

# Fault prediction of combine harvesters based on stacked denoising autoencoders

Zhaomei Qiu<sup>1</sup>, Gaoxiang Shi<sup>1</sup>, Bo Zhao<sup>2</sup>, Xin Jin<sup>1\*</sup>, Liming Zhou<sup>2</sup>, Tengfei Ma<sup>1</sup>

(1. College of Agricultural Equipment Engineering, Henan University of Science and Technology, Luoyang 471003, Henan, China;

2. China Academy of Agricultural Mechanization Science, Beijing 100083, China)

**Abstract:** Accurate fault prediction is essential to ensure the safety and reliability of combine harvester operation. In this study, a combine harvester fault prediction method based on a combination of stacked denoising autoencoders (SDAE) and multi-classification support vector machines (SVM) is proposed to predict combine harvester faults by extracting operational features of key combine components. In general, SDAE contains autoencoders and uses a deep network architecture to learn complex non-linear input-output relationships in a hierarchical manner. Selected features are fed into the SDAE network, deep-level features of the input parameters are extracted by SDAE, and an SVM classifier is then added to its top layer to achieve combine harvester fault prediction. The experimental results show that the method can achieve accurate and efficient combine harvester fault prediction. In particular, the experiments used Gaussian noise with a distribution center of 0.05 to corrupt the test data samples obtained by random sampling of the whole population, and the results showed that the prediction accuracy of the method was 95.31%, which has better robustness and generalization ability compared to SVM (77.03%), BP (74.61%), and SAE (90.86%).

**Keywords:** fault prediction, combine harvester, stacked denoising autoencoders, support vector machines

**DOI:** 10.25165/ijabe.20221502.6963

**Citation:** Qiu Z M, Shi G X, Zhao B, Jin X, Zhou L M, Ma T F. Fault prediction of combine harvesters based on stacked denoising autoencoders. *Int J Agric & Biol Eng*, 2022; 15(2): 189–196.

## 1 Introduction

At present, agricultural machinery and equipment are constantly developing in the direction of large-scale, intelligent and high-speed. Combine harvesters have been widely used as the main agricultural machinery harvesting equipment. Owing to unavoidable malfunction and downtime of the combine in the process of operation, fault diagnosis is of great significance to the reliability and safety of the combine harvester<sup>[1-3]</sup>.

In general, fault diagnosis methods are divided into the followings, signal-based threshold judgement, algorithmic model-based methods and composite methods<sup>[4]</sup>. Conventional combine harvester fault monitoring relies on the operator and a sensor-based threshold judgement method. This method has problems with high operator skill requirements and delayed alarms causing damage to the combine. The health of combine harvester during operation has a direct impact on harvest yield and service life, which is the reason why many experts have conducted in-depth research into methods for monitoring and diagnosing faults in

agricultural machinery<sup>[5-7]</sup>. For example, Zhang et al.<sup>[8]</sup> proposed a new intelligent defect detection framework based on time-frequency transformation, which can be effectively applied to agricultural equipment fault detection by adaptively mining available fault features. Zheng et al.<sup>[9]</sup> optimized the video monitoring technology of agricultural machinery based on the artificial neural network algorithm. The method realized the scientificity and transparency of the agricultural machinery fault identification data, ensured the scientificity of agricultural machinery fault identification and improved the efficiency of agricultural machinery management by establishing an online monitoring and visualization detection platform for agricultural machinery faults without real-time monitoring records. Gupta et al.<sup>[10]</sup> propose an on-board multi-signal classification algorithm and pseudo-spectral analysis for the analysis of hydraulic system faults in agricultural machinery on mobile equipment. Wattanajitsiri et al.<sup>[11]</sup> used FMEA techniques to perform a risk assessment of key components of combine harvester, indicating the causes and effects of faults. The risk priority number was rated and preventive maintenance strategies were proposed.

In recent years, fault monitoring and diagnosis systems based on principles such as deep learning and neural networks have been widely used in the field of machinery fault diagnosis. The algorithms trained by multivariate balanced data sources can make combine harvester faults more accurate and generalized. Janotta et al.<sup>[12]</sup> used the temperature as the monitoring object and proposed to build a mechatronic system to monitor the bearing temperature of combine harvester and warn of exceeding the temperature threshold. Xiao et al.<sup>[13]</sup> proposed a multi-group co-evolutionary particle swarm optimization BP neural network-based fault diagnosis method for diesel engines of agricultural machinery. A Mohammed et al.<sup>[14]</sup> used FBG sensors for monitoring the condition of bearings in motor operation, using the time and frequency domain

**Received date:** 2021-08-01 **Accepted date:** 2021-12-10

**Biographies:** **Zhaomei Qiu**, PhD Associate Professor, research interests: modern agricultural equipment and intelligence, Email: qzm381@126.com; **Gaoxiang Shi**, MS candidate, research interests: agricultural engineering and information technology, Email: sgx19031904@163.com; **Bo Zhao**, Research Fellow, research interests: image processing and agricultural machinery intelligence, Email: zhaoboshi@126.com; **Liming Zhou**, Research Fellow, research interests: precision agriculture and measurement and control technology research, Email: Haibo1129@163.com; **Tengfei Ma**, MS candidate, research interests: agricultural mechanization engineering, Email: mtf20032004@163.com.

