and Transformer models. Additionally, the generalization and robustness of the improved approach are demonstrated through promising results in a classification experiment on the MNIST dataset. In general, the model provides a certain feasibility for early warning of grain storage risks by predicting its temperature trends.

INDEX TERMS Grain storage temperature forecasting, grain storage security, broad learning network, multi-head self-attention, convolutional neural network.

I. INTRODUCTION

With the increase in world population and the development of animal husbandry, our demand for grain is increasing. However, every year, a significant proportion of stored grain is lost due to mildew and insect pests, primarily caused by inappropriate storage practices [1]. It is estimated that approximately 1-2% of stored grain is lost due to poor management of grain storage facilities [2], with temperature fluctuations being a key factor influencing mildew in stored grain [3], [4], [5], [6]. Therefore, it's necessary to develop an accurate and rapid grain temperature forecasting method to

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

guide grain temperature control in advance, such as mechanical ventilation or grain cooling to ensure the safety of grain storage.

For the grain storage temperature prediction model, the current research mainly includes mathematical methods and machine learning methods. By using mathematical methods to conduct mathematical modeling of heat and moisture transfer in stored grain, and obtaining temperature field information by solving equations [7]. Because the factors that affect grain storage temperature have complexity and randomness, being unable to carry on effective mathematical modeling or modeling is very difficult. In practice, the inefficiency and inaccuracy of mathematical methods limit their applicability in grain storage temperature prediction. Machine learning methods focus on analyzing the inherent laws in data rather than the coupled processes of heat and humidity. Although traditional machine learning methods can better represent the nonlinear relationship between data than modeling, it is difficult to effectively process large amounts of complex grain situation data because the time correlation of stored grain temperature data is not considered [8].

The development of deep learning has led to the proposal of various neural networks, enabling the consideration of both time and nonlinear data in grain storage temperature prediction [9]. For example, deep belief network (DBN), convolutional neural network (CNN), and recurrent neural network (RNN) represented by long short-term memory (LSTM) network, and so on. [10], [11], [12]. However, these deep models are much more complex than traditional machine learning, and all have long training times and large numbers of parameters.

To cope with the training difficulties of deep learning, especially for small samples, broad learning has appeared. It is a shallow network structured by the random vector functional-link neural network (RVFLNN) [13], which realizes regression and classification by expanding the learned features to the width, and has fast training and good learning ability [14], [15], [16], [17], [18]. Broad learning can be an alternative to deep learning in some cases.

In this paper, an improved approach (BLN-1DCNN-MHSA) is proposed, namely broad learning network (BLN) is combined with one-dimensional convolution (1DCNN) and multi-head self-attention mechanism (MHSA). This method with low-cost consumption and rapid forecasting can be used to further improve the accuracy of grain storage temperature forecasting, so to guide the temperature control of granary staff in advance and reduce the risk of grain storage. The main contributions of our work are as follows:

(1) The proposed grain storage temperature forecasting model combines the MHSA at the enhancement nodes of BLN to process the enhancement nodes of non-stationary and nonlinear stored grain temperature data, which effectively improves the accuracy and robustness of prediction and reduces the error.

(2) The introduction of 1DCNN in the BLN feature nodes fully leverages the advantages of convolution for feature extraction, enhances the extraction of time-dependent features, and strengthens the interconnections between features. This further improves the predicting accuracy of the model for grain storage temperature.

(3) Our model training is more convenient and faster, which provides a strong basis for timely and effective grain storage temperature prediction. At the same time, the model has generalization and can classify MNIST image data. Compared with the BLN, the classification accuracy is also improved, but because it is not time series data, the Timing factor extraction operation is not used at the feature nodes.

The remainder of this paper is organized as follows. Section II reviews the related literature on grain storage temperature forecasting. Section III introduces the proposed model and related knowledge. Section IV describes the grain storage temperature dataset as well as the evaluation method. Section V presents the experimental results and analysis of the proposed model. Finally, the conclusion is given in Section VI.

II. RELATED WORK

Under closed grain storage conditions, the micro-airflow inside the stored-grain will cause temperature diffusion, and the magnitude of the temperature is not only affected by the temperature inside and outside the granary, but also the influence of the temperature outside the granary usually lags one to two months. Therefore, accurate prediction of grain storage temperature is a relatively difficult problem to be studied.

The field of grain temperature prediction, mathematical methods were initially employed. Rocha et al. [19] proposed a three-dimensional heat transfer model (i.e., conservation equations for heat, mass, and momentum) in a flat maize bin. Based on the above model, Quemada et al. [20] provided the flow of relative humidity and temperature in silos through the simulation of temperature changes and moisture changes in grain piles. The two-dimensional mathematical model of barley temperature in cylindrical silos was simulated by Francisco [21]. The standard error between the predicted barley germination temperature and the measured barley temperature ranged from 6.6% to 12.4%, depending on the grain depth. Due to the limitations of mathematical modeling and simulation of various data in grain warehouses, traditional machine learning has been developed in grain situation forecasting, such as linear regression (LR) [22], support vector regression (SVR) [23], random forest (RF) [24], grey model (GM) [25], etc. Wang et al. [26] predicted the temperature of the grain storage pile through the least square method model, which was effective and feasible with an average root-mean-square error of 1.9. By forecasting the temperature of the grain storage pile with different SVR models, the obtained best root-mean-square error is 1.26 [27]. However, the forecasting accuracy of the above method is not high and it tends to make short-term forecasting.

Nowadays, deep learning models are becoming popular. LSTM network has the characteristics of long-term memory information, can avoid gradient disappearance or explosion problems, and is specially designed for sequence prediction problems. And many variants have emerged to improve prediction performance, such as Bidirectional LSTM, CNN-LSTM, stacked-LSTM, and so on. [28]. The Gated Recurrent Unit (GRU) is simplified vision of the internal structure of the LSTM, and has fewer training parameters and faster speed, but there is no obvious performance gap between the two in practice [29]. Hence, the use of LSTM in granary temperature prediction, indoor temperature prediction, and other fields is on the rise. Wu et al. [30] using Bidirectional LSTM forecast grain stack temperature with an error of 0.42, the prediction accuracy was improved compared with the LSTM error of 1.49. Compared with a fully connected network, CNN needs to learn fewer parameters. Furkan et al. used the CNN-LSTM architecture to combine the excellent feature extraction of convolutional layers with the ability of LSTM to learn sequential dependencies, which solved the limitations of a single model and effectively improved the accuracy of indoor temperature prediction by nearly half [31].

Here, the Transformer model is considered superior to LSTM, it has completely abandoned the traditional cycle of the neural network structure, adopted the structure of the attention mechanism [32], speeding up the development progress in the field of deep learning [33]. However, a Transformer model alone does not appear to be very effective for handling time series prediction tasks [34]. Therefore, other scholars usually adopt combination models. Sharma et al. [35] adopted a combined deep learning model for feature extraction of nonlinear grain temperature data. Duan et al. [36] proposed an encoder-decoder model with spatiotemporal attention mechanism, which has achieved good results in forecasting the stored grain temperature with RMSE of 0.28 by considering the relationship between grain temperature and external meteorological factors. Qu et al. [37] used graph Convolutional neural Network (GCN) and Transformer to solve the grain temperature prediction considering the spatial topology of sensor networks with RMSE of 0.25.

However, these methods are either overly time-consuming for prediction or lack the desired accuracy, which motivates us to develop new approaches. Hence, this paper introduces an improved approach (BLN-1DCNN-MHSA) based on the BLN, offering a rapid prediction method with high accuracy for forecasting storage grain temperature.

III. METHODOLOGY

Population growth brings greater demand for grain. To ensure grain demand, it is necessary to rely on storage in addition to increasing arable land. However, for grain storage, it is inevitable to ensure both safety and low cost. For temperature forecasting, our method not only has good prediction accuracy but also does not need expensive GPU for accelerated training, so it meets the demand well. This section introduces the basic background theory related to BLN-1DCNN-MHSA model, in order to understand the proposed model in this paper.

