

Combine Harvester Cooling Water Temperature Prediction Based on CDAE-LSTM Hybrid Model

Yining Fu, Baoyan Xu, Xindong Ni, Yehong Liu, Xin Wang*

*Beijing Key Laboratory of Optimized Design for Modern Agricultural Equipment,
College of Engineering, China Agricultural University (East Campus),
Beijing, China
wangxin117@cau.edu.cn*

Abstract—Cooling water temperature of the combine harvester during operations can reflect the changes of its power consumption and even overloads caused by extreme workload. There is an existing problem when extracting water temperature information from harvesters: data redundancy and the loss of time series feature. To solve such problem, a Convolutional denoising autoencoder and Long-Short Term Memory Artificial Neural Network (CDAE-LSTM) hybrid model based on parameter migration is proposed to predict temperature trends. Firstly, the historical data of the combine harvester are taken into account to perform correlation analysis to verify the input rationality of the proposed model. Secondly, pre-training has been performed to determine the model's initial migration parameters, along with the adoption of CDAE to denoise and reconstruct the input data. Finally, after the migration, the CNN-LSTM hybrid model was trained with a real dataset and was able to predict the cooling water temperature. The accuracy of the model has been verified by field test data gathered in June 2019. Results show that the root mean squared error (RMSE) of the model is 0.0817, and the mean absolute error (MAE) is 0.0989. Compared with the performance of LSTM on the prediction data, the RMSE improvement rate is 2.272 %, and the MAE improvement rate is 20.113 %. It is proven that the adoption of CDAE stabilizes the model, and the CDAE-LSTM hybrid model shows higher accuracy and lower uncertainty for time series prediction.

Index Terms—Prediction model; CDAE; Combine harvester; CNN-LSTM.

I. INTRODUCTION

Combine harvesters are the most widely used harvesting machinery in crop harvesting, and their performance often has a direct impact on harvest quality. In actual operation, the harvester load is not constant due to factors, such as crop density, surface undulation, and crop moisture content. This also makes the engine output power of the harvester show a trend of dynamic changes accordingly. When the rated working load is exceeded, it may even cause the harvester to fail severely [1]–[5]. When the engine power changes throughout its operations, the temperature of the cooling water changes accordingly. Therefore, the harvester's load

status can be judged by the temperature of the cooling water. How to collect and analyse the relations between water temperature changes and harvester performance is a problem that needs to be solved urgently [6], [7].

In recent years, machine learning and deep learning combined with traditional feature extraction methods, such as frequency-domain features [8]–[12], wavelet energy entropy, have performed well in fault diagnosis and prediction. Since the introduction of Alex Net [13] in 2012, the deep neural network has been widely used in fault diagnosis and prediction. Therefore, the diagnosis and prediction of faults can be realized by analysing the characteristic parameters of related faults (such as speed, torque, vibration, etc.). Building a data-driven predictive model will be a new way of solving traditional problems. During parameter tuning, a large number of noises and vibrations were detected in the combine harvester's working situation [14]–[17]. They interfere with the feature extraction process, resulting in the final extracted key feature parameters showing nonlinearity, no marking, and mutual interference status. Therefore, how to choose a feature extraction model is one of the current research issues in detection data processing.

For noise issues, which interfere with the feature extraction process, which they causes the abnormal data of key feature parameters to be extracted. Some parameters also show nonlinearity, no marking, and mutual interference status. On the one hand, the time sequence processing was usually ignored in the analysis of working condition parameters. The fault parameter labelling requires manual marks, which is difficult to be used in the working conditions of the fault post-analysis [18]–[20]. On the other hand, different acquisition parameters were viewed separately in traditional research, which makes the variable load correlation difficult to reflect. On the contrary, the deep feature differences of the whole system and the analysis objects, which are mostly vibration signals, are easy to extract [21]–[24].

After viewing the above problems, this paper proposes an improved hybrid forecasting model to explore the relationship between the engine cooling water temperature and the combine harvester's performance. The model is based on the transfer learning convolutional autoencoder.

Manuscript received 11 October, 2020; accepted 16 May, 2021.

This research was funded from the National Key Research and Development Program of China under Grant No. 2017YFD0700603.

First, reconstruct the original data, as well as apply Convolutional Neural Networks (CNN) to extract the working condition data features. Then, create LSTM model to complete the engine cooling water's regression fitting. The accuracy and feasibility of the model were verified through real experimental data.

II. INFLUENCE FACTORS OF WORKING PARAMETERS

A. Input Selection of Prediction Model

The relations of working parameters of combine harvesters present multiple possible causes or one cause with various results. Different researchers have studied the key parts of the combine harvester [25]–[27]. To conduct correlation calculations, the independent time series characteristics of key components like speed, temperature, and engine working index, are selected. To couple the feature information affecting the parameters and fully excavate the potential internal relations and the spatial law of these features, the correlation calculations happen between the parameters are carried out to determine the acquisition parameters. We select four Rotation speeds of important load parts, fuel consumption, and other engine parameters as the inputs of

our model.

B. Correlation Calculation

The correlation analysis is carried out for the original data. As for the combine harvester, a working condition parameter acquisition system is developed combined with a field experiment acquisition point, data from the operation or maintenance management platform. Pearson's correlation coefficient is selected as an indicator to calculate the correlation, the expression is as follows

$$\text{corr}(x, y) = \frac{\text{cov}(x, y)}{\sqrt{\text{Var}(x) \times \text{Var}(y)}}, \quad (1)$$

where $\text{Var}(i)$ is the variance of the corresponding parameter, and $\text{cov}(x, y)$ is the covariance between parameters x and y . Details of the dataset and the calculated image are shown in the Fig. 1; as illustrated, the selected key part parameters correlate $0.2 \leq |\text{Corr}| < 0.52$, which can be used as predictive input parameters for the actual distribution of training data over time. Figure 2 shows the correlation between the parameters.