**\*Corresponding author:** **Xin Jin**, Associate Professor, research interests: agricultural equipment and intelligent technology. Henan University of Science and Technology, Luoyang 471003, Henan, China. Tel: +86-379-64877817, Email: jx.771@163.com.

fault characteristics of the bearings to enable fault diagnosis and trend prediction. Shi et al.<sup>[15]</sup> proposed an improved Hilbert-Huang transform (IHHT) combined with convolutional neural network (CNN) for the diagnosis of rolling bearing composite faults in agricultural machinery and equipment, and realized the intelligent diagnosis of bearing composite faults under different working conditions. Trinh et al.<sup>[16]</sup> proposed a new framework based on a collection of genetic algorithms (GA) that can be used to automatically optimize the optimal parameter values for each submodule by combining them with specific machine learning algorithms, which can be used for the classification of failure types in mechanical systems and for the prediction of the remaining life of machinery. Chang et al.<sup>[17]</sup> developed a fault diagnosis methodology using the signals from a motor servo driver. Based on the servo driver signals, the demagnetization fault diagnosis of permanent magnet synchronous motors (PMSMs) was implemented by using an autoencoder and K-means algorithm. Liu et al.<sup>[18]</sup> proposed an intelligent fault diagnosis method combining local mean decomposition (LMD) and stacked noise reduction self-encoder (SDAE). The Fault Diagnosis Method for Reciprocating Compressors was implemented. The SDAE model-based approach has also been successfully applied to a number of areas such as remote sensing image detection, lithium battery monitoring, bearing fault diagnosis and power and energy system fault detection<sup>[19-22]</sup>.

To improve the accuracy of combine harvester fault prediction, this paper proposes a combine harvester fault prediction model, which combines a stack denoising autoencoder (SDAE)<sup>[23,24]</sup> with a multiclassification SVM classifier, using multiple denoising autoencoders with SDAE stacks to extract deep-level features of the data and establish a nonlinear between various fault data and the fault type of the combine mapping. To enhance the robustness of the model, balanced sample data and Gaussian noise were used in the model training to enhance the generalization ability of the model and to get rid of the reliance on manual prediction experience.

## 2 Data and models

### 2.1 Study Subjects

The Xinjiang-7 combine harvester (Luoyang Zhongshou Machinery Equipment Co., Ltd, China) was selected as the object of study for the experiment, by collecting data on the speed and temperature of the combine's components during the wheat harvesting process as a combine fault prediction feature. Based on the combine's drive roadmap and combine fault conditions, it was determined that excessive speed, excessive feed and lack of power due to belt slippage were the main causes of combine fault. The sensor assembly is shown in Figure 1.



1. Cutter 2. Feeding auger 3. IPC 4. Threshing cylinder 5. Secondary Stirring 6. Rethresher 7. Vibrating Sieve 8. Conveyor chain harrow

Figure 1 Diagram of the sensor installation position

To provide a more comprehensive and accurate picture of the combine's working conditions, speed sensors are installed to detect threshing cylinder speed, conveyor chain harrow speed, feed auger speed, secondary Stirring speed, rethresher speed and cutter reciprocation frequency, and temperature sensors are installed to detect threshing cylinder and rethresher shaft temperature. The speed sensor uses the Aotero TRD18-7DN with short-circuit protection, reverse polarity protection and IP67 protection and operates at DC10-30 V. The temperature sensor uses the SMD Amsik WZP-PT100 with IP68 protection. The data acquisition module transmits the sensor data to the on-board embedded IPC. The controller adopts Delta DVP40R2 PLC controller with rated voltage AC100-240V. The maximum output pulse string can reach 100 kHz with Delta's high-performance expansion unit, which can meet the sensor input point requirements. The combine harvester information processing adopts 8-inch touch multifunctional traveling computer with i5-6200U processor, 4GB memory, 128GB storage and RS485/232 communication interface. The communication between the upper computer and PLC adopts Modbus communication protocol, which can meet the needs of data transmission and processing.

### 2.2 Test protocol and methodology

In order to test the performance of the model in a real operating environment, a wheat harvesting test of the combine harvester was conducted at the Matun test field, Mengjin County, Luoyang City in June 2020. It was found that the combine harvester was not prone to failure at normal harvesting speed, so the combine harvester was tested for characteristic failure phenomena by human intervention with the fault settings as shown in Table 1. Through the above test protocol, the speed and temperature information of the corresponding components of the combine harvester were obtained when different faults occurred. Among them, Threshing cylinder fault can be divided into slight blockage and severe blockage. Slight blockage refers to the situation that the speed of threshing cylinder is obviously decreasing and less than 20%, which can be solved by artificially reducing the driving speed or decreasing the feeding volume. Severe blockage refers to the situation that the speed is reduced by 20%, the temperature of the drive shaft is rapidly increased, and continuous operation may lead to threshing cylinder damage. In addition to threshing cylinder, a 20% reduction in speed of the components is considered as a fault.

Table 1 Fault setting of the combine harvester test

Fault position	Manual fault setting	Fault phenomenons
Rethresher	1. Place grain debris in the rethresher. 2. Reduce the rubbing plate of the rethresher.	The speed of the rethresher drops and not working
Cutter	1. Reduce stubble height. 2. Increase the gap between moving blade and fixed blade.	The speed of the Cutter drops and not working
Secondary Stirring	1. Increasing the feeding amount. 2. Put in crop stems.	The speed of the Secondary stirring reduced
Feeding auger	1. Reduced cutting table height. 2. Increase the speed of combine harvester.	Paddy stalk build-up and reduced feeding auger speed
Conveyor chain harrow	1. Increasing the feeding amount or driving speed.	The speed of the Conveyor chain harrow reduced
Vibrating Sieve	1. Adjusting nut of loose vibrating screen. 2. Placing excess debris in the vibrating screen.	Vibrating Sieve speed reduced or not working
Threshing cylinder	1. Reduce the speed of the plate tooth roller 2. Engine running at low throttle to increase driving speed	The speed of the threshing drum drop and the belt heats up

In order to avoid irreversible damage to the combine caused by artificially set faults, different fault data collection times were set for different parts during fault sample data collection. For the test, the data collection frequency was set to 5 Hz. To ensure the stability and generalization ability of the algorithm, the ratio of the number of healthy and faulty states in the sample was artificially controlled to be close to or at 1:1, and a five-day period was set as the test period to ensure the scale of the data

### 2.3 The proposed method

#### 2.3.1 Overview of the method

Figure 2 shows the architecture of the proposed method, which incorporates combine fault-feature selection and fault prediction. In this work, knowledge-based methods are used to select effective fault features for fault prediction, while an SDAE network is constructed to predict Combine harvester fault by using selected features as its input.