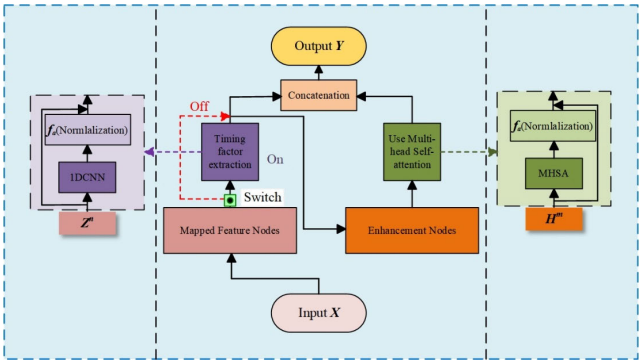


FIGURE 1. The structure diagram of the proposed BLN-1DCNN-MHSA model.

A. OVERALL MODEL STRUCTURE

In this subsection, we introduce BLN-1DCNN-MHSA the overall structure of the network model, as shown in Fig. 1.

As depicted in Fig. 1, the main structure of the model is composed of BLN, mainly feature mapping and enhancement mapping, corresponding to two feature groups (Z and H). Firstly, after the input data X undergoes the feature mapping, then 1DCNN module was performed according to the time sliding window composed of simple exponential smoothing to extract feature vectors to enhance the local temporal correlation for the feature nodes Z, and a residual connection following normalization is used. Next, the core of the proposed model was to introduce the MHSA at the enhancement transformation of the enhancement nodes H. This includes normalizing the global temporal correlation features and applying a nonlinear transformation through the activation function to adaptively enhance the extraction of effective features from the enhancement nodes, followed by a residual connection. Lastly, the outputs of the two (referring to the new Z and the new H) were concatenated, and the connection weights between it and the output Y were obtained by ridge regression. The proposed model also has the characteristics of broad expansion of features and fast training. Moreover, with the help of two improvement modules, the learning ability of the model for time series and the representation ability of features are strengthened, so the prediction accuracy of the model is improved.

Based on the proposed prediction model, the test set is predicted independently, yielding the final predictions for storage grain temperature using the model. At the same time, good results are obtained in the classification problem on the MNIST data set (the Timing factor extraction operation is turned off after mapped features). Moreover, this proposed model capitalizes on the structural benefits of the BLN rapid training and can be applied to both regression and classification tasks.

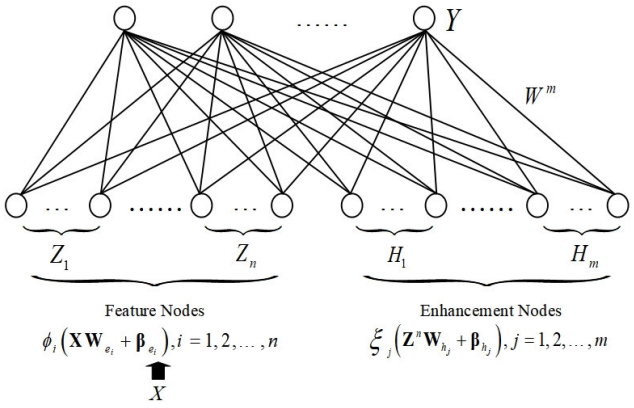


FIGURE 2. The illustration of BLN.

B. BLN BASIC MODULE

In this paper, the main model we used for grain storage temperature forecasting is broad learning network (BLN) is inspired by the random vector functional-link neural network (RVFLNN) [13], [38]. Due to its distinctive broad network structure instead of deep structure to learn features, and the use of ridge regression generalized inverse to directly calculate the output connection matrix, the BLN can predict quickly and has a small number of parameters. Our model takes full advantage of the BLN that the time-non-consuming training process. The general structure of the BLN is illustrated in Fig. 2, comprising a three-layer structure: the input layer, hidden layer, and output layer. The hidden layer comprises both feature nodes and enhancement nodes. The input X is mapped to the feature nodes Z by the mapping process. Subsequently, the feature nodes are enhanced to generate the enhancement nodes H, where the feature nodes and the enhancement nodes are extracted and nonlinear transformed respectively. Moreover, the connection weights between it and the output Y were obtained by a rapid training method.

In this structure, the input vector is represented as X and the output is Y. $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N]^T \in \mathbb{R}^{N\times D}$, it has N samples with D dimensions, and the output matrix $\boldsymbol{Y} \in \mathbb{R}^{N\times d}$. The feature and the enhancement nodes of the BLN can be obtained by

$$\boldsymbol{Z}_i = \phi_i(\boldsymbol{X}\boldsymbol{W}_{e_i} + \boldsymbol{\beta}_{e_i}), i = 1, 2, \ldots, n \tag{1}$$

$$\boldsymbol{H}_j = \xi_j\left(\boldsymbol{Z}^n\boldsymbol{W}_{h_j} + \boldsymbol{\beta}_{h_j}\right), j = 1, 2, \ldots, m \tag{2}$$

where ϕ_i can be a linear or nonlinear activation function, ξ_j is nonlinear activation function, $\boldsymbol{W}_{ei}$ and $\boldsymbol{W}_{hj}$, $\boldsymbol{\beta}_{ei}$ and $\boldsymbol{\beta}_{hj}$ are the randomly generated weights and biases, respectively, which are kept constant during training, and denote n and m as the number of feature and enhancement nodes. All feature nodes $[\boldsymbol{Z}_1, \boldsymbol{Z}_2, \ldots, \boldsymbol{Z}_n]$ denoted by $\boldsymbol{Z}^n$ and all enhancement nodes $[\boldsymbol{H}_1, \boldsymbol{H}_2, \ldots, \boldsymbol{H}_m]$ denoted by $\boldsymbol{H}^m$.

The rapid training method that only needs to train weights matrix of the BLN to reduce the training errors, which greatly reduces the training amount of the model. The rapid training method with the following formula:

$$\arg\min_{\boldsymbol{W}^m} :\| \boldsymbol{A}\boldsymbol{W}^m - \hat{\boldsymbol{Y}} \|_2^2 + \lambda \| \boldsymbol{W}^m \|_2^2 \tag{3}$$

where the combination matrix of the feature nodes and the enhancement nodes is denoted as $\boldsymbol{A} = [\boldsymbol{Z}^n \mid \boldsymbol{H}^m]$. $\hat{\boldsymbol{Y}}$ is the true value, $\boldsymbol{W}^m$ is the output weights of the BLN, and $\boldsymbol{W}^m = [\boldsymbol{Z}^n \mid \boldsymbol{H}^m]^+\boldsymbol{Y} = \boldsymbol{A}^+\boldsymbol{Y}$, using ridge regression can be computed as follows:

$$\boldsymbol{W}^m = \left(\lambda I + \boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T\boldsymbol{Y} \tag{4}$$

where is λ a regularization coefficient with constraints, I is a unit matrix. Pseudo-inverse is a very convenient method to solve the weights of the output layer, and $\boldsymbol{A}^+$ is the pseudo-inverse of $\boldsymbol{A}$. According to (4), the $\boldsymbol{A}^+$ can be calculated, which is:

$$\boldsymbol{A}^+ = \lim_{\lambda\to 0}(\lambda I + \boldsymbol{A}\boldsymbol{A}^T)^{-1}\boldsymbol{A}^T \tag{5}$$

Thus, the BLN can be represented as follows:

$$\begin{aligned}\boldsymbol{Y} &= [\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n \mid \xi_1(\boldsymbol{Z}^n\boldsymbol{W}_{h_1} + \boldsymbol{\beta}_{h_1}), \ldots, \xi_m(\boldsymbol{Z}^n\boldsymbol{W}_{h_m} \\ &\quad + \boldsymbol{\beta}_{h_m})]\boldsymbol{W}^m \\ &= [\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n \mid \boldsymbol{H}_1, \ldots, \boldsymbol{H}_m]\boldsymbol{W}^m \\ &= [\boldsymbol{Z}^n \mid \boldsymbol{H}^m]\boldsymbol{W}^m \\ &= \boldsymbol{A}\boldsymbol{W}^m\end{aligned} \tag{6}$$