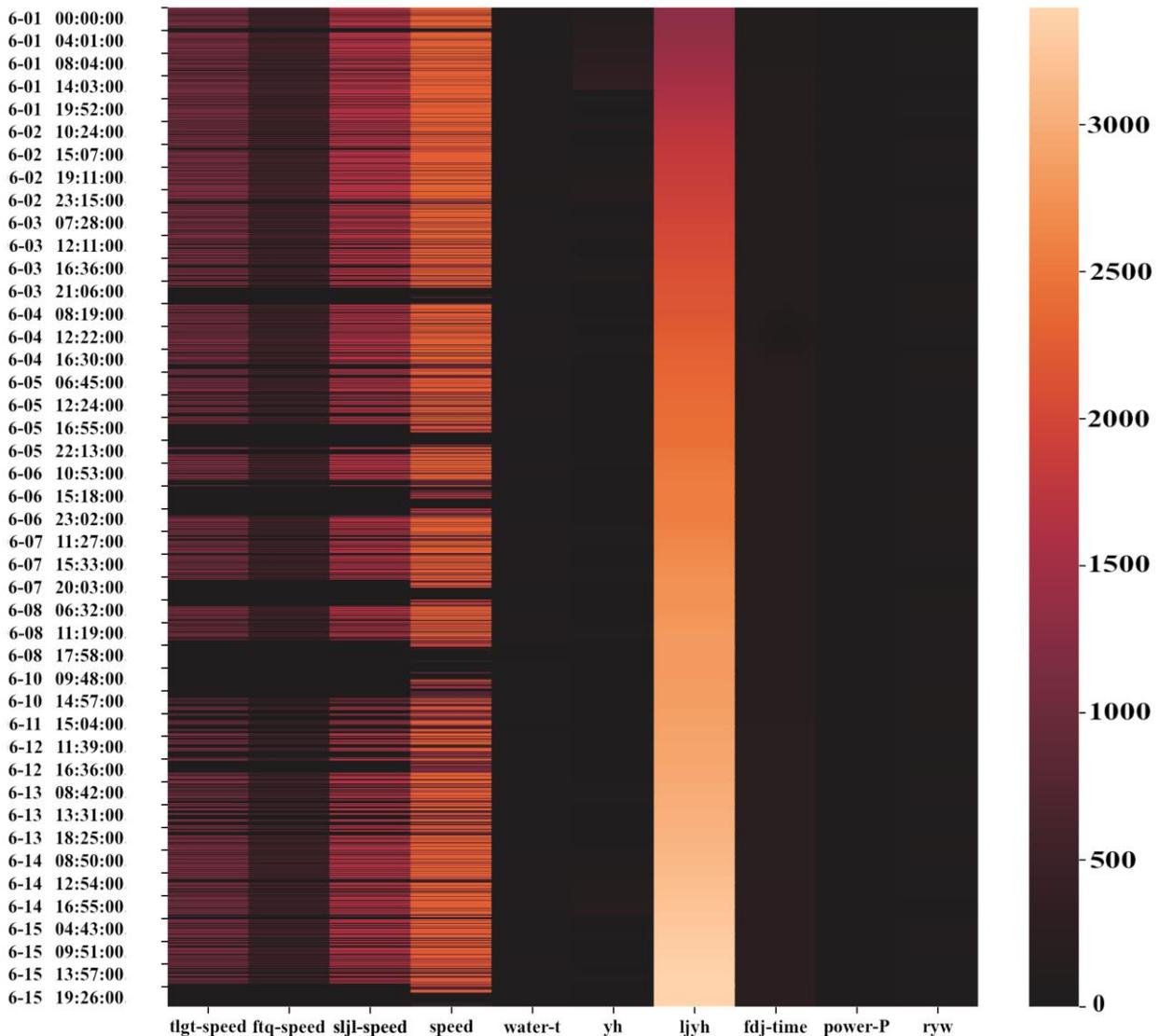


Fig. 1. The details of the Combine harvester dataset.

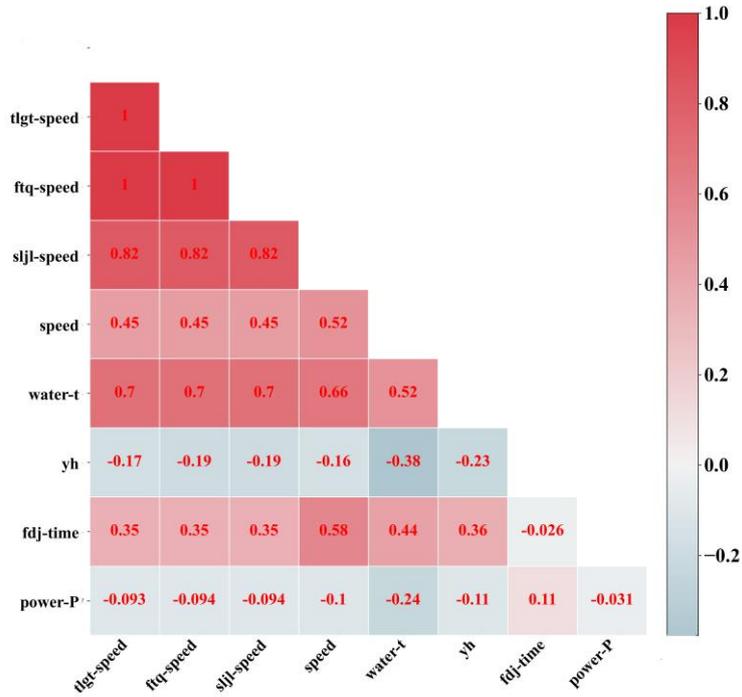


Fig. 2. Correlation calculation diagram of parameters.

C. Time Series Analysis

In this paper, the data are obtained from different sensors based on Controller Area Network-bus (CAN-bus); speed, fuel consumption, and cooling water temperature all belong to multivariate time series, which contain multiple unary time series as portions. The sampling time points of each unary time series are the same. All data can be represented in matrix form, each row represents a time point, and each column stands for a unary time series. The time series representation of the variables of multiple processes at the same time is as follows

$$x_t = (x_{t1}, x_{t2}, \dots, x_{tn})^T. \quad (2)$$

The Pearson correlation matrix in the previous section is adopted as the time series correlation matrix. The prediction of the cooling water temperature needs to learn the univariate process of each portion and to learn the relations and variation laws between each portion to make the prediction and controlling of the time series of cooling water temperature affected by the workload.

In fact, the use of LSTM is the combination of multiple regression and time series analysis. It needs to determine the prediction model's structure according to the cooling water temperature parameters before using the LSTM model. Multi-step prediction needs to be conducted to meet the requirements of operation and maintenance management for combine harvesters. In this case, we need to predict the data from the current time to a period. The input data can be obtained through a sliding window. Therefore, the persistence model can be built, which specific details will be described in Section III.

III. RESEARCH METHOD

First, using a convolution noise reduction in the model design, we follow the original autoencoder to reduce noise, build the pre-train model as closely as possible. Then the data

were inputted into the improved CNN-LSTM model to make the prediction. The data format is one-dimensional sensor data. Adopting CNN-LSTM training parameters as the model parameters of the Convolutional denoising autoencoder (CDAE) can reduce the training difficulty. Adopting the feature extraction of deep data as the input of LSTM can reduce the influence of abnormal data on the time series and improve the prediction accuracy.

Specifically, the initial parameter range of the model structure is obtained through pre-training. The optimal hyperparameter combination is obtained by searching randomly from the search function. The pre-trained model parameters are used as the input of transfer learning for the initialization of the CDAE and CNN-LSTM model.

Transfer learning is a type of machine learning, which aims to obtain general knowledge representation from existing machine learning tasks. It is applied to other tasks to optimize the model and accelerate convergence. In transfer learning, parameter transfer is to transfer the weights of nodes in a partial layer network from a trained network to an untrained network with the same structure, rather than train a model for a certain task from the beginning. Training and fine-tuning a new model based on transferred model parameters can reduce the training time of hyperparameters, saving computing resources and making the model more robust. The generalization ability is stronger than the retrained model.

A. One-Dimensional Convolutional Network

The Convolutional layer in CNN is divided into one-dimensional and two-dimensional convolutions according to the dimensions of the input. We use one-dimensional convolution layer (Covn1D) to construct a hybrid model. Covn1D is used to excavate the deep features and potential information of input multi-parameter time series data. The Covn1D appearing in following are all one-dimensional convolution single-layer, and Convolutional Neural Networks are formed by multiple

Covn1D layers and other Operation layers. Complete convolution operation adding the pooling layer, automatically extract the useful feature representation of the original data and output the feature vector. The convolution layer operation formula is as follows

$$\mathbf{x}_j^l = f\left(\sum_{i \in M_j} \mathbf{x}_i^{l-1} \times \mathbf{k}_{ij}^l + \mathbf{b}_j^l\right) \quad (3)$$

where \mathbf{x}_i^{l-1} is the output characteristic of the upper layer network, \mathbf{k}_{ij}^l is the learnable convolution kernel of the current layer. The output of the current layer network \mathbf{x}_j^l can be obtained through the activation function $f(x)$; \mathbf{b}_j^l is the offset of the current layer M_j feature set.