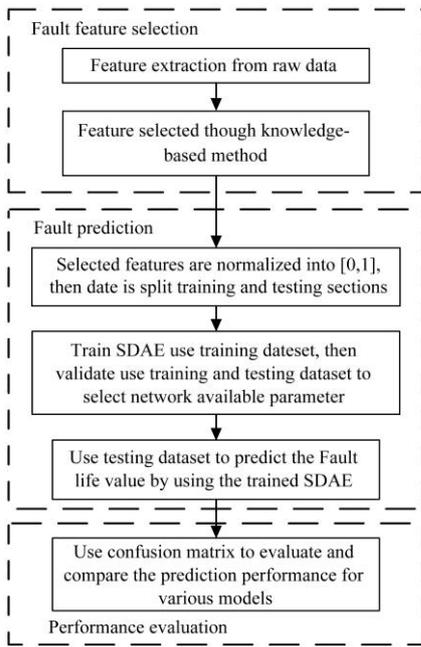


Figure 2 Architecture of the proposed method

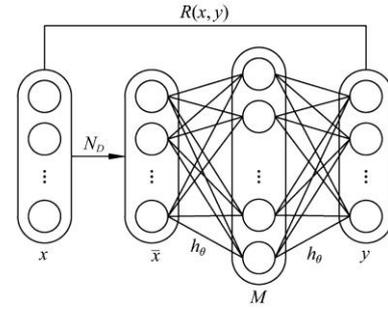
#### 2.3.2 SDAE model

The SDAE is based on a denoising auto-encoder (DAE) and improves the robustness of the auto-encoder (AE) by mixing the training data with noise. After normalizing the sample input data, the corrupted data  $\bar{x}$  is obtained by adding noise  $N_D$ . The corrupted data is encoded by the encoder  $f_\theta$  to obtain the deep feature  $M$ . The deep feature  $M$  is then decoded by the decoder  $h_\theta$  to obtain the reconstruction result  $y$  of the data sample  $x$ . The DAE model is obtained by minimizing the error  $R(x, y)$  between the data sample  $x$  and the reconstruction result  $y$  using the stochastic gradient descent (SGD) algorithm, where the DAE model is shown in Figure 3. The SDAE is then obtained by stacking multiple DAE models.

The SDAE model consists of three parts: the input layer, the hidden layer and the output layer.

**Input layer:** For a combine dataset,  $X=\{X_m\}(1\leq m\leq S)$  with  $S$  features for combine. The data  $X$  is corrupted by adding a Gaussian noise  $N_D(\bar{x}|X, \mu_0)$  given an initial Gaussian noise distribution centred on  $\mu_0$  to the original input data  $x$  to obtain the corrupted data  $\bar{x}$ , thus connected to the first hidden layer.

**Hidden layer:** Applied to extract features and reconstruct the original input from the unlabeled samples. There are multiple DAE units superimposed on it.



Note:  $x$  is the working parts' rotational speed of combine harvester, r/min;  $\bar{x}$  is  $x$  after adding noise, r/min;  $y$  represents the reconstructed data, r/min;  $R(x, y)$  is the mean square error between  $x$  and  $y$ , r/min;  $M$  is the deep feature;  $f_\theta$  represents encoder;  $h_\theta$  represent decoder;  $N_D$  is the add noise.

Figure 3 Basic structure of DAE

**Output layer:** After initial training of SDAE, the test dataset is used as input to SDAE and the deep features extracted after training are connected to the SVM classifier for fault prediction.

## 3 Experiments

### 3.1 SDAE model data pre-processing and initial training

The Combine harvester fault prediction process is divided into two parts: data preprocessing, and the use of SDAE for fault prediction.

**Data preprocessing:** The combine harvester data acquisition system is affected by the sensor's own drift and the environmental impact of data acquisition, and the raw data collected has 0 values, outliers and missing values, so it is necessary to pre-process the experimental data. If all values in a set of samples are 0, the sample will be deleted and any outlier that exceeds the maximum speed of the corresponding component will be replaced by linear interpolation according to the normal operating parameter range of the combine.

The 12 800 combine harvester data samples obtained from the pre-processing were divided in the ratio of 3:2 into a training set containing 7680 data samples and a test set of 5120 data samples. The data set was normalized to the input data using the maximum-minimum normalization method, which was calculated using the following Equation (1)

$$x_m = \frac{x_m - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

where,  $x_{\max}$  and  $x_{\min}$  are denoted as the maximum and minimum values of  $x$ , respectively. After normalisation, the numerical ranges of different combine fault features to  $[0, 1]$ .

Fault prediction is divided into SDAE pre-training and fault prediction.

Given an initial Gaussian noise distribution center on  $\mu_0$ , the Gaussian noise  $N_D(\bar{x}|X, \mu_0)$  with variance  $\sigma=1$  corrupts the original input data  $x$  to obtain the corrupted data  $\bar{x}$ , which is thus connected to the first hidden layer, where  $\bar{x} \sim N_D(\bar{x}|X, \mu_0)$ .