The pseudocode of Algorithm1 is shown as below:

Algorithm 1 Pseudocode of the BLN algorithm

Input: training samples X;
  for $i = 0; i \leq n$
  Initialize $\boldsymbol{W}_{e_i}$, $\boldsymbol{\beta}_{e_i}$ randomly;
  Calculate $\boldsymbol{Z}_i = \phi_i(\mathbf{X}\boldsymbol{W}_{e_i} + \boldsymbol{\beta}_{e_i})$;
end
Set the feature nodes $\boldsymbol{Z}^n = [\boldsymbol{Z}_1, \boldsymbol{Z}_2, \ldots, \boldsymbol{Z}_n]$;
  for $j = 0; j \leq m$
  Initialize $\boldsymbol{W}_{h_j}$, $\boldsymbol{\beta}_{h_j}$ randomly;
  Calculate $\boldsymbol{H}_j = \xi_j\left(\boldsymbol{Z}^n\boldsymbol{W}_{h_j} + \boldsymbol{\beta}_{h_j}\right)$;
end
Set the enhancement nodes $\boldsymbol{H}^m = [\boldsymbol{H}_1, \boldsymbol{H}_2, \cdots, \boldsymbol{H}_m]$;
  Get $\boldsymbol{A}$ by concatenating the $\boldsymbol{Z}^n$ and the $\boldsymbol{H}^m$;
  Calculate $\boldsymbol{W}^m$ by ridge regression;
Output: $\boldsymbol{W}^m$;
Test testing samples by the $\boldsymbol{W}^m$.

C. TIMING FACTOR EXTRACTION MODULE

Convolutional neural network (CNN) is frequently used in both research and application because of its efficient feature extraction capability, enabling convolution operations for various data types and extract the features of data [31], [39], [41]. Since the BLN mainly learns features horizontally, and the learning of the temporal relationship between features is not sufficient, this study one-dimensional convolution is selected to extract the local temporal correlation of features at feature

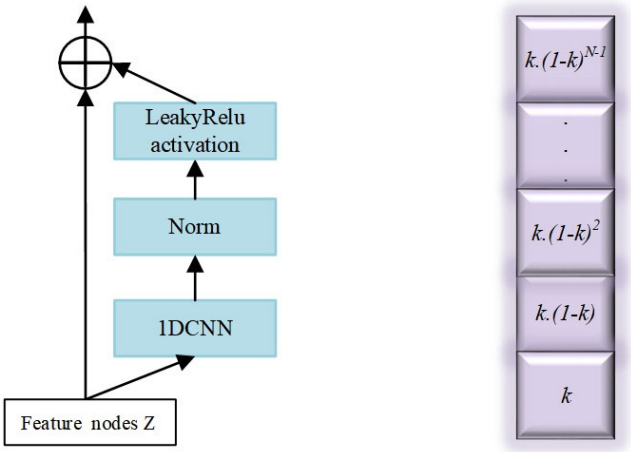


FIGURE 3. Structure of timing factor extraction module (left) and 1DCNN convolution kernel (right).

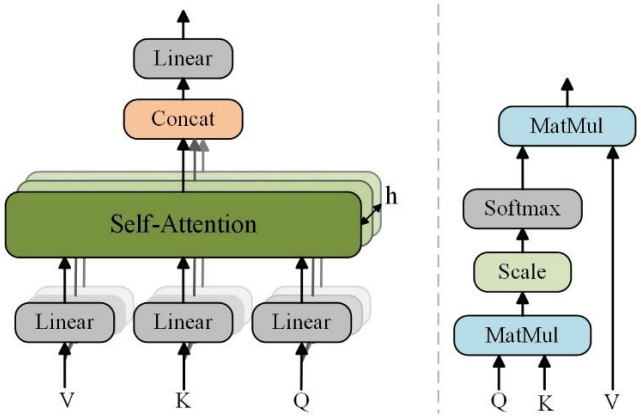


FIGURE 4. Multi-head self-attention (left), and attention mechanism (right).

nodes of the BLN, which can enhance the predicting of the time series of grain storage temperature, and it does not take the form of stacked convolutional layers. It is called the Timing factor extraction module because it enhances the learning of the time factor aspect of the network. The pseudocode of Timing factor extraction module (denoted as1DCNN for ease of expression) is shown Algorithm 2. We employed the exponential weight of Simple Exponential Smoothing (SES) [43] as the convolution kernel of the 1DCNN in this model. The module constructed by convolution, normalization, leaky relu activation and residual connection, it effectively improves the representation ability of the trend of temperature series, as depicted in Fig. 3.

Algorithm 2 Pseudocode of the Timing factor extraction module(1DCNN)

Input: training samples X;
  for $i = 0; i \leq n$
  Initialize $\boldsymbol{W}_{e_i}$, $\boldsymbol{\beta}_{e_i}$ randomly;
  Calculate $\boldsymbol{Z}_i = \phi_i(\mathbf{X}\boldsymbol{W}_{e_i} + \boldsymbol{\beta}_{e_i})$;
end
Set the feature mapping group $\boldsymbol{Z}^n = [\boldsymbol{Z}_1, \boldsymbol{Z}_2, \ldots, \boldsymbol{Z}_n]$;
  Calculate the convolutional feature mapping group $\boldsymbol{Z}_c^n$ by 1DCNN convolution kernel;
  The $\boldsymbol{Z}_c^n$ for normalization and Leaky Relu activation;
  Calculate *new* $\boldsymbol{Z}^n$ by residual connection for $\boldsymbol{Z}^n$ and $\boldsymbol{Z}_c^n$;
Output: *new* $\boldsymbol{Z}^n$;

SES intensifies the influence of near observations on the predicted value and assigns varying weights to observations at different times, with the weights decreasing in a proportional manner. As follows:

$$\hat{y}_t = k \cdot y_{t-1} + k \cdot (1-k) \cdot y_{t-2} + \ldots + k \cdot (1-k)^{N-1} y_{t-N} \tag{7}$$

where k ($0< k<1$) is a smoothing constant, (1-k) is the common ratio (of a geometric series), N is the number of observations to consider, and is the predicted value. Within the observed values, the weight changes exponentially from the near to the distant, with the largest weight assigned to the nearest value, and strengthen the influence of temperature on the forecast in nearer days.

D. MULTI-HEAD SELF-HEAD SELF-ATTENTION MECHANISM

The enhancement nodes represent significant features resulting from the nonlinear transformation in BLN. Therefore, utilizing the Multi-Head Self-Attention (MHSA) mechanism on the enhancement nodes can lead to mine global correlations between them. The pseudocode of the using MHSA module is shown Algorithm 3.

