Research on Anomaly Detection Method of Non-Standard Machinery Production Process Based on Semi-Supervised Learning

Longlong Lv[1], Helong Qiu[1], Huihui Wu[2]
Jun Dou[1], Qiang Cao[1], Qiyuan Zhang[1], Wenchao Zhao[1], and Xinghui Zhu[1]
[1]School of Computer Science, Xidian University, Xi'an, Shaanxi, 710162, China
[2]Information Science and Technology (BNRist), Department of Automation, Tsinghua University, Beijing 100084, China

The non-standard machinery refers to customized machinery produced to meet specific customer demands. The mainstream research direction in data stream anomaly detection focuses on deep learning, which involves learning data distribution through a large amount of training data. However, non-standard machinery equipment has the characteristics of a small production scale and sparse samples, making it difficult to obtain sufficient annotated training sets. This inadequacy in training data results in the model not learning enough, thereby rendering it unable to effectively detect abnormal events. In this paper, we propose a semi-supervised learning (SSL) based anomaly detection method. We employ a hybrid C-LSTM network based on the self-attention mechanism as an abnormality prediction model, where the convolutional neural network (CNN) and long short-term memory network (LSTM) extract spatiotemporal features of industrial data streams. The self-attention mechanism calculates the relationship weights between different positions in the input data, capturing long-term dependencies in time series data to fully learn data distribution. To improve the training effectiveness of the prediction model, we use an updating algorithm based on weighted fuzzy rough set (WFDA) to update the prediction model in a reverse manner. This algorithm can classify data streams in real-time, compare the classification results of the prediction model, and retrain unreliable data. The experimental results show that our proposed method achieves an F1 score of 0.955 and a recall value of 0.957 on a real-world data set, which is a 4.1% improvement in F1 score and a 6.4% improvement in recall compared to similar anomaly detection algorithms that do not use our proposed method.

Index Terms—**Non-standard achinery anomaly detection, Semi-supervised learning, Fuzzy-rough-set, CNN, LSTM**

I. Introduction

INDUSTRIAL Internet of Things (IIOT) refers to the application of IoT technology in the industrial field. Non-standard machinery production is a specific scenario under IIOT, which provides enterprises with a comprehensive method of monitoring the production process [1], optimizing resource allocation, and reducing costs by covering various customized machinery equipment, sensors, and computer systems.

With the vigorous development of China's manufacturing industry, various industrial sensors are widely used in non-standard mechanical equipment, including non-standard environmental monitoring equipment [2], customized CNC (Computerized Numerical Control) machine tools [3], and industrial production workshops [4], [5]. How to ensure the safety and stability of industrial production is the focus of research. Anomaly detection of industrial equipment is an effective solution, effective anomaly data detection technology is of great significance for ensuring the reliability of non-standard mechanical production processes and decision-making support for management departments.

The goal of our work is to provide practical solutions to the production process of non-standard mechanical equipment. Concretely, there is a need for an anomaly detection method that can operate with high real-time performance and low false positive rate in non-standard mechanical production environments, to meet the practical requirements of industrial production.

Manuscript received June 13, 2023; revised August 2, 2023. Corresponding author: Xinghui Zhu (email: xhzhu@xidian.edu.cn).

The challenges of our study are two-fold: Firstly, In the production process of non-standard machinery, there are numerous environmental noises such as electromagnetic interference, machine vibration, temperature, humidity changes, etc. Secondly, the difficulty in obtaining a large amount of pre-labeled data for certain non-standard mechanical equipment is another challenge. The combination of deep learning models and time series data for anomaly detection has been a hot research topic. However, common methods require a large amount of pre-training for deep learning models to learn the distribution characteristics of the data. In summary, the contributions in this paper are:

- We propose a model update algorithm based on WFDA (Weighted Fuzzy-Rough Density Algorithm), an anomaly detection algorithm suitable for non-standard mechanical production environments, with good noise resistance and anomaly detection performance.
- We propose a semi-supervised anomaly detection method that employs a hybrid C-LSTM network based on a self-attention mechanism as the anomaly prediction model, and combines it with the WFDA-based update algorithm to reverse-update the prediction model. This method can process data streams in real-time for data classification, compare the classification results of the prediction model, and retrain untrustworthy data.