The noise-laden  $\bar{x}$  is encoded by Equation (2) and mapped to the nonlinear space  $M=M_m$  to obtain the deep feature encoding  $M_1 = f_\theta(\bar{x})$ .

$$M_1 = f_\theta(\bar{x}) = s_f(W_1\bar{x} + b_1) \quad (2)$$

where,  $W_1$  and  $b_1$  denote the connection weight matrix and bias term between the original data  $X$  and the hidden features  $M$ ,  $\theta$  is the set of parameters of the coded network and  $\theta=\{W, b\}$ ,  $s_f$  is the nonlinear activation function and the sigmoid function  $\delta=1/1+e^{-x}$ . In addition, the ReLU function  $Y=X$  performs a linear mapping of data, while the input data has non-linear characteristics.

Therefore, comprehensively considered, we chose the sigmoid function as the activation function for data mapping.

Decoding from  $M_1$  by Equation (3) and mapping it to nonlinear output space  $Y$ , the reconstructed output of original input data is obtained.

$$\hat{x} = h_{\theta}''(M_1) = s_h(W_1' M_1 + b_1') \tag{3}$$

where,  $W_1'$  is the decoder weight matrix;  $b_1'$  is the decoder bias vector;  $\theta'$  is the set of parameters of the decoding network;  $\hat{x}$  is the reconstructed output and  $s_h$  is the non-linear activation function.

Initial training of the denoising self-encoder is completed by using the gradient descent algorithm (SGD) to reduce reconstruction errors.

$$J(W, b) = \frac{1}{N} \sum_{l=1}^L \frac{1}{2} (Y - X)^2 \tag{4}$$

where,  $L$  is the total number of hidden layers. The parameter  $(W, b)$  is iteratively updated at  $l$ th hidden layer using Equations (5) and (6), respectively:

$$W_{l-1} = W_l - \eta \frac{\partial}{\partial W} J_{l-1} \tag{5}$$

$$b_{l-1} = b_l - \eta \frac{\partial}{\partial b} J_{l-1} \tag{6}$$

where,  $\eta$  is the learning rate. Now,  $\frac{\partial}{\partial W} J_{l-1}$  and  $\eta \frac{\partial}{\partial b} J_{l-1}$  are the partial derivatives of  $W$  and  $b$  in the SGD algorithm at the  $(l-1)$ th hidden layer.

When the reconstruction error of DAE1 model is small, the initial optimal weight  $W_1$  and the optimal bias  $b_1$  are obtained, and the initial deep feature  $M_1$  of the original input data is obtained. Similarly, the initial deep feature  $M_1$  is used to train DAE2, and its initial optimal weight  $W_2$ , optimal bias value  $b_2$  and feature  $M_2$  are obtained.

After the initial training of SDAE, label information from the input data was used to supervise the fine-tuning training of SDAE models. Supervised fine-tuning is to adjust the parameters of each layer obtained during the initial training process. The result of supervised fine-tuning will make each parameter of the model closer to the global optimal solution, thus enhancing the intra-class aggregation and inter-class differentiability of the features and improving the classification ability of the model.

### 3.2 Determination of network parameters

In this step, the 8 selected features are input to the SDAE network, and a multi-class classifier SVM was used to classify them for combine harvester fault prediction. In the SDAE data processing, the number of features of the input samples determines the input size of the input layer, e.g. if each input sample contains 8 selected features, the input size of the input layer is 8. For the structural setting of the SDAE network, references [25,26] show that the network structure with multiple hidden layers has a better representation of the data due to its ability to learn patterns from the raw data and represent the data more abstractly through layer-by-layer feature extraction, thus improving the accuracy of the classification.

While the accuracy increases as the number of hidden layers increases, it may decrease when the number of hidden layers is too large, so the network structure with four hidden layers {10, 8, 6, 4} was chosen to make a stable triangular structure. Compared with the hidden layers with the same number of nodes in each layer, the triangular structure network with a gradually decreasing number of nodes in the hidden layers had a higher accuracy rate.

Taking the selected 8 features as an example, the number of

neural nodes in the input layer is 8. Therefore, the number of neural nodes for input  $X$  and output  $Y$  in the first DAE unit is 8, and the neural node dimension for the extracted feature  $M$  in the 1st DAE unit is 10. Therefore, the structure of the 1st DAE unit is {8, 10, 8}, and then the 10 features extracted from the 1st DAE unit are selected as input to the 2nd DAE unit, setting the number of features extracted to 8. Therefore, the structure of the 2nd DAE unit is {10, 8, 10}, and in addition the rest of the structures of the 2nd DAE unit are {8, 6, 8}, {6, 4, 6}, {8, 10, 8}, {4, 2, 4}.

The determination of key hyperparameters of the SDAE network, such as the number of iterations, learning rate, denoising rate and batch size, can have a significant impact on the predictive performance of the model. In the process of hyperparameter determination, the batch size should be selected first before tuning the other hyperparameters. The batch size refers to the number of samples fed into the model at each training session. In the process of model optimization, a small batch size means that the number of samples fed into the model at a time is too small, the statistics are not represented, and the noise increases accordingly, making it difficult for the model to converge; a large batch size affects the generalization ability of the model and reduces accuracy. The batch size is 128 as the dividing line, and the test results are compared after training by scaling, and the final batch size is 64 after considering the hardware conditions and network training time.

The experiment used a single-factor test to analyze the effect of changes in learning rate, denoising rate, and number of iterations on the accuracy of the network to determine the optimal hyperparameters. When the other hyperparameters are the same, the number of iterations varies as [100, 500, 800, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10 000]. When the difference between the test error rate and the training error rate is small, the current number of iterations is considered appropriate; when the test error rate becomes smaller and then larger, the number of iterations is too large and needs to be reduced, otherwise, overfitting is likely to occur. By considering the prediction accuracy and time consumption, we set the number of iterations to 4000.