Algorithm 3 Pseudocode of the Multi-Head Self-Attention (MHSA)

Input: the feature nodes $\boldsymbol{Z}^n$ or *new* $\boldsymbol{Z}^n$;
  for $j= 0; j \leq m$
  Initialize $\boldsymbol{W}_{h_j}$, $\boldsymbol{\beta}_{h_j}$ randomly;
  Calculate $\boldsymbol{H}_j = \xi_j\left(\boldsymbol{Z}^n\boldsymbol{W}_{h_j} + \boldsymbol{\beta}_{h_j}\right)$;
end
  Set the enhancement nodes $\boldsymbol{H}^m = [\boldsymbol{H}_1, \cdots, \boldsymbol{H}_m]$;
  Calculate the $\boldsymbol{H}_c^m$ by using Multi_Head(Q, K, V);
  The $\boldsymbol{H}_c^m$ for normalization,
  then, for nonlinear transformation by f_a(activation function);
  Calculate *new* $\boldsymbol{H}^m$ by residual connection for $\boldsymbol{H}^m$ and $\boldsymbol{H}_c^m$;
Output: *new* $\boldsymbol{H}^m$;

The attention mechanism associates different positions of a single input sequence and then calculates the features of this sequence [32]. It can pay more attention to the key information in the input for accurate prediction [33]. For notation, denote each input multiplied by a set of weights for the key as K, by a set of weights for the query as Q, and by a set of weights for the value as V. The attention

mechanism then learns how to map the query Q to a sequence of key-value pairs (K, V). The importance of the extracted V values is reflected by the similarity calculated from Q and K, namely the weights, which are then weighted and summed to obtain the attention value. Nevertheless, the particular of the self-attention mechanism compared to the attention mechanism is that Q = K = V:

$$\text{Attention(Q,K,V)} = \text{softmax}\left(\frac{\text{QK}^{\text{T}}}{\sqrt{\text{d}_{\text{K}}}}\right)\text{V} \tag{8}$$

The MHSA can solve the defect of the self-attention that excessively focus on its own position. The MHSA mechanism can be understood as that the self-attention operation is performed on multiple heads simultaneously, h(the number of heads) different (Q, K, V) are obtained, each self-attention value is calculated, and the results are concatenated for projection, as can be seen from Fig. 4. This can be expressed in formula (10):

$$\text{head}_i = \text{Self-Attention}\left(QW_i^{\text{Q}}, KW_i^k, VW_i^v\right) \tag{9}$$

$$\text{Multi\_Head}(Q, K, V) = \text{Concat}\left(\text{head}_1 \ldots, \text{head}_h\right) W^o \tag{10}$$

where the parameter matrices $W_i^{\text{Q}} \in \mathbb{R}^{d\times d_q}$, $W_i^k \in \mathbb{R}^{d\times d_k}$, $W_i^v \in \mathbb{R}^{d\times d_v}$, and $W^o \in \mathbb{R}^{hd_v\times d}$, $dq = d_k = d_v = d/h$, and dq, d_k, and d_v denote the dimensions of Q, K, and V, respectively. The number of dimensions can be thought of as the number of nodes, and Concat(•) is a concatenation operation that puts the individual heads together.

The MHSA mechanism can extract more important information. After normalizing the output generated by the MHSA mechanism at the f_a, the model in this paper also considers the activation function for nonlinear transformation. In this way, the network has a better fitting ability is a common means to improve the prediction results.

IV. DATA SETS AND ESTIMATION METHODS

The experimental evaluation of the above-mentioned models in the granary temperature forecasting and MNIST is carried out using the two datasets described below.

A. GRANARY TEMPERATURE DATA

For the temperature data of the experiment analysis, the granary temperature data from May 1, 2015 to September 30, 2015 of a national storage in Yiwu city were selected as the original data in vector format (four groups of the highest grain pile temperature data, one group of the average grain pile temperature data, and one group of the temperature data inside the warehouse, and one group of the temperature data outside the warehouse, and the highest grain pile temperature data corresponding to the grain pile was output). The granary temperature data was sampled once a day, and the adjacent point interpolation method was used to solve the temperature data due to abnormal or missed sensor detection.

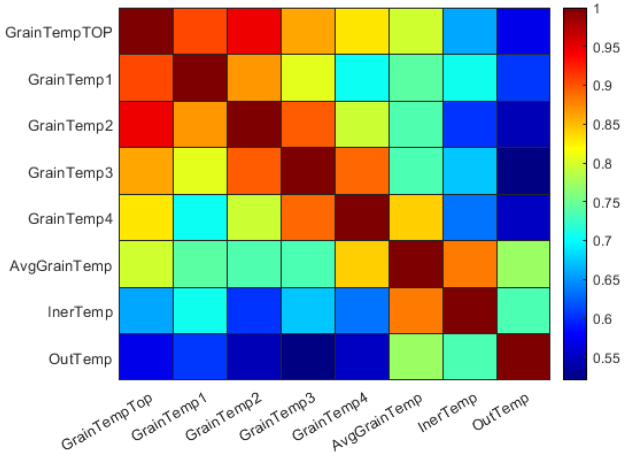


FIGURE 5. Correlation matrix showing the correlation among different selected temperature samples.

Pearson correlation coefficient was performed on the data to determine the interaction degree between different continuous temperature variables. As shown in the Fig. 5, the correlation coefficients among different selected temperature samples. According to the following equation:

$$r = \frac{\sum_{t=1}^{k}\left(x_t - \bar{x}\right)\left(y_t - \bar{y}\right)}{\sqrt{\sum_{t=1}^{k}\left(x_t - \bar{x}\right)^2}\sqrt{\sum_{t=1}^{k}\left(y_t - \bar{y}\right)^2}} \tag{11}$$

where x_t and y_t represent the temperature samples at time t, respectively; $\bar{x}$ and $\bar{y}$ are the mean values of the temperature samples, respectively. The numerical size of r values obtained by Pearson analysis can provide reference for the selection of data samples. The correlation analysis of grain temperature data shows that there is a strong interaction between the selected temperature samples, and the overall temperature presents a dynamic change trend.

In order to avoid the negative impact of the singular sample data on the prediction accuracy and to make the model converge quickly, the MNIST data set, and the grain storage temperature data set were normalized to the range [0,1] before the experiment. The min-max normalization method is adopted as follows:

$$\hat{y}_i = \frac{y_i - y_{min}}{y_{max} - y_{min}} \tag{12}$$

where $\hat{y}_i$ is the normalized result, y_i is the data to be normalized at a certain time, y_{max} and y_{min} are the maximum and minimum values in the data to be normalized, respectively.

B. PERFORMANCE ESTIMATION

In this study, the commonly used evaluation methods, mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean square error (RMSE) are used to evaluate the prediction quality of the model for grain storage temperature prediction problem. The evaluation metrics are

defined as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | y_{act}(i) - y_{pred}(i)| \tag{13}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} | \frac{y_{act}(i) - y_{pred}(i)}{y_{act}(i)}| * 100\% \tag{14}$$

$$RMSE = \sqrt{\frac{1}{n}\sum\nolimits_{i=1}^{n} \left(y_{act}(i) - y_{pred}(i)\right)^2} \tag{15}$$

where n represents the number of test samples, $y_{act}(i)$ and $y_{pred}(i)$ are the actual and predicted values at time i, respectively. MAE is a common way to calculate the error to represent the prediction accuracy. MAPE considers the relative gap between the actual and predicted values. RMSE evaluates accuracy by comparing the deviation between the actual and predicted values. For the above three evaluation metrics MAPE, MAE, and RMSE, the smaller the evaluation metrics value is, the more accurate the prediction result is.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} * 100\% \tag{16}$$

where TP (True positive), TN (True negative), FP (False positive), and FN (False negative) are part of the confusion matrix. The test set of MNIST is relatively uniform and can be regarded as a balanced data set. Accuracy used in the evaluation is a commonly used evaluation metrics in multi-classification problems, and the larger the value of the metrics, the better.

To compare the robustness of some models in the literature, a simplified robustness evaluation method based on random sampling is used to evaluate the robustness of the model to the influence of input perturbation. The robustness index F_m is defined as

$$F_m = 1 - \frac{\Delta f_m}{\Delta f} \tag{17}$$

where Δf_m is the difference between the maximum and minimum values of the loss function f sampled in the robust space of model m; Δf is the difference between the maximum and minimum value of the loss function f sampled in the global space. The larger the F_m is, the better the model m is resistant to input disturbance.

V. EXPERIMENTAL RESULTS AND ANALYSIS

This section, we conduct experiments and result analysis related to the proposed model.