The remainder of this paper is organized as follows. Section II presents the related work. Section III gives a brief introduction to C-LSTM and fuzzy rough set theory and a description of the System Architecture and Module Design. In Section IV, We conducted an experimental implementation of the proposed methodology and performed a theoretical analysis of the experimental results. Finally, the paper is concluded in Section V.

II. Related Works

This section categorizes and summarizes the relevant techniques in anomaly detection, and analyzes their advantages and disadvantages.

There has been a flurry of recent works in the area of abnormal detection. Generally, the current research in the field is mainly focused on predictive-based anomaly detection algorithms, which are capable of discovering abnormal samples in a dataset that do not conform to expected behavior, by forecasting possible anomalies based on historical data. This approach has wide-ranging applications in fields such as financial risk prediction, medical diagnosis, and industrial equipment failure. The commonly used predictive-based anomaly detection methods are based on machine learning models, which aim to learn the feature distribution of normal data from the dataset and use these features to detect anomalous points. Various machine learning algorithms, such as Support Vector Machine (SVM) [6] and Random Forests [7], can be employed in this approach. Suleman Shahzad used a unique sequential deep learning method based on SVM to detect anomalies in sperm samples [8], while Qingyang Zhang achieved good results in detecting financial data anomalies by introducing Random Forests algorithm [9].

With the development of deep learning, some researchers have used deep neural networks to model data and adaptively learn the structure of the data in order to detect anomalies in new data. The core idea of these methods is to extract high-level features from the data by multiple layers of nonlinear transformations, capturing the distribution of the data. For example, HOPFIELD J J first proposed recurrent neural networks (RNN) [10], which consist of a large number of equivalent components, i.e., neurons, to form a physical system with interactions, achieving significant progress in time series prediction and causing a huge sensation in related fields, but ignoring the long-term dependence between time series data. HOCHREITER S and SCHMIDHUBER J. proposed long short-term memory (LSTM) [11], [12], an improved model to solve the problems of gradient vanishing and exploding in RNN for long sequence data. LSTM, as an improved model of RNN, has the advantages of strong long-term memory capability, flexible model structure, fast training speed, and wide application. Furkan Elmaz et al. combined convolutional neural networks (CNN) and LSTM to extract complex features, using the CNN layer to extract spatial features in time series data, and further using LSTM to extract temporal features, effectively predicting indoor temperature. Zhuqing Wang et al. effectively predicted the remaining useful life (RUL) of lithium-ion batteries using LSTM based on adaptive self-attention mechanism [13].

Currently, cutting-edge deep learning-based anomaly prediction algorithms require a large amount of pre-training data sets, which can be difficult to obtain in industrial Internet of Things (IoT) settings where it may be challenging to acquire large amounts of pre-labeled data and where the labels may not be accurate enough to extract sufficiently accurate and informative features. To reduce dependence on annotated training sets, scholars have conducted extensive research on this issue. Aldo Glielmo and others proposed an unsupervised learning method for molecular simulation data [14] that enables online learning without the need for a pre-training set. However, unsupervised learning still faces challenges in eliminating the impact of complex noise. Vikas Verma and others proposed a semi-supervised method [15] using interpolation consistency training that can effectively train deep neural network models and achieve excellent performance in classification problems, but the method's effectiveness in extracting spatial features of high-dimensional data is limited. This article focuses on anomaly detection technology and proposes a semi-supervised anomaly detection solution based on fuzzy rough granularity and a hybrid C-LSTM neural network (HFC-LSTM) that combines a mixed C-LSTM neural network with a self-attention mechanism and the WFDA anomaly detection algorithm based on fuzzy rough granularity. This approach allows for real-time online learning of the spatiotemporal features of new anomalous data with a small training set, further improving the accuracy of anomaly identification. The proposed solution is well-suited to industrial IoT environments and has been demonstrated in multiple industrial IoT scenarios with good detection results. This research provides an effective solution for anomaly detection in non-standard mechanical industrial production environments and has practical application value.