The learning rate, another important factor in the predictive performance of the network, can greatly reduce the speed of convergence and increase the time cost if it is too small; if it is too large, it may cause the parameters to oscillate on either side of the optimal solution. Setting the learning rate to vary as [0.001, 0.003, 0.005, 0.008, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, 0.2, 0.3], by considering the prediction accuracy, the learning rate was set to 0.02. Similarly, setting the range of denoising rate as [0.01, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1], comparing the magnitude of the prediction accuracy of the analysis network, it was found that the effect of denoising rate was not significant, so the denoising rate was set as 0.05. Finally, the detailed structure information about the SDAE model is listed in Table 2.

**Table 2 Parameter configuration for SDAE model**

Parameters	Values
Active function	Sigmoid
Total number of hidden layers	4
Structure of hidden layers	[10, 8, 6, 4]
Learning rate	0.02
Maximum iteration number	4000
Denoising rate	0.05
Batch size	64

### 3.3 Performance evaluation

The confusion matrix shown in Table 3 was established, and the accuracy of classification results is obtained by comparing the position and classification of each measured pixel with the corresponding position and classification in the classification image, so as to evaluate the fault prediction performance of combine.

**Table 3 Confusion matrix**

Confusion Matrix		True Value	
		Positive	Negative
Predicted value	Positive	TP	FP
	Negative	FN	TN

Each column in the graph represents the predicted category and the number of data in each column indicates the number of data predicted to be in that category. Each row represents the true category to which the data belongs and the number of data in each row indicates the number of instances of data in that category. The value in each column indicates the number of instances of real data predicted to be in that category. TP, TN, FP and FN in the table are the first level base indicators of the confusion matrix and the concepts are as follows:

True Positive, TP: The true category of the sample is a positive example and the model predicts a positive example. True Negative, TN: The sample's true category is negative example and the model predicts it as negative example. False Positive, FP: The true category of the sample is negative example, but the model predicts it to be positive example. False Negative, FN: The sample's true category is positive example, but the model predicts it as negative example.

In order to accurately judge the strengths and weaknesses of the model, the model is measured more accurately by comparing the four secondary indicators extended from the primary indicators and the tertiary indicators extended from the secondary indicators. The specific meaning and formulae are listed in Table 4.

**Table 4 Equations and significance of confusion matrix**

Abbreviation	Formula	Implication
ACC	$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$	The proportion of all correct results in the total observed values in the classification model.
PPV	$Precision = \frac{TP}{TP + FP}$	The weight of the model prediction pair in all results for which the model prediction is Positive.
TPR	$Recall = \frac{TP}{TP + FN}$	The weight of model prediction pairs among all results for which the true value is Positive.
TNR	$Specificity = \frac{TN}{TN + FP}$	The weight of model prediction pairs among all results for which the true value is Negative.
$F_1$ score	$F_1 \text{ score} = \frac{2 \times Precision \times Recall}{Precision + Recall}$	The reconciled average of the model's accuracy and recall.

### 3.4 Results and analysis

Table 5 shows the confusion matrix for fault classification and identification by SDAE method, which totally contains nine categories: normal, re-thresher blocked, cutter fault, secondary stirring blockage, feeding auger blocked, threshing cylinder serious blocked, threshing cylinder slightly blocked, conveyor chain rakes blocked, vibrating sieve fault denoted as normal,  $F_1$ ,  $F_2$ ,  $F_3$ ,  $F_4$ ,  $F_5$ ,  $F_6$ ,  $F_7$ , and  $F_8$  respectively. The horizontal axis of the confusion matrix represents the true categories and the vertical axis represents the predicted categories. The values on diagonal line represent the number of samples correctly classified by SDAE network in

each category of the test samples, and the values at off-diagonal positions represent the number of samples incorrectly classified by the network in each category. A total of 5120 samples were taken, and the ratio of normal samples to error samples was 1:1, which satisfied the requirements of balanced sampling, made the model more general and eliminated the influence of chance of the experiment to the greatest extent.

**Table 5 Confusion matrix of SDAE model**

Category	Normal	$F_1$	$F_2$	$F_3$	$F_4$	$F_5$	$F_6$	$F_7$	$F_8$
Normal	2536	3	0	2	9	0	6	4	0
$F_1$	13	213	0	3	0	0	7	0	0
$F_2$	3	0	187	0	0	0	0	0	0
$F_3$	7	0	0	225	4	0	11	2	0
$F_4$	5	0	0	5	400	0	10	6	0
$F_5$	0	0	0	0	0	497	5	0	0
$F_6$	7	0	0	6	0	8	281	4	0
$F_7$	2	0	0	0	14	5	19	369	0
$F_8$	1	0	0	0	0	0	0	0	241

When the actual category is normal, the number of samples misclassified to  $F_1$  is 3, and the number of samples misclassified to  $F_3$ ,  $F_4$ ,  $F_6$  and  $F_7$  are 2, 9, 6 and 4, respectively. When the actual category is  $F_1$ , the number of samples misclassified to be normal is 13, and the number of samples misclassified as  $F_3$  and  $F_7$  is 3 and 7 respectively. When the actual category is  $F_2$ , the number of samples misclassified to be normal is 3. When the actual category is  $F_3$ , the number of samples misclassified to be normal is 7, and the number of samples misclassified as  $F_4$ ,  $F_6$  and  $F_7$  is 4, 11 and 12 respectively. When the actual category is  $F_4$ , the number of samples misclassified to be normal is 5, and the number of samples misclassified as  $F_3$ ,  $F_6$  and  $F_7$  are 5, 10 and 6 respectively. When the actual category is  $F_5$ , the number of samples misclassified to be normal is 0, and the number of samples misclassified as  $F_6$  is 5. When the actual category is  $F_6$ , the number of samples misclassified to be normal is 7, and the number of samples misclassified as  $F_3$ ,  $F_5$  and  $F_7$  When the actual category is  $F_7$ , the number of samples misclassified to be normal is 2, and the number of samples misclassified as  $F_4$ ,  $F_5$  and  $F_6$  is 14, 5 and 19 respectively. When the actual category is  $F_1$ , the number of samples misclassified to be normal is 1, and the number of samples misclassified as other is 0.