In this study, device configuration used for model training and testing: Windows 10, 64-bit operating system, Intel (R) core (TM) i7-10700 CPU, 64GB RAM, NVIDIA Tesla T4 graphics card, Python 3.10 and pytorch2.0 operating environment. A Long Short-Term Memory (LSTM) network [29] is suitable for the problem of predicting time series. A Convolutional Neural Network-Long Short-Term Memory (CNN-LSTM) network [31] to combine the prominent feature extraction of CNN layers with the LSTM's capability of learning sequential dependencies. A Transformer [33] has

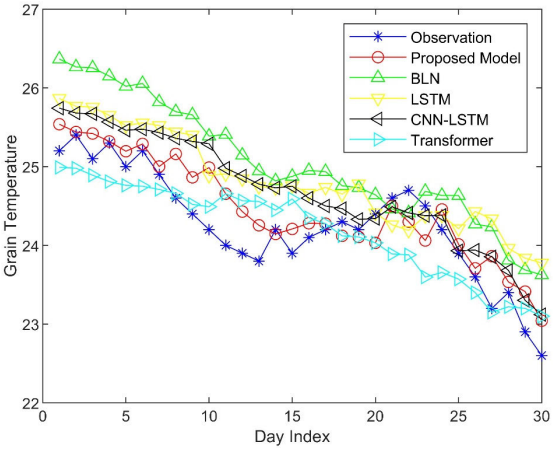


FIGURE 6. Comparison prediction diagram for experiment I.

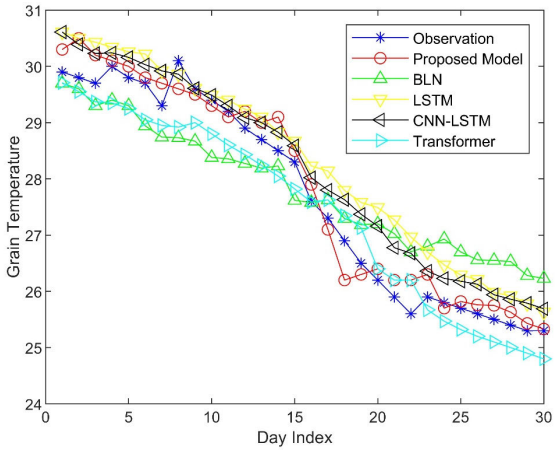


FIGURE 7. Comparison prediction diagram for experiment II.

an attention mechanism that enables it to learn to focus on important parts of the time series to accurate predicting. For LSTM, CNN-LSTM and Transformer on the number of hidden layers is set to 2, 2, 4, the number of hidden nodes and learning rate is set to 128 and 0.001 respectively, and set the dropout parameter to 0.1. The Adam optimizer is used, the loss function is the root mean square error function, and the cross-entropy loss function is used for classification, while the number of self-attention heads in the Transformer is 8, and the maximum number of epochs of the above models is 100.

A. GRANARY TEMPERATURE EXPERIMENT

The granary temperature data from June 1, 2015 to September 30, 2015 and May 1, 2015 to August 30, 2015 were selected as the original data, and were divided into training set and test set according to the ratio of 3:1, so that two independent experiments (denoted as I and II) were conducted to verify the accuracy of the forecasting. The regularization coefficient λ is [-1,1]. The proposed model chooses the tanh activation function after using the MHSA mechanism, the dropout is 1e-10, the number of training batches is 3, the convolution

TABLE 1. The comparison forecasting results for experiment I.

Evaluation index	BLN	Proposed Model	LSTM	CNN-LSTM	Transformer
RMSE	0.879	**0.366**	0.679	0.592	0.495
MAPE (%)	3.21	**0.71**	2.24	1.77	1.22
MAE	0.81	**0.29**	0.61	0.48	0.41
Training time (s)	1.91	2.83	64.22	85.85	71.34

TABLE 2. The comparison forecasting results for experiment II.

Evaluation index	BLN	Proposed Model	LSTM	CNN-LSTM	Transformer
RMSE	0.813	**0.341**	0.689	0.561	0.445
MAPE (%)	2.11	**0.54**	2.06	1.81	1.05
MAE	0.74	**0.28**	0.60	0.46	0.39
Training time (s)	2.01	2.88	61.53	84.43	72.51

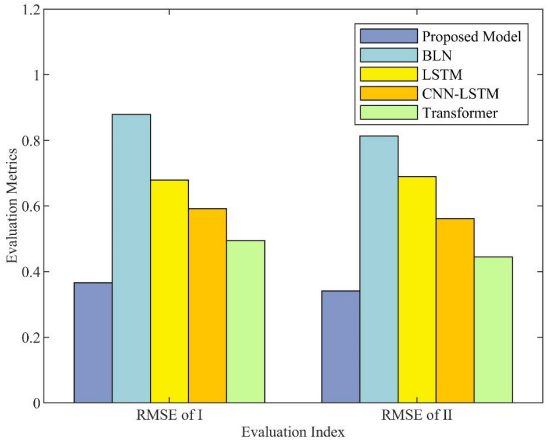


FIGURE 8. The comparison of RMSE in experiment I and experiment II.

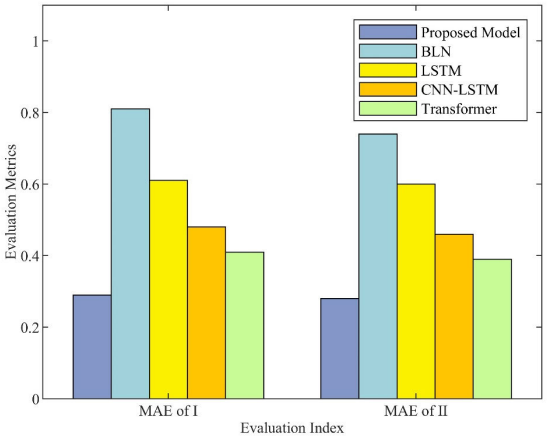


FIGURE 10. The comparison of MAE in experiment I and experiment II.

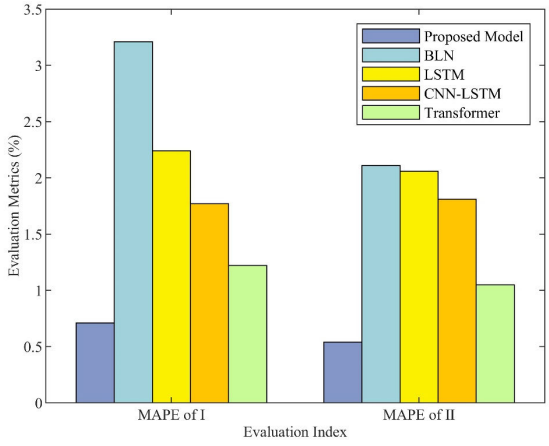


FIGURE 9. The comparison of MAPE in experiment I and experiment II.

kernel size is 1×5, and the smoothing constant is 0.5. Furthermore, to reduce unnecessary computational costs, the number of enhancement nodes above 1000 is not considered.

The grain storage temperature predicting results of experiments I and II are shown in Fig. 6 and Fig. 7. As can be seen from Table 1, the RMSE, MAPE, and MAE of the proposed model in experiment I is 0.366, 0.71%, and 0.29, respectively. In the BLN, LSTM, CNN-LSTM, and Transformer, the proposed model in this paper has the smallest error, and its training time of 2.83s is second only to the BLN.

From Table 2 we observed that in experiment II, the RMSE, MAPE, and MAE of proposed model are 0.341, 0.54%, and 0.28 respectively, which are also the smallest compared with other models, and its training time of 2.88s is second only to BLN as well. The BLN model has the lowest prediction accuracy but the fastest training speed among the single models. However, the LSTM model can remember the past interactions of the data due to its unique periodic unit. Therefore, it has better prediction accuracy RMSE (0.689), MAPE (2.06%), and MAE (0.60), but the training time is more than 20 times that of the BLN. Transformer relies on self-attention mechanism to capture long-term dependence and interaction of data. It has better prediction RMSE (0.445), MAPE (1.05%), and MAE (0.39), but also increases the training time by about 20% compared to the LSTM. For the CNN-LSTM model, the prediction accuracy is better than the LSTM due to the use of convolution to extract important features, but the training time is also increased. In a certain time period (Day Index), other models predict fairly well, but for grain temperature prediction, we take the overall forecasting effect as the standard. Experiment II is a little better than experiment I, which is also caused by the uncertainty of the grain storage temperature data, but the error is within the acceptable range.