III. Semi-Supervised Hybrid Network

In this section, we first propose a semi-supervised anomaly detection method (HFC-LSTM) based on a hybrid C-LSTM network and the WFDA algorithm. Then, we use a hybrid C-LSTM model with a self-attention mechanism as the prediction model and design a model update algorithm based on the fuzzy rough set theory of the WFDA algorithm proposed in the previous chapter. The algorithm updates the prediction model via backpropagation to learn new anomalies and further improve the detection success rate.

A. System Model Design

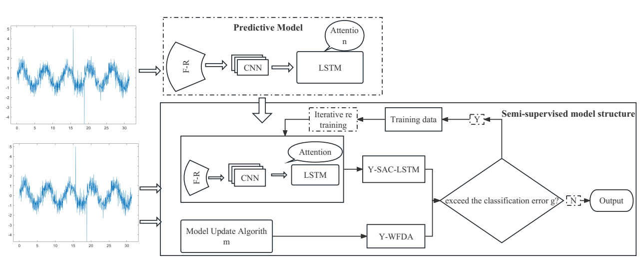


Fig. 1. Semi-supervised hybrid network framework design diagram

According to Fig. 1, the proposed semi-supervised anomaly detection model architecture in this chapter is mainly divided into two modules. The first module is the HC-LSTM (Hybrid Convolutional LSTM) pre-training module, which mainly performs pre-training on labeled time-series data streams. The pre-training model includes a fuzzy roughening module and a C-LSTM model based on a self-attention mechanism. The second module is the prediction model updating algorithm based on weighted rough-fuzzy density anomaly detection (WFDA), which mainly functions to update the pre-training model network in the semi-supervised process.

B. Hybrid C-LSTM Model

The C-LSTM model is a deep learning neural network that combines convolutional neural networks and long short-term memory networks and is used for sequence classification. The model adopts a multi-layer feedforward network structure, which can effectively extract high-dimensional data features. Specifically, the model first uses a convolutional neural network to extract high-dimensional features of the input data and uses pooling operations to obtain the most significant features of the convolutional layer output as the input to the LSTM network. The LSTM network further extracts temporal features and performs further feature extraction and transformation on the input data through a fully connected layer. Finally, the model uses the Softmax classifier function to classify the input data, predict its result, and compare it with the true value to determine whether it is abnormal.

In order to filter out low-weight features and enhance the model's focus on key information in the input sequence, a self-attention mechanism is introduced. In addition, a fuzzy rough set processing module is further added to classify the input data features, as shown in Fig. 2.

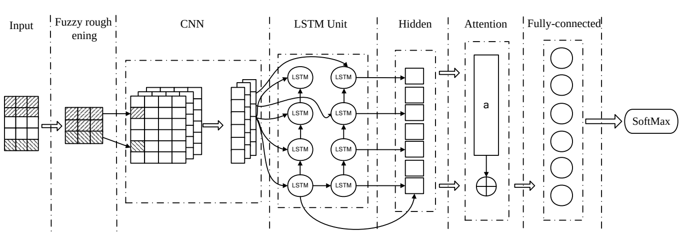


Fig. 2. Hybrid C-LSTM model diagram

C. Fuzzy-Rough Calculation

The fuzzy roughness calculation module aims to select the most discriminative subset of features from the original feature set. The rough fuzzy degree of a feature is an indicator used in fuzzy rough set theory to describe its discriminative ability, we first set time series data series $X = (x_1, x_2, \cdots, x_T)$, where $x_t \in R^n$ represents the n-dimensional data vector at time t, and the specific steps are as follows:

1) Fuzzification

Fuzzification is the process of transforming a data vector x_t into a fuzzy set A_t, where A_t represents the degree of membership of x_t on each attribute. For each attribute i, its membership function is$\mu_{x_{i,t}}(x)$, which represents the degree of membership of x on attribute i. The sigmoid function is used as the membership function to fuzzify the data vector x_t into a fuzzy set A_t, which represents the degree of membership of x_t on attribute i.

$$A_t(x_i) = \mu_{x_{i,t}}(x_i) \tag{1}$$

2) Roughification

Roughification refers to the process of transforming each fuzzy set A_t at every time step into an equivalence class E_t, where E_t represents the equivalence class partition of x_t based on each attribute. For each attribute i, its equivalence R_i is defined to indicate whether two data vectors x_{t_1} and x_{t_2} are equal on attribute i. $R_i(x_{t_1}) = R_i(x_{t_2})$ if and only if $x_{i,t_1} = x_{i,t_2}$. Therefore, the fuzzy set A_t can be roughified into an equivalence class E_t, where $E_t(x_i) = x_{i,t'}$,$x_{i,t'}$ has the same value as $x_{i,t}$ on all attributes.