Where the correct classification rates for normal, rethresher blocked, cutter fault, secondary stirring blockage, feeding auger blocked, threshing cylinder serious blocked, threshing cylinder slightly blocked, conveyor chain rakes blocked and vibrating sieve fault were 99.10%, 90.25%, 98.42%, 90.36%, 93.90%, 99.00%, 91.83%, 90.22%, and 99.59%, respectively, and the overall fault recognition accuracy was also as high as 96.66%.

Figure 4 shows the results of the fault sample identification. The experimental results show that the prediction accuracy of the model for faulty samples is significantly lower than the prediction accuracy for normal samples.

Among them, the misjudgment of slight blockage of threshing cylinder is one of the main reasons for the low prediction accuracy of faulty samples. There is no significant difference in the transition data from slight blockage to severe blockage of threshing cylinder, which leads to a decrease in the prediction accuracy. The cutter and vibrating sieve are located at the head and tail ends of the combine harvester drive structure, where failure is evident and therefore the prediction rate for cutter fault and vibrating sieve failure is high. The main reason is that when the threshing cylinder is severely blocked, the conveyor chain rakes located in

front of the threshing cylinder and connected with it are directly affected by the blockage, which results in the prediction error. The secondary stirring is misclassified as slightly blocked threshing cylinder, mainly because the threshing cylinder is slightly blocked, while the secondary stirring directly connected to the threshing cylinder receives more debris. The secondary stirring receives more debris, which is easy to block. The feeding auger is blocked due to overfeeding and the corresponding conveyor chain rakes are blocked due to its influence, so some data of the blocking of the feeding auger blocked is incorrectly judged as the blocking of conveyor chain rakes blocked. For rethresher blockage, there are also misjudgments. Most of them are being slightly blocked by the threshing cylinder and few are misjudged as being blocked by secondary stirring. The main reason is that secondary stirring blockage is a major cause of the blockage of the threshing cylinder, and a slight blockage of the threshing cylinder will affect the normal operation of the rethresher.

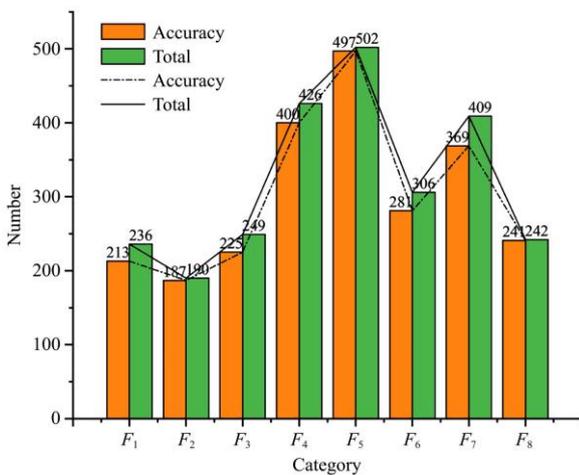


Figure 4 Fault sample prediction results

#### 4 Comparison analysis and discussion

To verify the effectiveness of SDAE, we adopted various models to predict the fault of combine harvester, including Support SVM, Back Propagation Neural Network (BP), Sparse Auto Encode (SAE) and SDAE. The hyperparameters of these models are summarized in Table 6.

Table 6 Hyperparameters of various models

Models	Parameters
SVM	<ol style="list-style-type: none"> <li>1. Penalty parameter c: 2.8633</li> <li>2. Kernel parameter g: 4.5423</li> <li>3. Kernel function: radial basis function</li> </ol>
BP	<ol style="list-style-type: none"> <li>1. Input size: 8</li> <li>2. Output size: 8</li> <li>3. Number of neural nodes at hidden layer: 12</li> <li>4. Learning rate: 0.02</li> </ol>
SAE	<ol style="list-style-type: none"> <li>1. Input size: 8</li> <li>2. Active function: sigmoid</li> <li>3. Total number of hidden layers: 4</li> <li>4. Structure of hidden layer: [10 8 6 4]</li> <li>5. Learning rate: 0.02</li> <li>6. Maximum iteration number: 4000</li> <li>7. Output size: 8</li> </ol>
SDAE	<ol style="list-style-type: none"> <li>1. Input size: 8</li> <li>2. Active function: sigmoid</li> <li>3. Total number of hidden layers: 4</li> <li>4. Structure of hidden layer: [10 8 6 4]</li> <li>5. Learning rate: 0.02</li> <li>6. Maximum iteration number: 4000</li> <li>7. Denoising rate: 0.05</li> <li>8. Output size: 8</li> </ol>

The SDAE model in this paper is compared with SAE networks, BP neural networks and SVM for analysis. To ensure the accuracy of the experimental results, the same 5120 data samples as the SDAE model experiments were selected for the experiments. The prediction accuracies of the four models are shown in Table 7.

Table 7 Identification results of different models

Model	Accuracy/%	Normal sample accuracy/%	Fault sample accuracy/%
SVM	84.04	87.78	80.31
BP	80.12	87.50	72.73
SAE	90.20	94.53	85.86
SDAE	96.66	99.10	94.26

The experimental results show that the prediction accuracy of SDAE with triangular hidden layer structure is 6.46% higher than that of SAE model with the same number of hidden layers., and 16.54% and 12.62% higher than that of BP and SVM models respectively. During normal operation of the combine harvester, the speed and temperature information corresponding to the components changes steadily and feature extraction is easier. In view of the transmission structure of the combine harvester, in the event of a combine harvester fault, the components will interact with each other, leading to difficulties in fault feature extraction, so the SDAE model is significantly more accurate in identifying fault samples than the rest of the models.