TABLE 3. Comparison of our model with recent studies in forecasting temperature of grain.

Models	Precision	Complexity	Robustness	Training difficulty
In [27]	Low	Medium	Average	Moderate
In [36, 37]	High	High	Good	Difficult
In [44]	Medium	Medium	Good	Moderate
Our model	High	Medium	Excellent	Moderate

Moreover, the two experimental results are the best on the evaluation metrics. The evaluation metrics of RMSE, MAPE, and MAE for experiments I and II can also be observed more intuitively from Figs. 8, 9, and 10, and it can be found that the proposed model is the best among these models. It is very valuable in the field of grain storage temperature prediction that the proposed model takes less than half the time and improves more than twice the performance metrics compared to the fastest BLN. Although the proposed model loses a small part of its speed advantage over the BLN, it still saves about 90% of the training time compared with the LSTM model. In short, other models cannot surpass the proposed model by considering the temporal correlation and interaction of features for the BLN.

A summary comparison of recent studies is presented in Table 3. In [21], through meteorological factors, an adaptive boosting (AdaBoost) combined with principal components analysis method was used to predict grain pile temperature, and the best RMSE was 1.26. However, the RMSE of this method is more than 3 times that of our method because this model performance is limited and the correlation between the temperature in the grain pile is not considered. In [30] and [31], the improved deep learning model uses a large amount of data to achieve good prediction results, but the model training speed is slow and the complexity is relatively large. In [39], a prediction model based on broad learning network was proposed to predict the risk point of stored-grain situation. Benefiting from the processing of multi-modal features of stored-grain situation data, good risk point prediction results are obtained, but its accuracy still needs to be improved in the predicting only temperature of stored-grain. Therefore, our model in this paper still has application space.

In this part, experiment I is conducted on the proposed model with varying numbers of MHSA heads (briefly referred to as Heads) and enhancement nodes (referred to as Nodes) to assess the impact of the selection of the above parameters. Table 4 shows that the three accuracy indicators of RMSE, MAPE and MAE have achieved good forecasting results of grain storage temperature under different Heads and Nodes, and head 2 and node 500 are the best in this comparative experiments, RMSE (0.366), MAPE (0.71%), and MAE (0.29). It is also observed that merely increasing the number of Heads and Nodes does not necessarily yield better results. However, the effectiveness of the proposed model is still obvious compared to the LSTM. In this experiment, it is shown that the two parameters can be flexibly selected, and the prediction results are relatively acceptable in the grain temperature prediction scenario.

TABLE 4. Comparison of the proposed model with different numbers of Heads & Nodes in experiment I.

Heads & Nodes	RMSE	MAPE (%)	MAE	Training time (s)
2 & 300	0.416	0.91	0.39	**2.62**
2 & 400	0.395	0.84	0.35	2.75
2 & 500	**0.366**	**0.71**	**0.29**	2.83
2 & 600	0.382	0.82	0.33	2.92
2 & 800	0.391	0.83	0.33	2.99
4 & 300	0.396	0.82	0.32	2.88
4 & 400	0.413	0.92	0.38	2.95
4 & 500	0.429	0.98	0.41	3.06
4 & 600	0.398	0.82	0.34	3.16
4 & 800	0.391	0.81	0.32	3.22
6 & 300	0.441	1.02	0.39	3.03
6 & 420	0.432	0.96	0.38	3.14
6 & 540	0.416	0.92	0.37	3.22
6 & 660	0.423	0.94	0.37	3.33
6 & 840	0.402	0.88	0.36	3.41

B. ABLATION EXPERIMENT OF TEMPERATURE PREDICTING

In this subsection, in order to verify the effectiveness of each improvement point of the proposed model, the ablation experiment is carried out on the grain storage temperature forecasting of experiments I and II in the previous subsection.

This experiment is intended to reflect the specific effects of the two modules added to BLN individually and together. Similarly, two independent time periods of grain storage temperature forecasting were carried out to ensure the robustness of the experiment. Firstly, as shown in Table 5, the improvement effect of BLN + 1DCNN by considering the time factor is about 25% compared with the BLN, the results boosted by BLN + 1DCNN are not prominent. However, the addition of the MHSA mechanism BLN + MHSA predicted results show that the MHSA mechanism has played a key role, for experiment I and experiment II, RMSE decreased by 0.467 and 0.418, MAPE decreased by 2.29% and 1.25%, MAE decreased by 0.44 and 0.40, respectively. Next, the 1DCNN module is introduced at the feature nodes, BLN + 1DCNN + MHSA, which is the proposed model in this paper. Compared with only adding the MHSA mechanism RMSE (i.e., BLN + MHSA) is reduced by 0.046 and 0.054 respectively, and MAPE is reduced by 0.21% and 0.32% respectively. The MAE is reduced by 0.08 and 0.06 respectively, and the training time increased by the proposed model is less than 1s compared with the fastest BLN. Moreover, the proposed model takes less than half the time and improves more than twice the performance metrics compared to the BLN. So, the effectiveness of the proposed model is further demonstrated by the results of ablation experiments.

C. EXPERIMENT ON MODEL GENERALIZATION

In this paper, besides the analysis of granary temperature data, we will also conduct classification experiment to ensure

TABLE 5. Ablation experiment of experiment I and II.

Experiment	Evaluation index	BLN	BLN+1DCNN	BLN+MHSA	BLN+1DCNN+MHSA
I	RMSE	0.879	0.613	0.412	0.366
I	MAPE (%)	3.21	1.82	0.92	0.71
I	MAE	0.81	0.55	0.37	0.29
I	Training time (s)	1.91	2.14	2.62	2.83
II	RMSE	0.813	0.598	0.395	0.341
II	MAPE (%)	2.11	1.79	0.86	0.54
II	MAE	0.74	0.51	0.34	0.28
II	Training time (s)	2.01	2.25	2.66	2.88

TABLE 6. Experiment on MNIST when the enhancement nodes are 11000.

Models	Feature nodes	Enhancement nodes	Accuracy (%)	Training time (s)	Params (M)	Fm	FPS
BLN	100	11000	98.74	83.64	1.1	0.967	501.5
Proposed Model	100	11000	98.98	282.57	1.8	0.969	482.8
DBN	-	-	98.87	53129.77	19.5	0.955	53.4
CDBN	-	-	99.12	82169.24	28.6	0.962	45.6
VGG16	-	-	98.91	68243.56	140.4	0.958	82.5

TABLE 7. Experiment on MNIST when the enhancement nodes are 500.

Models	Accuracy (%)	Training time (s)	Params (M)	FPS
BLN	96.25	3.52	0.5	512.3
Proposed Model	96.47	5.29	1.2	489.6
CNN	95.63	21793.93	62.6	120.4
VGG11	96.28	49375.28	133.5	95.8

the effectiveness and generalization of the proposed model. In this subsection, we mainly conduct comparative experiments on classical MNIST handwritten digit images with the BLN, Convolutional Deep Belief Network (CDBN) [40], CNN [41], and VGG [42].