$$E_t\left(x_i\right) = \left\{x_{i,t'} \mid R_1\left(x_{t'}, x_t\right) \wedge R_2\left(x_{t'}, x_t\right) \wedge \cdots \wedge R_n\left(x_{t'}, x_t\right)\right\} \tag{2}$$

3) Feature Extraction

Feature extraction refers to the process of computing a fuzzy feature vector, denoted as f_t, for each equivalence class E_t. Each dimension of the fuzzy feature vector f_t represents the fuzzy mean value of the corresponding attribute for the equivalence class E_t. For each attribute i, the fuzzy mean value $\bar{A}_t(x_i)$ is computed for the equivalence class E_t, which represents the average membership degree of E_t on attribute i.

$$\bar{A}_t\left(x_i\right) = \frac{\sum_{x_{i,t'} \in E_t} A_t\left(x_{i,t'}\right)}{\left|E_t\left(x_i\right)\right|} \tag{3}$$

Secondly, the fuzzy mean values on all attributes can be assembled into a fuzzy feature vector f_t, where $f_t(x_i) = \bar{A}_t(x_i)$, representing the fuzzy feature vector of the equivalence class E_t.

$$f_t\left(x_i\right) = \bar{A}_t\left(x_i\right) \tag{4}$$

Finally, the obtained fuzzy feature vector sequence $f = (f_1, f_2, \cdots, f_T)$ is used as the input for the C-LSTM network to perform downstream task training and prediction.

D. Update algorithm based on WFDA model

In this section, we propose an anomaly detection algorithm based on a fuzzy rough granulation model. The algorithm addresses the challenge faced by the semi-supervised model proposed in this paper, which struggles to identify novel anomalies with limited training data and achieves low detection rates. By iteratively updating the pre-trained model in a reverse manner, the algorithm improves the accuracy and precision of the semi-supervised approach.

Based on literature [16], the algorithm process is shown in the algorithm 1:

In the model update algorithm, firstly, the non-empty finite set of data attributes A is initialized. Then, a loop is performed for each attribute to calculate the fuzzy granularity structure and fuzzy entropy, which determines the influence of each attribute on anomalies. Next, a loop is conducted for each data object to calculate its fuzzy rough density and corresponding weight value. Subsequently, the ratio between local density and global density is computed, followed by the calculation of the anomaly score for each object. Through analysis, the number of iterations in the algorithm is given by $|A||O||O| + |O||A|$. Therefore, in the worst-case scenario, the time complexity of the algorithm is $O(|A||O|^2)$.

Algorithm 1 Model Update Algorithm

Input: $sample\ set$: $\mathrm{D} = \{O, A\}, parameter: \delta$
Output: $abnormal\ score$: AS

```
procedure GETSCORE(a, o)
    for a ∈ A do
        STR(R_a) = {[o_1]_a, [o_2]_A, [o_3]_A, ..., [o_n]_a};
        FE(a) = -(1/|O|) Σ_{o∈O} log2 (|[o]_a| / |O|);
    end for
    for o ∈ O do
        for a ∈ A do
            FRD_a(o) = (Σ_{a∈A} Density_a(o)) / |A|;
            W(a) = FE(a) / Σ_{i=1}^{|A|} (FE(a_i)) ∈ [0, 1];
        end for
        AS(o) = Σ_{a∈A} W(a) (1 - FRD_a(o));
    end for
    return AS;
end procedure
```

Line 3: $STR\left(R_a\right) = \left\{[o_1]_a, [o_2]_A, [o_3]_A, \ldots, [o_n]_a\right\}$; line 4: $FE(a) = -\frac{1}{|O|}\sum_{o\in O}\log 2\frac{|[o]_a|}{|O|}$; line 8: $FRD_a(o) = \frac{\sum_{a\in A}\mathrm{Density}_a(o)}{|A|}$; line 9: $W(a) = \frac{FE(a)}{\sum_{i=1}^{|A|}(FE(a_i))} \in [0,1]$; line 11: $AS(o) = \sum_{a\in A} W(a)\left(1 - FRD_a(o)\right)$.