B. Convolutional Noise Reduction Autoencoder

The autoencoder is greatly affected by the original data. To learn the original data with superimposed noise, we can use the DAE to get better results. Convolutional noise reduction autoencoder conducts the process of data compression and decompression, the original data are mixed with noise, and then the original data are reconstructed through unsupervised learning, the back-propagation optimization algorithm is used to make the reconstruction error rate meet the noise reduction requirement [28]. In this paper, the mean absolute error (MAE) is regarded as the reconstruction loss function in the autoencoder, DAE structure is shown in Fig. 3. Autoencoder uses the mapping relationship between the input layer and the output layer to realize sample reconstruction and extract features, the formula is as follows

$$L_1 = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i). \quad (4)$$

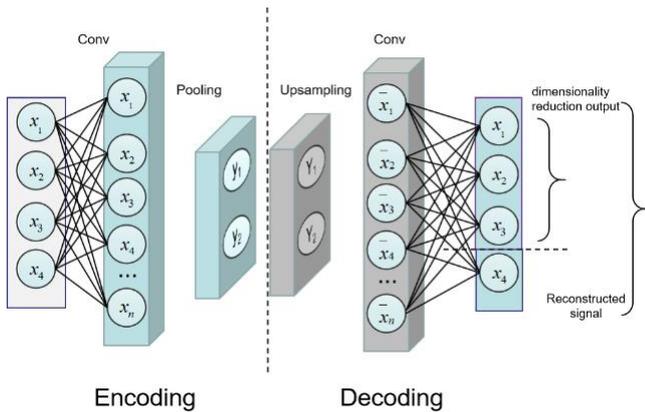


Fig. 3. Structure diagram of simple DAE.

The structure of the convolutional autoencoder is based on the model structure of the encoder CNN in the pre-model, which adopts the same number of one-dimensional convolution layer and pooling layer, and the LSTM layer is regarded as the outermost layer of the decoder to replace the deconvolution layer. The hyperparameters are adjusted based on the initial state parameters of the transfer CNN-LSTM. The next-step adjustment is carried out after the deconvolution of the inner parameters of each layer and the

outermost layer, and finally, the convolution noise reduction autoencoder, which can effectively meet the noise reduction requirement, can be obtained.

C. Long-Short Term Memory Artificial Neural Network

Long-Short Term Memory Artificial Neural Network (LSTM) is an improved feedback neural network based on Recurrent Neural Network (RNN), which can effectively deal with the sensitive problems of time series and solve gradient disappearance's problems on the premise of inheriting the characteristics of the RNN model [29]. For the problems that combine harvester working parameters belong to time series, LSTM can effectively simulate the combine harvester working logic and learn the implicit characteristics between parameters. The specific structure of single-layer multi-step LSTM is shown in Fig. 4.

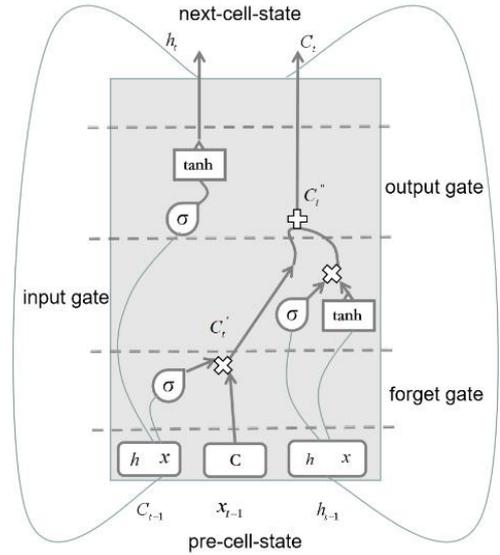


Fig. 4. Structure diagram of Single layer.

LSTM adds or removes information to the cell state through the threshold, which includes the forgetting gate, input gate, and output gate. In the threshold structure, σ represents the sigmoid layer, \tanh represents the \tanh layer. Cell state C and hidden layer state h are updated through vector calculation, the output vector of LSTM is as follows:

$$\mathbf{I}_t = \sigma(\mathbf{w}_{xI} \mathbf{x}_t + \mathbf{w}_{hI} \mathbf{h}_{t-1} + \mathbf{w}_{CI} \mathbf{C}_{t-1} + \mathbf{B}_I), \quad (5)$$

$$\mathbf{F}_t = \sigma(\mathbf{w}_{xF} \mathbf{x}_t + \mathbf{w}_{hF} \mathbf{h}_{t-1} + \mathbf{w}_{CF} \mathbf{C}_{t-1} + \mathbf{B}_F), \quad (6)$$

$$\mathbf{O}_t = \sigma(\mathbf{w}_{xO} \mathbf{x}_t + \mathbf{w}_{hO} \mathbf{h}_{t-1} + \mathbf{w}_{CO} \mathbf{C}_{t-1} + \mathbf{B}_O), \quad (7)$$

$$\mathbf{c}_t = \mathbf{F}_t \mathbf{c}_{t-1} + \mathbf{I}_t \tanh(\mathbf{w}_{xc} \mathbf{x}_t + \mathbf{w}_{hc} \mathbf{h}_{t-1} + \mathbf{B}_c), \quad (8)$$

$$\mathbf{h}_t = \mathbf{O}_t \tanh(\mathbf{C}_t). \quad (9)$$

In the formulas, \mathbf{w}_{xI} , \mathbf{w}_{xF} , \mathbf{w}_{xO} , and \mathbf{w}_{xc} represent the input gate, forgetting gate, output gate, and the weight matrix from the memory unit to \mathbf{x}_t , respectively, while \mathbf{w}_{hI} , \mathbf{w}_{hF} , \mathbf{w}_{hO} , and \mathbf{w}_{hc} are the weight matrix to \mathbf{h}_{t-1} . \mathbf{B}_i represents the bias vector, σ represents the sigmoid function $1/(1+e^{-x})$; the forgetting gate, \mathbf{x}_t , \mathbf{C}_{t-1} , and intermediate output jointly determine the forgetting operation of the memory unit, the updated \mathbf{O}_t and \mathbf{C}_t jointly determine the output \mathbf{h}_t .

IV. EXPERIMENTS AND ANALYSIS

To verify the stability and effectiveness of the prediction method, field experiments were carried out (as shown in Fig. 5). The experimental object was a certain combine harvester. The nominal feeding rate is 8 kg/s, the cutting width is 2560 mm, and the matching power is 140 kW.



Fig. 5. Field experiment.

A. Equipment and Steps of Experiment

The experimental equipment includes a combine harvester working condition acquisition system based on CAN bus of SAE J1939 protocol (as shown in Fig. 6). To obtain the key working condition parameters of the combine harvester, a maintenance management platform was developed to upload real-time data. The acquisition device's sampling frequency was 1 HZ, and the sensor's acquisition accuracy error is shown in Table I.

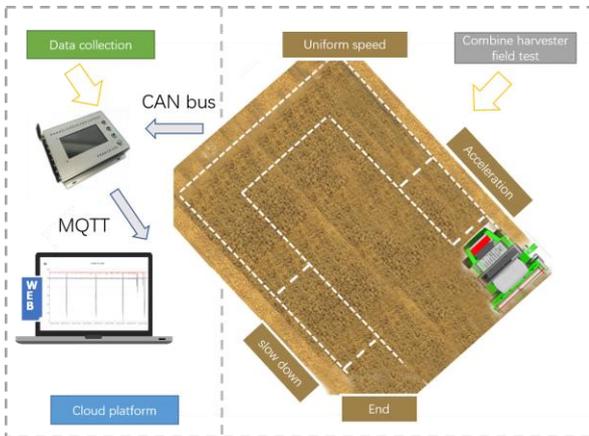


Fig. 6. Data acquisition system.

TABLE I. SENSOR PERFORMANCE PARAMETERS.

Specification of Field Device		
Sensor model	NCB4-12GM40-N0	HSTL-PT100S
Collection object	Rotating speed	Temperature
Type	Hall	Thermal resistance
Supply voltage	DC 10 V–30V	DC 12 V–40 V
Accuracy	3 % FS	0.2 % FS
Measuring range	4 mm	-100 °C–500 °C

Before the harvester started, the driver controlled the engine speed by manual throttle and kept the engine speed stable under the rated condition. After starting the operation, the driver accelerated to 6 km/h–8 km/h and kept the speed relatively stable, followed by the working condition

acquisition system started to collect the rotating speed of the engine, threshing roller, re-thresher, grain conveyor screw, and other components, and the temperature of key components. The engine parameters were transmitted to the acquisition system through CAN bus.