The harsh operating environment and complex working process of combine harvesters lead to abnormalities in their data collection and transmission processes, as well as the influence of factors such as the drift of the sensors themselves, all of which will make the fault characteristics of the system random, fuzzy and uncertain, deep data mining is needed to achieve effective fault identification. The quality and size of the dataset largely determine the performance of deep learning fault diagnosis methods. The SDAE model in this paper takes a noise-additive approach to the original dataset and tests the denoising ability of the algorithm using a Gaussian noise contaminated signal. The performance of the SDAE model for combine harvester fault prediction was verified by introducing Gaussian noise with a distribution center  $\mu_0$  of 0.05 during testing as shown in Table 8, compare the recall, specificity, F1 score and average recognition accuracy of different models for samples.

Table 8 Recall rate, specificity, F1 score and noise addition accuracy results of different models

Model	Recall/%	Specificity/%	F1 Score	Noise addition accuracy/%
SVM	64.84	93.13	0.704	77.03
BP	62.34	91.88	0.679	74.61
SAE	87.03	96.88	0.889	90.86
SDAE	92.50	100.31	0.939	95.31

After adding the noise treatment, the recognition accuracy of all four models decreased by different magnitudes, with the SDAE model in this paper only decreasing by 0.09%, while the other models decreased significantly. The SDAE model outperformed the SVM, BP and SAE models by 27.66%, 30.16% and 5.47% respectively in predicting the accuracy of faulty samples. In terms of Specificity, the SDAE model is 100.31%, indicating that the model has a high prediction accuracy for non-faulty samples and there is a phenomenon that the prediction of faulty samples is misjudged as normal operation. F<sub>1</sub> Score, as the four secondary indicators extended from the primary indicators and the tertiary

indicators extended from the secondary indicators, the size of the score can more accurately measure the goodness of the model, and the results show that The  $F_1$  Score of the SDAE model is 0.939, which is a significant performance improvement compared to SVM (0.704), BP (0.679) and SAE (0.889).

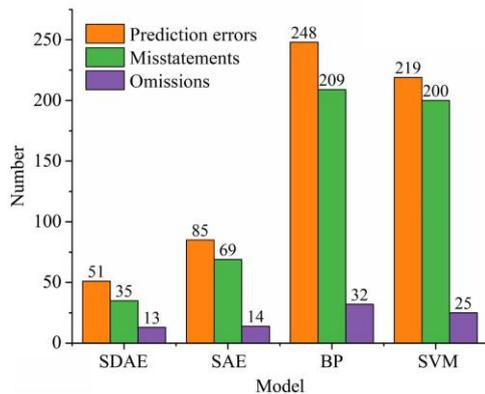


Figure 5 Identification results of different models

Figure 5 shows the distribution of the number of errors predicted by the different models. In particular, the BP and SVM models are much higher than the SAE and SDAE models in terms of the number of fault errors. However, there is no significant deterioration in the normal misclassification as faults compared to the situation before the inclusion of noise. The experimental results show that the SDAE model outperforms the other models in terms of the number of fault misses and fault false positives. The reasons for the large gap between the four models are mainly three:

(1) Different from the “shallow” network model, the model is trained to integrate fault feature extraction and classification models; (2) The model is a multi-hidden layer network, which can avoid the limitations of “dimensional disaster” and insufficient prediction capability; (3) The model can learn the global coarse-grained features and local fine-grained features of the operating condition data set of the combine harvester, and has strong robustness. The experimental results show that the model effectively reduces the influence of load and noise factors on the accuracy of the combine harvester in actual operation, thus improving the accuracy of combine harvester fault prediction.

## 5 Conclusions

In this paper, a fault prediction model of combine harvester based on a deep SDAE neural network was proposed. By extracting the operational characteristics of key combine harvester components in SDAE network and inputting a multi-class SVM classifier to classify and identify the fault of combine harvester, a deep-level representation of the fault state features of combine harvester was learned. The fault prediction method proposed in this study had significant advantages in terms of data processing and classification accuracy. The main conclusions are as follows.

(1) By comparing the model recognition accuracy and training time under different model structures and hyperparameters, the optimal connection weight matrix, bias term, number of nodes in the hidden layer, hidden layer structure, number of iterations, learning rate, denoising rate and batch size were selected. For a specific problem of combine harvester fault diagnosis, a feature extraction based on deep learning was proposed. Combining SDAE network with multi-class SVM, a model based on SDAE feature extraction and SVM classification was established, which provided a new method for combine harvester operation fault prediction.

(2) The SDAE model was trained to ensure the accuracy of the test accuracy by ensuring the scale and quality of the data during sample data collection. Ensure the integrity, comprehensiveness, legality and uniqueness of the data during sample data cleaning; and ensure a balanced sampling during sample selection. The experimental results showed that the prediction accuracy of the model is 96.66%, which proved that the model had a good effect in the application of combine harvester fault prediction.

(3) Adding Gaussian noise with distribution center of 0.05 to the test sample data of whole group random sampling, the recognition accuracy of SDAE model was 95.31%, which was 20.7%, 18.28% and 4.45% higher than that of BP neural network, SVM and SAE models respectively.  $F_1$  scores were 0.26, 0.235 and 0.05 respectively. The experiments showed that the model can effectively reduce the influence of load and environmental noise on the model accuracy during the actual operation of the combine harvester, thus improving the fault prediction accuracy and extending the operating environment of the combine harvester.

## Acknowledgements

The work was sponsored by the Intelligent Manufacturing Comprehensive Standardization Project (No. 2018GXZ1101011), the National Key Research and Development Program of China Sub-project (No.2016YFD0701802), the Natural Science Foundation of Henan (No.202300410124).