E. Semi-Supervised Scheme Detailed Design

The overall design of the scheme consists of two modules, namely the hybrid C-LSTM module and the WFDA-based model update algorithm. The main process is illustrated in Fig.3.

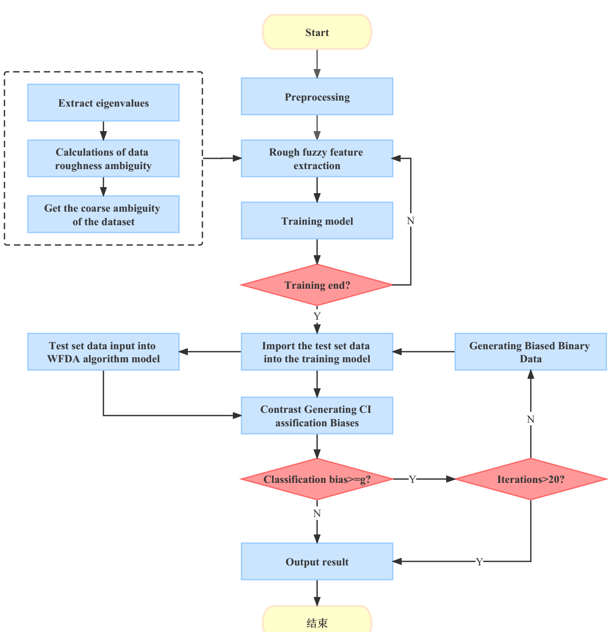


Fig. 3. Schematic diagram of the semi-supervised scheme

1) Pre-training process

The process of fuzzy roughening described above results in a sequence of fuzzy rough feature vectors $f = (f_1, f_2, \cdots, f_T)$ for the dataset X, which serve as inputs to the convolutional neural network (CNN). The CNN performs convolution and pooling operations to further extract important spatial features of the time series data. The long-term dependent time features are then extracted using the LSTM. The proposed hybrid C-LSTM network structure includes two convolutional layers, two max-pooling layers, two shared LSTM networks, and a fully connected layer with the tanh function as the activation function.

The main steps of the process include:

(1) Label the time-series data stream, the input time window length 1 is determined by a sliding window based on time intervals. The input data consists of a feature vector sequence from an industrial sensor data stream, with each data point containing four attributes. The data is distributed to a CNN network as input, with a convolutional kernel size of 3, convolution stride of 1, and no padding used for input data.

(2) Use a convolutional neural network to extract high-dimensional features as input to a long short-term memory (LSTM) network. Max-pooling is used to compress the number of features, with a pooling kernel size of 2 and a pooling stride of 2. The output of the last pooling layer of the CNN network is used as input to the LSTM network.

(3) The LSTM network takes the high-dimensional features extracted in step (2) as input and continues to extract their temporal features. The tanh function is used as the activation function for the hybrid C-LSTM network. The tanh function maps a real number to the range of [-1,1], with an output of 0 when the input is 0. The tanh function is sensitive to changes in the middle region and can effectively suppress both ends, which is very beneficial for classification tasks.

(4) In this paper, the tanh function is used as the activation function, which can effectively capture the subtle changes in the sensor data stream after anomalies occur. Additionally, it can compress the network output data into the range of [-1,1], ensuring that data does not spread or exceed limits between layers of the convolutional neural network, accelerating network fitting and increasing network robustness.

$$\tanh x = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{5}$$

After compressing the data using fully connected layers, it is passed through the Softmax function for anomaly detection. Steps (1) to (4) are repeated until all training data is processed.

2) Semi-Supervised Process

This section introduces the semi-supervised real-time detection module. The semi-supervised module mainly deals with unlabeled time-series data streams, and the semi-supervised process is as follows:

(1) The unlabeled time-series data stream is used as input to the pre-trained model, which predicts the input time-series data and classifies the processed data into labels denoted as Y_{hclstm}. At the same time, the real-time online anomaly judgment is performed by the fuzzy-rough density-based anomaly detection module, and the data is labeled as Y_{wfda}.