B. Experimental Index and Date Processing

According to the requirements of the China national standard GB/T 8097-2008 “Equipment for harvesting Combine-Harvester-Test procedure” on crop and field conditions, the experimental field should be flat, and the crops should be in uniform growth.

During the experiment, the environment temperature was between 20 °C–34 °C, and the wind speed was between 1.6 m/s–3.3 m/s. CNN-LSTM was designed to predict model performance, regarding R^2 as the regression coefficient and MAE, and the Root Mean Squared Error (RMSE) as experimental indexes. The calculation formulas are as follows:

$$R^2 = 1 - \frac{SSE}{SST}, \quad (10)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \bar{y}_i)^2}, \quad (11)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \bar{y}_i|, \quad (12)$$

where SSE is the sum of squares of residuals, $SSE = \sum_{i=1}^N (y_i - \hat{y}_i)^2$, SST is the sum of squares of total deviations, $SST = \sum_{i=1}^N (y_i - \bar{y}_i)^2$, and y is the sample characterization variable.

The data set is designed according to the data uploaded from sensors. Because parameter resources and the acquisition frequency are different, the data should be pre-processed. Thus, the collected 9900 data entries were divided into 8000 training sets and 1900 verification data sets. The original water temperature and speed sensor data were processed evenly through a sliding window, which was obtained using convolution operation. The convolution formula is as follows

$$(\mathbf{a} \times \mathbf{v})[n] = \sum_{m=-\infty}^{\infty} \mathbf{a}[m] \mathbf{v}[n-m], \quad (13)$$

where \mathbf{a} and \mathbf{v} are input arrays, and the size of the convolution kernel takes a sliding step of 10. Min-Max scaling is used to improve the convergence speed and prediction accuracy of the model.

The computer hardware configuration used in this paper is as follows: the processor is Inter(R) Core (TM)i7-9750H, 16G-RAM, the operating system is 64-bit Windows 10, and the GPU graphics board is NVIDIA GeForce GTX 1660 Ti. The software framework is the Keras deep learning framework, and the deep learning framework TensorFlow 2.0 is the back-end support, the programming language version is Python 3.7, and the integrated development environment is PyCharm. The overall process is shown in Fig. 7.

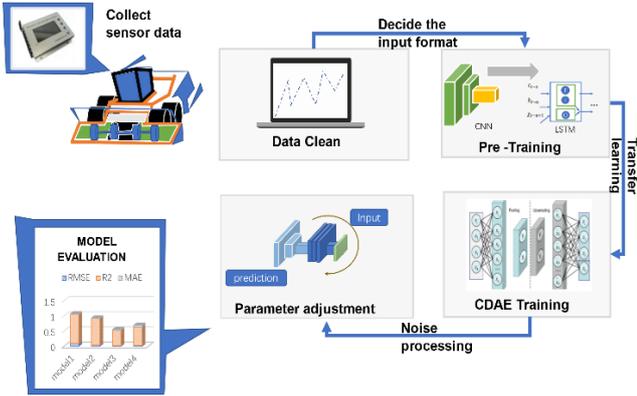


Fig. 7. Process of experiment of the predictive model from a field trail to model evaluation.

C. Model Training

We evaluated the statistics of our model according to the indexes. Fine-tuning shows that the RMSE and MAE were greatly affected by abnormal values. First, the regression coefficient R^2 is regarded as the optimization standard to determine the model's fitting degree. The result can determine the approximate range of the parameters through presenting parameters. The filters and the number of layers may be increased according to the condition of pre-training. It is found that when the number of layers was increased, the loss function decreased, and R^2 decreased as well. However, RMSE first increased and then decreased and MAE, on the contrary, decreased first and then followed by increase.

Different layers and different hyperparameters have a certain effect on R^2 , making it difficult to determine the best structure because the layer is too low. Further parameter and structure adjustments are needed. Besides, R^2 cannot fully reflect the prediction ability of the model. Taking the example of the water temperature of combine harvesters, the range of the independent variable of the actual value of water temperature is 0°C – 80°C , which is too small, and the R^2 of the model is closed to 1. Therefore, the best prediction model can be determined by RMSE and MAE.

The above analysis is the basis for training after the model was designed. As described in the third section of the text, pre-training is required, and the initial transferred value to CDAE and CNN-LSTM for the next step of training.

1. Pre-training of Initial Model

CNN-LSTM model [30] layers and initialization parameters need to be trained in the pre-training, where the specific training steps of Conv1D convolution layer and pooling layer are as follows:

- Determine the size and dimension of the sample according to the input sample and parameters and initialize the kernel size;
- Input pre-processed data into Conv1D according to the input format of one-dimensional convolutional neural network;
- Adjust the kernel size and the number of filters to determine the best CNN extraction model parameters.

In this paper, the grid search “GridSearchCV” method is adopted to traverse batch_size and epochs and then continue to adjust other parameters [31]. The range of the parameters is shown in Fig. 8. When the epochs are 100 and batch_size is 3, the combination has better results. Then we used the

random sampling tool to adjust other parameters.

RandomizedSearchCV is adopted to adjust the hyper-parameters of the CNN-LSTM model to obtain the optimal structure. The main parameters are introduced in the third section. After getting these key parameters, the search is carried out taking R^2 as the objective function. The hyper-parameters of the hybrid model determine the parameter types shown in the figure below. The training cost increases due to the correlation of the parameters, the optimization strategy with relatively low learning cost and high accuracy is selected for optimization to achieve pre-training.

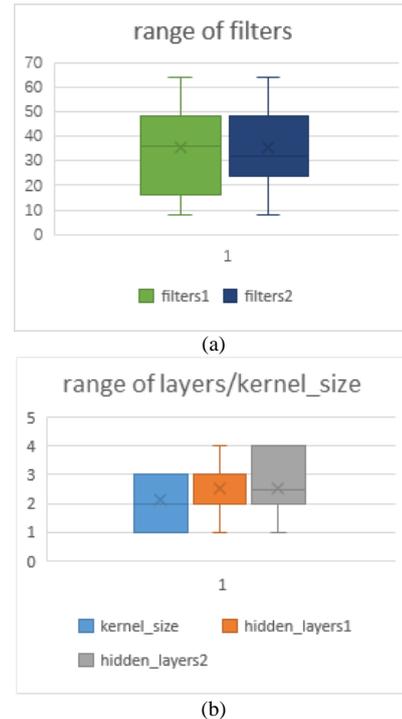


Fig. 8. Pre-adjusted Parameter Range: (a) range of filters and (b) range of layers/kernel_size.

First, we get the best combination of initialization parameters obtained through pre-training. The structure and parameters of the pre-training model are taken as transfer objects (as shown in Table II).

TABLE II. STRUCTURE OF PROPOSED CNN-LSTM MODEL.

CNN-LSTM		
S. No.	Layers	All-params
Conv1D	3	7
LSTM	2	27,945
Max pooling	2	-
Flatten	1	-
Dense	1	55,361

Pre-training can reduce the number of calculations needed and improve the accuracy in the initial training of the convolutional noise reduction autoencoder and CNN-LSTM model. The final model structure can be obtained after further adjustment of the inner structure.

2. Convolutional Denoise Autoencoder Training

The structure and initialization parameters of CDAE conduct parameter transfer learning based on the convolutional coder of the pre-training model (as shown in Table III). To ensure that the lowest feature extraction, a layer of Conv1D is set at the outermost layer. The encoder

consists of Conv1D and UpSampling. The structure is shown in Fig. 9. The CDAE reconstructs the original data by inputting Gaussian noise in an autoencoder, the data will be generated after reconstructing by an autoencoder. The degree of

restoration is affected by the noise center. The reconstruction effect of the structure of different encoders is shown in Fig. 10.