## [References]

- [1] Wang L, Wang Y, Dai D, Wang X, Wang S M. Review of electro-hydraulic hitch system control method of automated tractors. *Int J Agric & Biol Eng*, 2021; 14(3): 1–11.
- [2] Yılmaz D, Gkduman M E. Development of a measurement system for noise and vibration of combine harvester. *Int J Agric & Biol Eng*, 2020; 13(6): 104–108.
- [3] Luo X W, Liao J, Hu I, Zang Y, Zhou Z Y. Improving agricultural mechanization level to promote agricultural sustainable development. *Editorial Office of Transactions of the CSAE*, 2016; 32(1): 1–11. (in Chinese)
- [4] Gao Z, Cecati C, Ding S X. A survey of fault diagnosis and fault-tolerant techniques—Part I: fault diagnosis with model-based and signal-based approaches. *IEEE Transactions on Industrial Electronics*, 2015; 62(6): 3757–3767.
- [5] Cecchini M, Piccioni F, Ferri S, Coltrinari G, Colantoni A. Preliminary investigation on systems for the preventive diagnosis of faults on agricultural operating machines. *Sensors*, 2021; 21(4): 1547. doi: 10.3390/s21041547.
- [6] Kamilaris A, Prenafeta-Boldu F X. Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, 2018; 147: 70–90.
- [7] Cheng L, Zhang Y. Analysis of intelligent agricultural system and control mode based on fuzzy control and sensor network. *Journal of Intelligent and Fuzzy Systems*, 2019; 37(2): 1–12.
- [8] Zhang Z W, Chen H H, Li S M, Wang J R. A novel sparse filtering approach based on time-frequency feature extraction and softmax regression for intelligent fault diagnosis under different speeds. *Journal of Central South University*, 2019; 26(6): 1607–1618.
- [9] Zheng J X, Li M, Hu S K, Xiao X W, Li H, Li W F. Research on optimization of agricultural machinery fault monitoring system based on artificial neural network algorithm. *INMATEH-Agricultural Engineering*, 2021; 64: 297–306.
- [10] Gupta S, Khosravy M, Gupta N, D Ar B Ar I H, Patel N. Hydraulic system onboard monitoring and fault diagnostic in agricultural machine. *Brazilian Archives of Biology and Technology*, 2019; 62(3): e19180363. doi: 10.1590/1678-4324-2019180363.
- [11] Wattanajitsiri V, Kanchana R, Triwanapong S, Kimapong K. Identifying preventive maintenance guideline for a combine harvester with application of fault mode and effect analysis technique. *MATEC Web of Conferences*, 2020; 319: 01004. doi: 10.1051/mateconf/202031901004.

- [12] Janotta R, Podsdek S, Bartoszuk M. The concept of a mechatronic system for monitoring the temperature of bearings in a combine harvester. 2nd International Conference on Chemistry, Chemical Process and Engineering (IC3PE), 2018. doi: 10.1063/1.5066485.
- [13] Xiao M H, Wang W C, Wang K X, Zhang W, Zhang H. Fault diagnosis of high-power tractor engine based on competitive multiswarm cooperative particle swarm optimizer algorithm. *Shock and Vibration*, 2020; 2020: 1–13.
- [14] Mohammed A, Djurovic S. Electric machine bearing health monitoring and ball fault detection by simultaneous thermo-mechanical fiber optic sensing. *IEEE Transactions on Energy Conversion*, 2021; 36(1): 71–80.
- [15] Shi J, Wu X, Liu T. Bearing compound fault diagnosis based on HHT algorithm and convolution neural network. *Transactions of the CSAE*, 2020; 36(4): 34–43. (in Chinese)
- [16] Trinh H C, Kwon Y K. A data-independent genetic algorithm framework for fault-type classification and remaining useful life prediction. *Applied Sciences*, 2020; 10(1): 368. doi: 10.3390/app10010368
- [17] Chang L K, Wang S H, Tsai M C. Demagnetization fault diagnosis of a PMSM using auto-encoder and K-means clustering. *Energies*, 2020; 13(17): 4467. doi: 10.3390/en13174467.
- [18] Liu Y, Duan L, Yuan Z, Wang N, Zhao J. An intelligent fault diagnosis method for reciprocating compressors based on LMD and SDAE. *Sensors*, 2019; 19(5): 1041. doi:10.3390/s19051041.
- [19] Li S, He H, Li J. Big data driven lithium-ion battery modeling method based on SDAE-ELM algorithm and data pre-processing technology. *Applied Energy*, 2019; 242: 1259–1273.
- [20] Lu N, Chen C, Shi W, Zhang J, Ma J. Weakly supervised change detection based on edge mapping and SDAE network in high-resolution remote sensing images. *Remote Sensing*, 2020; 12(23): 3907. doi:10.3390/rs12233907.
- [21] Rizwan-ul-Hassan, Li C, Liu Y. Online dynamic security assessment of wind integrated power system using SDAE with SVM ensemble boosting learner. *International Journal of Electrical Power & Energy Systems*, 2021; 125: 106429. <https://doi.org/10.1016/j.ijepes.2020.106429>.
- [22] Chen L, Wang Z Y, Qin W L, Ma J. Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification. *Signal Processing*, 2017; 130: 377–388.
- [23] Vincent P, Larochelle H, Lajoie I. Stacked de-noising auto-encoders: learning useful representations in a deep network with a local de-noising criterion. *J Mach Learn Res* 2010; 11: 3371–408.
- [24] Vincent P, Larochelle H, Lajoie I, Bengio Y, Manzagol P A. Extracting and composing robust features with de-noising auto-encoders. *Proceedings of the Twenty-Fifth Int Conference on Machine Learning*, 2008; pp. 1096–103. doi:10.1145/1390156.1390294.
- [25] Cai Z, Long Y, Shao L. Classification complexity assessment for hyper-parameter optimization. *Pattern Recognition Letters*, 2019; 125(7): 396–403.
- [26] Yang L, Shami A. On Hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing*, 2020; 415: 295–316.