The classic MNIST dataset is 70,000 handwritten digits, each with a grayscale pixel of 28 × 28 fixed-size image. Because MNIST is not a time series data, in the classification problem on the MNIST (the Timing factor extraction is operation is turned off after mapped features). The models use 7-fold cross validation, and the MNIST dataset divided into 60000 samples of training data set and test set of 10000 samples, which is beneficial to verify the robustness. One set of feature nodes was set to 100 and enhancement nodes were set to 11000, the other set of feature nodes was set to 100 and enhancement nodes were set to 500 (to verify the effectiveness of the model for a small number of enhancement nodes), and the number of MHSA heads was 10 and 2, respectively. The proposed model selects a hard-swish activation function after using the MHSA mechanism with dropout of 1e-10 and training batch number of 15. To save unnecessary calculation costs, more than ten MHSA heads are not considered.

As can be seen from Tables 6 and 7 above, that the number of parameters of the model proposed in this paper is still very small after improvement. Although the training time is increased compared with the BLN in the two experiments, the accuracy of the test classification is significantly improved, and it has a clear advantage over the training time, parameters, robustness, and FPS of other models. At the same time, when the number of enhancement nodes (500) is reduced, the accuracy is also improved by 0.22% compared with the BLN, the training time is greatly reduced compared with the number of enhancement nodes (11000) and all metrics are better than CNN and VGG11. Only the CDBN rely on deep structure to be more accurate than our model, but its other four metrics are hard to accept, which leads to the loss of timeliness and lightness. From this experiment, although the proposed model in this paper is not the best in each index, it is optimal in general comparison, especially the extremely fast training speed compared to these deep learning models. In conclusion, the proposed model has good generalization.

VI. CONCLUSION AND FUTURE WORK

The temperature forecasting of stored grain has important practical significance for the safety storage of grain. In the previous section, the validity of the proposed model was verified by the forecasting of grain storage temperature. BLN, LSTM, CNN-LSTM, and Transformer models were used as comparative models to conduct two independent periods of grain storage temperature prediction. Because traditional machine learning is more suitable for short-term forecasting, we did not experiment with it. Among the single models, the BLN model has the lowest prediction accuracy, whereas the LSTM model has better prediction accuracy due to its unique periodic unit, which enables the LSTM to remember past interactions and can deal with both time series and nonlinear problems. For the CNN-LSTM combination model, the prediction results are better than the two single models of BLN and LSTM. Transformer relies on the self-attention mechanism to capture the long-term dependence and interaction of data, to have good prediction accuracy, but it also increases a certain training time. Considering that grain storage temperature forecasting is susceptible to various factors, we focus on the data correlation and investigate its underlying patterns. The introduction of the multi-head self-attention mechanism plays a pivotal role in enhancing the accuracy of grain storage temperature prediction.

Therefore, this paper introduces a BLN-1DCNN-MHSA grain storage temperature prediction approach, the improvement of the model overcame the single model cannot effectively mine discontinuous data between the defect of hidden information, successful access to the inner relation of the temperature data improves the prediction accuracy effectively. At the same time, it retains the advantages of overcoming the limitations of deep learning methods that the time-consuming training process because of huge amounts of connecting parameters in layers and filters. Compared with the other four prediction models in the experiment, its accuracy is improved, and its operation time is very small. Although the proposed model sacrifices some computational time compared to BLN, it is worthwhile that it achieved the best prediction results. Through rapid forecasting and good forecasting results, reasonable granary temperature control can be carried out when there is a risk of stored grain. Therefore, the proposed BLN-1DCNN-MHSA has good practicability in forecasting the grain storage temperature, and can provide guidance for the temperature control of the granary temperature that will be in medium or high risk (the risk levels may be different for different regions and different granaries) through rapid and accurate forecasting.

However, there is still a gap between the proposed model and some optimized deep learning models in the accuracy metrics of grain storage temperature forecasting, and the factors to be considered need to be increased. In the future, other factors such as granary humidity and pests will be considered to optimize the model prediction for the risk conditions of stored grain. In the meantime, the prediction model should be further optimized, and the time-attention mechanism is planned to be introduced to make its prediction accuracy higher, reduce the difference caused by different granary temperatures in the prediction, and improve the practicability of the model. To achieve a more comprehensive forecasting of the grain storage risk situation and provide more rational control guidance. Finally, provides a more favorable guarantee for future grain security storage.

ACKNOWLEDGMENT

The authors would like to thank all the researchers who helped to improve this article.

REFERENCES

[1] C. B. Singh and J. M. Fielke, "Recent developments in stored grain sensors, monitoring and management technology," *IEEE Instrum. Meas. Mag.*, vol. 20, no. 3, pp. 32–55, Jun. 2017.

[2] D. S. Jayas, "Storing grains for food security and sustainability," *Agricult. Res.*, vol. 1, no. 1, pp. 21–24, Mar. 2012.

[3] C.-C. Jia, D.-W. Sun, and C.-W. Cao, "Mathematical simulation of temperature fields in a stored grain bin due to internal heat generation," *J. Food Eng.*, vol. 43, no. 4, pp. 227–233, Mar. 2000.

[4] F. Jian, D. S. Jayas, N. D. G. White, and K. Alagusundaram, "A three-dimensional, asymmetric, and transient model to predict grain temperatures in grain storage bins," *Trans. ASAE*, vol. 48, no. 1, pp. 263–271, 2005.

[5] D. Mohapatra, S. Kumar, N. Kotwaliwale, and K. K. Singh, "Critical factors responsible for fungi growth in stored food grains and non-chemical approaches for their control," *Ind. Crops Products*, vol. 108, pp. 162–182, Dec. 2017.

[6] G. R. Thorpe, "The application of computational fluid dynamics codes to simulate heat and moisture transfer in stored grains," *J. Stored Products Res.*, vol. 44, no. 1, pp. 21–31, Jan. 2008.

[7] F. Fleurat-Lessard, "Integrated management of the risks of stored grain spoilage by seedborne fungi and contamination by storage mould mycotoxins—An update," *J. Stored Products Res.*, vol. 71, pp. 22–40, Mar. 2017.

[8] F. Rodrigues and F. C. Pereira, "Beyond expectation: Deep joint mean and quantile regression for spatiotemporal problems," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 12, pp. 5377–5389, Dec. 2020.

[9] Z. Chen, M. Ma, T. Li, H. Wang, and C. Li, "Long sequence time-series forecasting with deep learning: A survey," *Inf. Fusion*, vol. 97, Sep. 2023, Art. no. 101819.

[10] T. van Klompenburg, A. Kassahun, and C. Catal, "Crop yield prediction using machine learning: A systematic literature review," *Comput. Electron. Agricult.*, vol. 177, Oct. 2020, Art. no. 105709.

[11] M. Abdar, F. Pourpanah, S. Hussain, D. Rezazadegan, L. Liu, M. Ghavamzadeh, P. Fieguth, X. Cao, A. Khosravi, U. R. Acharya, V. Makarenkov, and S. Nahavandi, "A review of uncertainty quantification in deep learning: Techniques, applications and challenges," *Inf. Fusion*, vol. 76, pp. 243–297, Dec. 2021.

[12] D. Fister, J. Pérez-Aracil, C. Peláez-Rodríguez, J. D. Ser, and S. Salcedo-Sanz, "Accurate long-term air temperature prediction with machine learning models and data reduction techniques," *Appl. Soft Comput.*, vol. 136, Mar. 2023, Art. no. 110118.

[13] Y.-H. Pao, G.-H. Park, and D. J. Sobajic, "Learning and generalization characteristics of the random vector functional-link net," *Neurocomputing*, vol. 6, no. 2, pp. 163–180, Apr. 1994.

[14] Q. Zhou and X. He, "Broad learning model based on enhanced features learning," *IEEE Access*, vol. 7, pp. 42536–42550, 2019.

[15] C. L. P. Chen, Z. Liu, and S. Feng, "Universal approximation capability of broad learning system and its structural variations," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 4, pp. 1191–1204, Apr. 2019.

[16] S. Kuok and K. Yuen, "Broad learning for nonparametric spatial modeling with application to seismic attenuation," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 35, no. 3, pp. 203–218, Mar. 2020.