TABLE III. STRUCTURE INFORMATION OF PROPOSED CDAE MODEL.

S. No.	Layer name	Filter × kernel size	Region	Activation	Parameter
Encoder section					
0	Input	-	-	-	0
1	Conv1D	128×3	-	Relu	512512
2	Conv1D	64×3-	-	Tanh	24640
3	Conv1D	32×3	-	Tanh	6176
4	Conv1D	16×3	-	Tanh	1552
5	MaxPooling1D	-	8	-	0
Decoder section					
6	UpSampling1D	-	-	-	0
7	Conv1D	16×3	-	Tanh	784
8	Conv1D	32×3	-	Tanh	1568
10	Conv1D	64×3	-	Tanh	6208
11	Conv1D	128×3	-	Sigmoid	24704
15	Dense	-	1	-	36

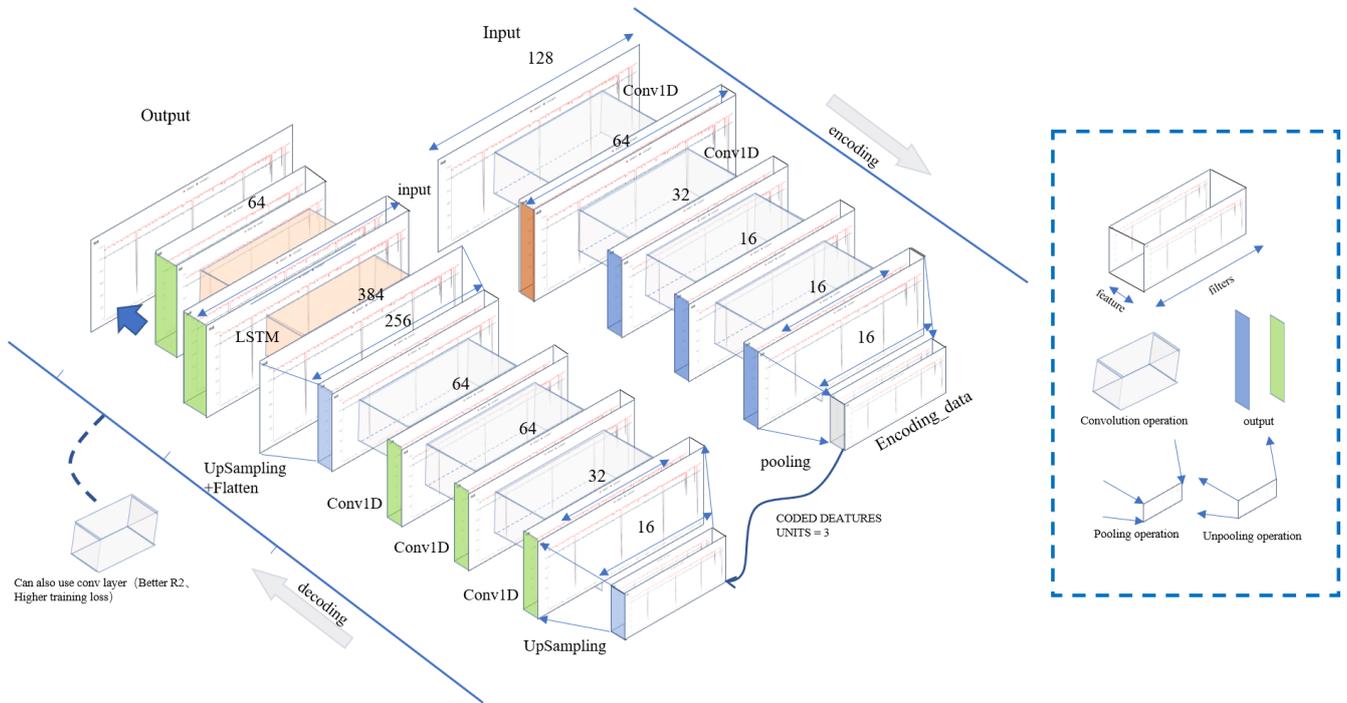
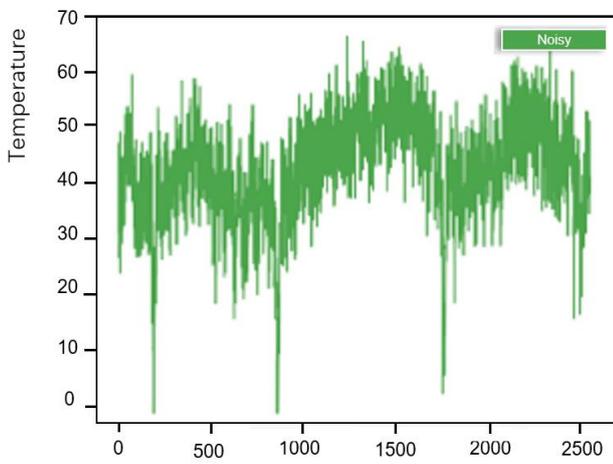
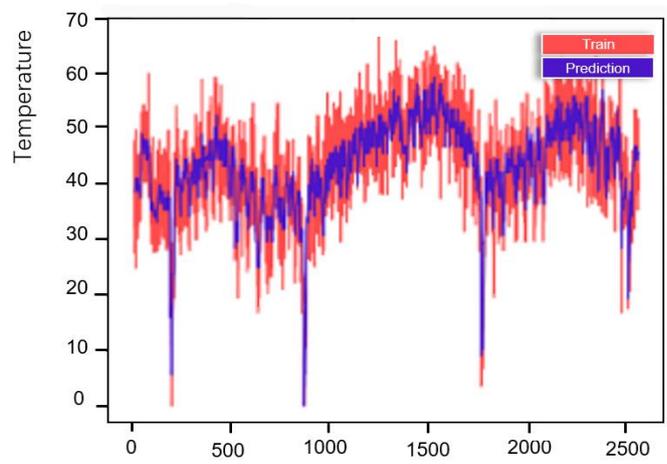


Fig. 9. Structure of different CDAE Models. In the figure, the structure of the hybrid model is transferred from pre-training (Conv1D is a one-dimensional convolutional layer, LSTM layers are finally replaced with Conv1D to obtain the structure in Table III).



(a)



(b)

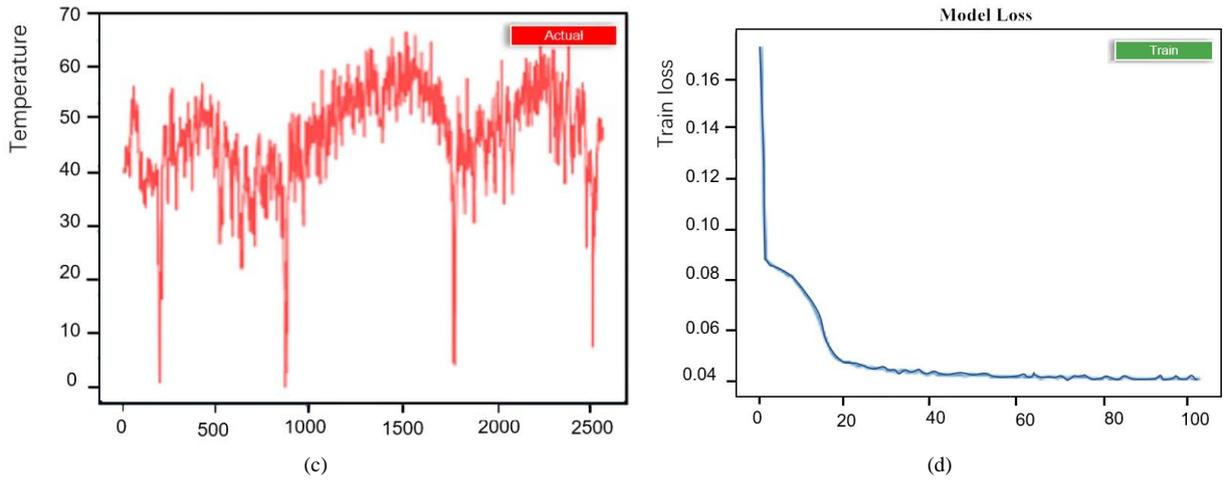


Fig. 10. Graphical representation denoised the results of the evaluated methods on record 2500 and Loss function of the CDAE: (a) Noisy water temperature signal; (b) Noisy & original signal; (c) Clean signal; (d) Train loss of CDAE on 0.05 noise center.

3. CNN-LSTM Model Training

In this paper, the pre-training model's structure is taken as the initialization of the CNN-LSTM. We set the time step as 5, the data from $t - 1$ to $t + 4$ are input, and the parameters of the model are adjusted continuously to optimize the performance of the model with Adam optimization algorithm. To avoid the model falling into a local minimum, the initial learning rate is set as 0.001, and the parameters are

adjusted until the result reaches the optimal. The parameter adjustment process is the same as the pre-training. The training process effect is shown in Fig. 11.

To verify the prediction accuracy of the model, RMSE and MAE were established as evaluation indicators, the model training effect can be evaluated by the promotion rate of RMSE and MAE. The calculation formulas are as follows:

$$RMSE_{pr} = \frac{RMSE_{(LSTM)} - RMSE_{(CNN-LSTM)}}{RMSE_{(LSTM)}} \times 100\%, \quad (14)$$

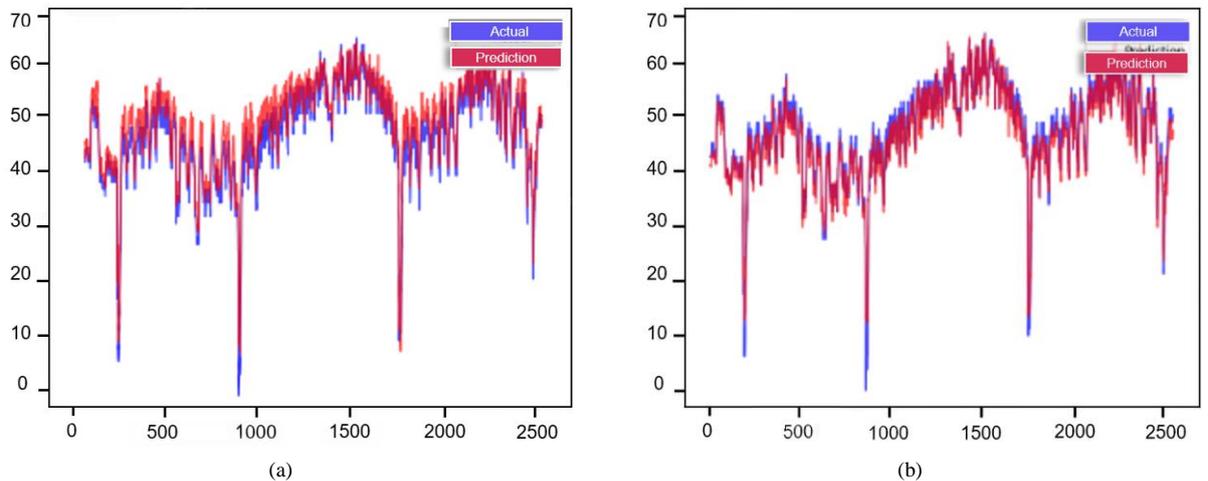
$$MAE_{pr} = \frac{MAE_{(LSTM)} - MAE_{(CNN-LSTM)}}{MAE_{(LSTM)}} \times 100\%. \quad (15)$$

4. Analysis of Model Prediction Effect

The training effect of the CDAE is shown in Fig. 12. The CDAE has a significant effect on data denoise. It improves the prediction efficiency combined with the data pre-processing ability. Different CDAE structures have different denoise effects on the data, which can be seen in Fig. 12. The CDAE based on transfer learning can denoise more effectively.

Figure 13(b) shows the degree of fit between the predicted cooling water temperature of the CDAE-CNN-LSTM model and the actual experimental results. To verify the accuracy and stability of the proposed prediction model, we select the

Back Propagation (BP) network model, the LSTM model Fig. 13(a), and the CDAE-CNN-LSTM model to compare and analyse. It can be seen in Fig. 13 that the CDAE-CNN-LSTM has a better fitting result. The RMSE, the MAE, and the R^2 are 0.0817, 0.0989, and 0.9785, respectively. The training loss function shows that the CNN-LSTM has a lower degree of overfitting. Compared with the LSTM model, the promotion rate of RMSE and MAE is 2.272 % and 20.113 %, respectively. The prediction effects of different models are shown in Table IV.



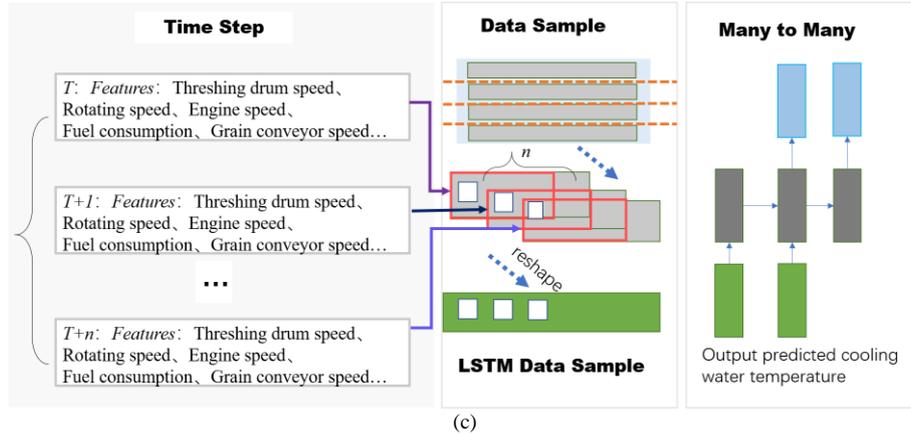


Fig. 11. Training process details of time steps and predictive effect: (a) LSTM Forecast comparison without smooth (volatility obviously); (b) CNN-LSTM Forecast comparison; (c) Time step for LSTM input selection.