[17] X. Cai, X. Feng, and H. Yu, "Broad learning algorithm of cascaded enhancement nodes based on phase space reconstruction," *Appl. Intell.*, vol. 53, no. 2, pp. 2321–2331, Jan. 2023.

[18] X. Shen, Q. Dai, and W. Ullah, "An active learning-based incremental deep-broad learning algorithm for unbalanced time series prediction," *Inf. Sci.*, vol. 642, Sep. 2023, Art. no. 119103.

[19] K. S. O. Rocha, J. H. Martins, M. A. Martins, J. A. O. Saraz, and A. F. L. Filho, "Three-dimensional modeling and simulation of heat and mass transfer processes in porous media: An application for maize stored in a flat bin," *Drying Technol.*, vol. 31, no. 10, pp. 1099–1106, Jul. 2013.

[20] L. I. Quemada-Villagómez, F. I. Molina-Herrera, M. Carrera-Rodríguez, M. Calderón-Ramírez, G. M. Martínez-González, J. L. Navarrete-Bolaños, and H. Jiménez-Islas, "Numerical study to predict temperature and moisture profiles in unventilated grain silos at prolonged time periods," *Int. J. Thermophys.*, vol. 41, no. 5, p. 52, Mar. 2020.

[21] F. Novoa-Muñoz, "Simulation of the temperature of barley during its storage in cylindrical silos," *Math. Comput. Simul.*, vol. 157, pp. 1–14, Mar. 2019.

[22] N. Fumo and M. A. Rafe Biswas, "Regression analysis for prediction of residential energy consumption," *Renew. Sustain. Energy Rev.*, vol. 47, pp. 332–343, Jul. 2015.

[23] Z. Zeng, W. W. Hsieh, A. Shabbar, and W. R. Burrows, "Seasonal prediction of winter extreme precipitation over Canada by support vector regression," *Hydrol. Earth Syst. Sci.*, vol. 15, no. 1, pp. 65–74, Jan. 2011.

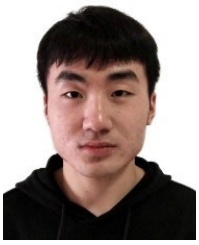
[24] F. Di Paola, E. Ricciardelli, D. Cimini, A. Cersosimo, A. Di Paola, D. Gallucci, S. Gentile, E. Geraldi, S. Larosa, S. T. Nilo, E. Ripepi, F. Romano, P. Sanò, and M. Viggiano, "MiRTaW: An algorithm for atmospheric temperature and water vapor profile estimation from ATMS measurements using a random forests technique," *Remote Sens.*, vol. 10, no. 9, p. 1398, Sep. 2018.

[25] S. Bahrami, R.-A. Hooshmand, and M. Parastegari, "Short term electric load forecasting by wavelet transform and grey model improved by PSO (particle swarm optimization) algorithm," *Energy*, vol. 72, pp. 434–442, Aug. 2014.

[26] Q. Wang, J. Feng, F. Han, W. Wu, and S. Gao, "Analysis and prediction of grain temperature from air temperature to ensure the safety of grain storage," *Int. J. Food Properties*, vol. 23, no. 1, pp. 1200–1213, Jan. 2020.

[27] S. Duan, W. Yang, X. Wang, S. Mao, and Y. Zhang, "Forecasting of grain pile temperature from meteorological factors using machine learning," *IEEE Access*, vol. 7, pp. 130721–130733, 2019.
[28] G. Van Houdt, C. Mosquera, and G. Nápoles, "A review on the long short-term memory model," *Artif. Intell. Rev.*, vol. 53, pp. 5929–5955, May 2020.
[29] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," 2014, *arXiv:1412.3555*.
[30] Y. Wu, H. Yang, K. Zhou, Y. Wang, and Y. Zhu, "Application of bidirectional LSTM neural network in grain stack temperature prediction," in *Proc. 17th Int. Conf. Bio-Inspired Comput., Theories Appl.*, Wuhan, China, Dec. 2022, pp. 385–395.
[31] F. Elmaz, R. Eyckerman, W. Casteels, S. Latré, and P. Hellinckx, "CNN-LSTM architecture for predictive indoor temperature modeling," *Building Environ.*, vol. 206, Dec. 2021, Art. no. 108327.
[32] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," 2014, *arXiv:1409.0473*.
[33] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," 2017, *arXiv:1706.03762*.
[34] A. Zeng, M. Chen, L. Zhang, and Q. Xu, "Are transformers effective for time series forecasting?" in *Proc. 23th Conf. AAAI Artif. Intell.*, Washington, DC, USA, Feb. 2023, pp. 11121–11128.
[35] G. Sharma, A. Singh, and S. Jain, "Hybrid deep learning techniques for estimation of daily crop evapotranspiration using limited climate data," *Comput. Electron. Agricult.*, vol. 202, Nov. 2022, Art. no. 107338.
[36] S. Duan, W. Yang, X. Wang, S. Mao, and Y. Zhang, "Temperature forecasting for stored grain: A deep spatiotemporal attention approach," *IEEE Internet Things J.*, vol. 8, no. 23, pp. 17147–17160, Dec. 2021.
[37] Z. Qu, Y. Zhang, C. Hong, C. Zhang, Z. Dai, Y. Zhao, X. Wu, Y. Gao, X. Jiang, J. Qian, and Z. Gu, "Temperature forecasting of grain in storage: A multi-output and spatiotemporal approach based on deep learning," *Comput. Electron. Agricult.*, vol. 208, May 2023, Art. no. 107785.
[38] P. Raiguru, M. Sahani, S. K. Rout, D. C. Panda, and R. K. Mishra, "A digital direction of arrival estimator based on fast on-line sequential random vector functional link network," *AEU - Int. J. Electron. Commun.*, vol. 142, Dec. 2021, Art. no. 153986.
[39] M. Rayhan Ahmed, S. Islam, A. K. M. Muzahidul Islam, and S. Shatabda, "An ensemble 1D-CNN-LSTM-GRU model with data augmentation for speech emotion recognition," *Expert Syst. Appl.*, vol. 218, May 2023, Art. no. 119633.
[40] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng, "Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations," in *Proc. 26th Annu. Int. Conf. Mach. Learn.*, Montreal, QC, Canada, Jun. 2009, pp. 609–616.
[41] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.
[42] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
[43] F. Poloni and G. Sbrana, "A note on forecasting demand using the multivariate exponential smoothing framework," *Int. J. Prod. Econ.*, vol. 162, pp. 143–150, Apr. 2015.
[44] L. Feiyu, Q. Yao, and F. Maixia, "Prediction model for the stored-grain situation risk point based on broad learning network," *IEEE Access*, vol. 11, pp. 82037–82049, 2023.

QIFU WANG was born in China. He received the Ph.D. degree from the Institute of Electronics, Chinese Academy of Sciences, Beijing, China, in 2011. He is currently an Associate Researcher with the Institute of Applied Physics, Henan Academy of Sciences, Zhengzhou, China. His research interest includes intelligent information processing.



MINGLEI HOU was born in Henan, China, in 2000. He is currently with the College of Information Science and Engineering, Henan University of Technology, and Henan Academy of Sciences, Zhengzhou, China. His research interests include intelligent information processing and machine learning.

YAO QIN was born in Henan, China, in 1981. She received the Ph.D. degree in engineering from the Graduate School, Chinese Academy of Sciences. She was a Visiting Scholar with the University of Houston. Her research interests include information and signal processing.

FEIYU LIAN was born in Zhengzhou, China. He received the B.S. degree in information and communication engineering from Zhengzhou University, in 2001, and the Ph.D. degree in computer application technology from Shanghai University, in 2012. Since 2002, he has been a Lecturer with Henan University of Technology, Zhengzhou. His research interests include the quantitative analysis of terahertz time-domain spectroscopy and the application of THz-TDS to the detection of food quality.