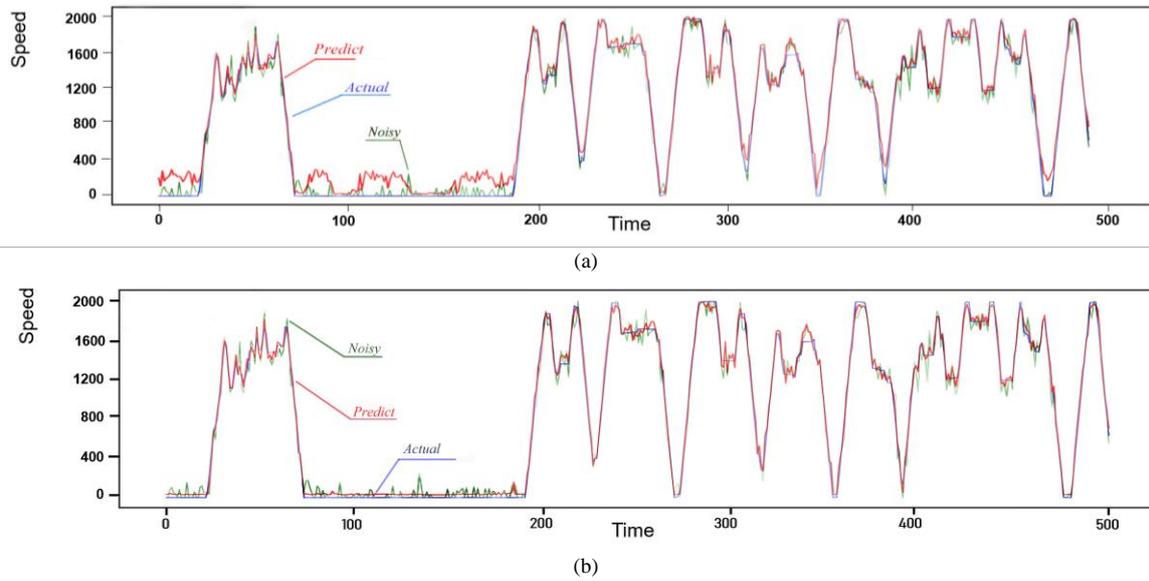
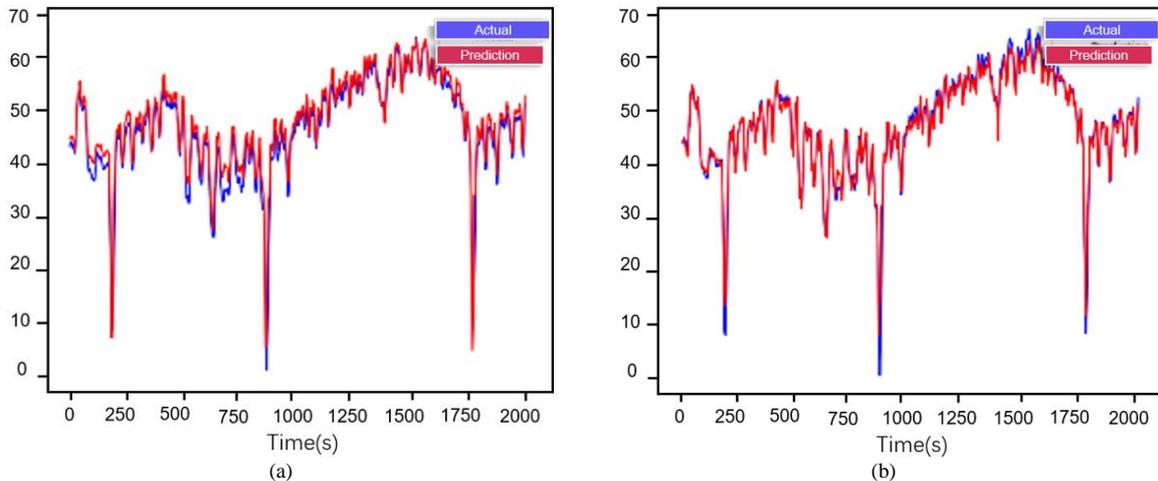


Fig. 12. Noise reduction effect comparison for different structures of CDAE: (a) Decoder part without transfer learning (The red predicted value is significantly affected by noise); (b) Decoder part using CNN structure based on transfer learning.

TABLE IV. THE SCORE OF DIFFERENT MODEL PREDICTION RESULTS.

Model	Parameter (Model)			R ²	RMSE	MAE
	Layers	Params	Epoch			
BP	9	4113	100	0.5373	0.0904	0.3111
LSTM	3	27,945	100	0.9818	0.0836	0.1238
CNN-LSTM	6 + 3	75,249	100	0.9785	0.0817	0.0989



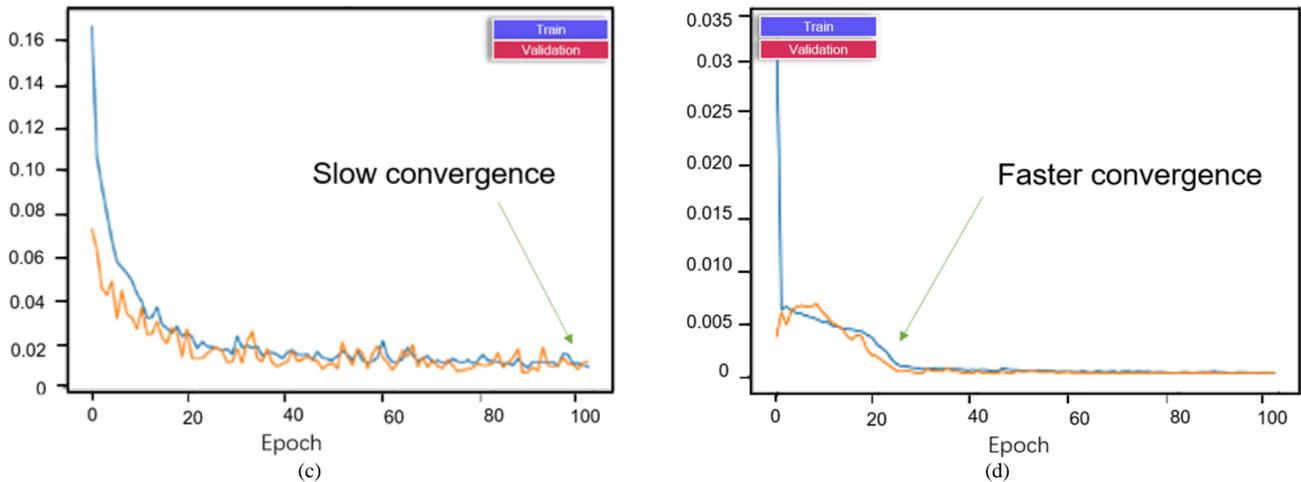


Fig. 13. Results of cooling water temperature prediction: (a) Predicted situation before parameter adjustment; (b) Final prediction model (The predicted value is close to the true value); (c) Before parameter adjustment model loss; (d) Final loss model (Fast convergence).

Combining chart analysis and the CDAE-LSTM hybrid model has led to a fabulous fit than the traditional BP neural network. It is almost difficult for the BP neural network to fit the time series data (R^2 shows poor-fitting effect). The hybrid model R^2 is slightly lower than the original LSTM, but the drop rate is less than 1%. Under the premise of fulfilling the requirements of fitting accuracy, the hybrid model is a lower risk, considering the impact of overfitting. As for accuracy analysis, on the one hand, the statistics RMSE can better reflect the influence of discrete points. From the table, we can see that CDAE-LSTM has good effects on abnormal data training. On the other hand, MAE can most intuitively reflect the prediction effects. Comparing CDAE-LSTM and LSTM from the fitting effect diagram, it shows that the hybrid model has better fitting accuracy.

V. CONCLUSIONS

1. The traditional regression model has poor processing efficiency within time series and it is unable to predict the parameters of unstable time series. Therefore, deep learning is introduced to the modelling and analysis of multi-source parameter time series. A neural network based on CDAE and CNN-LSTM is proposed. The cooling water temperature prediction model is proved by experiments that it has a high accuracy.
2. This paper proposes that the CDAE-LSTM model based on migration learning has a better prediction effectiveness than the original LSTM model. CDAE can effectively eliminate data anomalies caused by noise with limited sampling, as well as obtain the best migration process data. Since DAE is a lossy process for data processing, extended studies around output degradation may be considered in the future.
3. The prediction model's accuracy rate was analysed using the statistics of RMSE, MAE, and R^2 , which verified that the model had low overfitting based on the field test data. Under the premise of risk ($R^2 = 0.9785$), it has a higher accuracy rate (MAE = 0.0989, RMSE = 0.0817), which shows the model's robustness. The obtained regression model can provide an early warning basis for the subsequent changes in the combine harvester. Furthermore, the trend of the fault parameters can provide

potential ideas for the predictive diagnosis of the combine harvester.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

- [1] L. Xu, C. Wei, Z. Liang, X. Chai, Y. Li, and Q. Liu, "Development of rapeseed cleaning loss monitoring system and experiments in a combine harvester", *Biosystems Engineering*, vol. 178, pp. 118–130, 2019. DOI: 10.1016/j.biosystemseng.2018.11.001.
- [2] N. E. Isaac, G. R. Quick, S. J. Birrell, W. M. Edwards, and B. A. Coers, "Combine harvester econometric model with forward speed optimization", *Applied Engineering in Agriculture*, vol. 22, no. 1, pp. 25–31, 2006. DOI: 10.13031/2013.20184.
- [3] Y. Li and K. Wang, "Study on fault diagnosis and load feedback control system of combine harvester", in *Proc. of 7th International Conference on Electronics and Information Engineering (ICEIE)*, Nanjing, 2017. DOI: 10.1117/12.2265355.
- [4] V. Kuchtov, "Classification of failures and the reliability of parts of the variator combine harvesters", *Technical service of agriculture, forestry and transport systems*, vol. 6, no. 2, pp. 18–23, 2016.
- [5] B. Xu, X. Ni, Y. Wang, Y. Wang, Y. Liu, and X. Wang, "Optimization of threshing quality control strategy based on type-2 fuzzy logic controller", *Elektronika ir Elektrotechnika*, vol. 26, no. 2, pp. 15–23, 2020. DOI: 10.5755/j01.eie.26.2.25924.
- [6] A. A. Arkadan, S. Subramaniam-Sivasenan, and O. Douedari, "Design optimization of a capacitive transducer for displacement measurement", *IEEE Transactions on Magnetics*, vol. 35, no. 3, pp. 1869–1872, 1999. DOI: 10.1109/20.767398.
- [7] A. Grincova, M. Andrejiova, and D. Marasova, "Analysis of conveyor belt impact resistance data using a software application", *Acta Polytechnica Hungarica*, vol. 14, no. 2, pp. 113–130, 2017. DOI: 10.12700/APH.14.2.2017.2.6.
- [8] W. Xin, X. Miao, L. Weiwei, L. Hongsong, Z. Shaohong, and W. Shumao, "Study on orchard vehicle motion stability control system", *IFAC - papers online*, vol. 49, no. 16, pp. 451–456, 2016. DOI: 10.1016/j.ifacol.2016.10.082.
- [9] J. Yuanyuan, C. Jin, and Y. Guangjing, "A signal acquisition and failure alarm system of combine harvester based on touch screen technology", *Applied Mechanics and Materials*, vols. 347–350, pp. 28–31, 2013. DOI: 10.4028/www.scientific.net/AMM.347-350.28.
- [10] H. Abdi, "A neural network primer", *Journal of Biological Systems*, vol. 2, no. 3, pp. 247–281, 1994.
- [11] W. Yu and C. Zhao, "Robust monitoring and fault isolation of nonlinear industrial processes using denoising autoencoder and elastic net", *IEEE Transactions on Control Systems Technology*, vol. 28, no. 3, pp. 1083–1091, 2020. DOI: 10.1109/TCST.2019.2897946.
- [12] R. Huang, Y. Liao, S. Zhang, and W. Li, "Deep decoupling convolutional neural network for intelligent compound fault diagnosis", *IEEE Access*, vol. 7, pp. 1848–1858, 2019. DOI: 10.1109/ACCESS.2018.2886343.

- [13] B. Cai, H. Liu, and M. Xie, "A real-time fault diagnosis methodology of complex systems using object-oriented Bayesian networks", *Mechanical Systems and Signal Processing*, vol. 80, pp. 31–44, 2016. DOI: 10.1016/j.ymssp.2016.04.019.
- [14] G. Craessaerts, J. De Baerdemaeker, and W. Saeys, "Fault diagnostic systems for agricultural machinery", *Biosystems Engineering*, vol. 106, no. 1, pp. 26–36, 2010. DOI: 10.1016/j.biosystemseng.2009.12.004.
- [15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks", *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017. DOI: 10.1145/3065386.
- [16] S. Khaki, L. Wang, and S. V. Archontoulis, "A CNN-RNN framework for crop yield prediction", *Frontiers in Plant Science*, vol. 10, p. 1750, 2020. DOI: 10.3389/fpls.2019.01750.
- [17] D. Kolar, D. Lisjak, M. Pająk, and D. Pavković, "Fault diagnosis of rotary machines using deep convolutional neural network with wide three axis vibration signal input", *Sensors*, vol. 20, no. 14, 2020. DOI: 10.3390/s20144017.
- [18] J. Lei, C. Liu, and D. Jiang, "Fault diagnosis of wind turbine based on Long Short-term memory networks", *Renewable Energy*, vol. 133, pp. 422–432, 2019. DOI: 10.1016/j.renene.2018.10.031.
- [19] R. Zhao, R. Yan, J. Wang, and K. Mao, "Learning to monitor machine health with convolutional bi-directional LSTM networks", *Sensors*, vol. 17, no. 2, p. 273, 2017. DOI: 10.3390/s17020273.
- [20] Z. Liang, Y. Li, Z. Zhao, and L. Xu, "Structure optimization of a grain impact piezoelectric sensor and its application for monitoring separation losses on tangential-axial combine harvesters", *Sensors*, vol. 15, no. 1, pp. 1496–1517, 2015. DOI: 10.3390/s150101496.
- [21] J. Chen, P. Wu, and K. Xu, "Remote fault information acquisition and diagnosis system of the combine harvester based on LabVIEW", in *Proc. of 2015 International Conference on Applied Mechanics, Mechatronics and Intelligent Systems (AMMIS)*, 2016, pp. 285–292. DOI: 10.1142/9789814733878_0041.
- [22] Y. W. Zhang, D. Chen, S. M. Wang, and B. Lu, "Assembly quality detection software for combine harvester based on vibration spectra analysis", *Applied Mechanics and Materials*, vols. 635–637, pp. 870–873, 2014. DOI: 10.4028/www.scientific.net/AMM.635-637.870.
- [23] A. Laučka, V. Adaskeviciute, D. Andriukaitis, "Research of the Equipment Self-Calibration Methods for Different Shape Fertilizers Particles Distribution by Size Using Image Processing Measurement Method," *Symmetry*, vol. 11, no. 7, p. 838, Jun. 2019. DOI: 10.3390/sym11070838.
- [24] W. Xin, X. Miao, L. Weiwei, B. Miedzinski, and W. Shumao, "Optimal design of lifting platform balance control algorithm: Co-simulation and experiment", *Elektronika ir Elektrotechnika*, vol. 23, no. 5, pp. 3–10, 2017. DOI: 10.5755/j01.eie.23.5.19237.
- [25] Y. Hu, Y. Yan, L. Wang, X. Qian, and X. Wang, "Simultaneous measurement of belt speed and vibration through electrostatic sensing and data fusion", *IEEE Transactions on Instrumentation and Measurement*, vol. 65, no. 5, pp. 1130–1138, 2016. DOI: 10.1109/TIM.2015.2490958.
- [26] Z. G. Wei, "Torsional vibration dynamics modeling and simulation research on harvester", *Applied Mechanics and Materials*, vol. 443, pp. 160–163, 2013. DOI: 10.4028/www.scientific.net/AMM.443.160.
- [27] C. A. G. da Silva, "Diagnostic of failure in transmission system of agriculture tractors using predictive maintenance based software", *AgriEngineering*, vol. 1, no. 1, pp. 132–144, 2019. DOI: 10.3390/agriengineering1010010.
- [28] Y. Wu, C. DuBois, A. X. Zheng, and M. Ester, "Collaborative denoising auto-encoders for top-N recommender systems", in *Proc. of the Ninth ACM International Conference on Web Search and Data Mining*, San Francisco, CA, 2016, pp. 153–162. DOI: 10.1145/2835776.2835837.
- [29] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A Search space odyssey", *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 10, pp. 2222–2232, 2017. DOI: 10.1109/TNNLS.2016.2582924.
- [30] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization", *Journal of Machine Learning Research*, vol. 13, pp. 281–305, 2012.
- [31] J. Donahue *et al.*, "Long-term recurrent convolutional networks for visual recognition and description", in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 2625–2634. DOI: 10.1109/CVPR.2015.7298878.



This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 (CC BY 4.0) license (<http://creativecommons.org/licenses/by/4.0/>).