(2) The trustworthy data is labeled by combining the HC-LSTM model and the WFDA algorithm, which have the same classification results in labels Y_{hclstm} and Y_{wfda}. Unreliable data with different classification results are further classified using the WFDA-based model update algorithm.

(3) The classification results from step (2) are re-input into the pre-trained convolutional layer to extract features, and the results are used as input to the LSTM network. The LSTM network parameters are updated continuously through backpropagation until the output of the LSTM network, passed through the fully connected layer and Softmax classifier function, produces a classification result with an error smaller than the classification error threshold g of the original labeled data. At this point, the anomaly detection model has successfully learned the new data distribution and the iteration ends. If the iteration exceeds 25 times, the iteration ends, indicating that the anomaly detection model has failed to learn the new data distribution.

The classification error threshold g is defined as follows:

$$g=\sum_{i=0}\left|Y_{hclstm}(i)-Y_{wfda}(i)\right| \tag{6}$$

The LSTM backpropagation process is illustrated in Fig. 4.

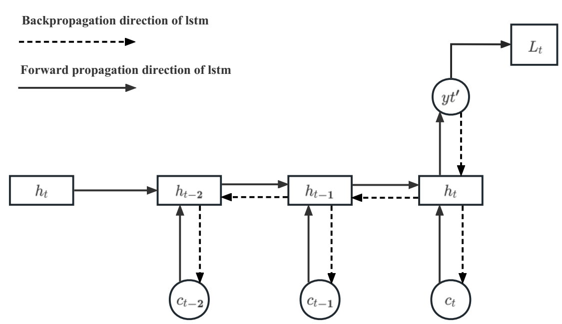


Fig. 4. LSTM schematic diagram of backpropagation

IV. Experimental Evaluation

A. Dataset Acquisition And Preprocessing

To validate the feasibility of the semi-supervised model, an actual sensor dataset was used from the Jinlu-Equipment-Health-Management-Platform. The experimental data was collected from indoor driving of multiple sensors on industrial machinery equipment from January 2023 to February 2023 by Henan Jinlu Network Technology Co., Ltd. Each data set includes four attributes: current, voltage, temperature, and pressure, and is unlabeled raw acquisition data with missing and abnormal values, requiring preprocessing before experimentation. The processing steps were as follows:

(1) Selecting 80,000 sets of data from the dataset as experimental data, using linear interpolation to fill in missing values for normal data of current, voltage, temperature, and pressure. The first 80,000 sets of data after preprocessing were taken as the experimental dataset and the preprocessed data were labeled as normal data.

(2) In the experimental dataset, randomly selecting some normal points and changing their four attribute values to abnormal values, and inserting 1%, 3%, 5%, and 10% abnormal points into the dataset respectively, obtaining four labeled datasets with abnormal percentages of 1%, 3%, 5%, and 10%. The experimental dataset is shown in Table I.

TABLE I
EXPERIMENTAL DATASET

Num	Dataset	Training	Test	Abnormal Ratio/%
I	JL_2023	5000	75000	1
II	JL_2023	5000	75000	3
III	JL_2023	5000	75000	5
IV	JL_2023	5000	75000	10

B. Evaluation Indexes

The anomaly detection model proposed in this section outputs detection results as a binary classification problem, with the classification results divided into the normal set P and the anomaly set N. As a binary classification problem, the classification results are generally divided into four categories according to the prediction situation [17], [18], as follows:

True Positive (TP): the positive samples correctly predicted by the model as positive;

False Positive (FP): the negative samples incorrectly predicted by the model as positive;

True Negative (TN): the negative samples correctly predicted by the model as negative;

False Negative (FN): the positive samples incorrectly predicted by the model as negative.

To further evaluate the performance of the model, $Precision$, $Recall$, and $F1_{\text{score}}$ are used as evaluation metrics for algorithm performance. Precision refers to the percentage of predicted positive samples that are actually positive.

Precission, the term precision refers to the percentage of predicted positive samples that are actually positive, calculated as the number of true positives divided by the total number of positive predictions;

$$\text{Precission }=\frac{TP}{(TP+FP)} \tag{7}$$

Recall, the recall rate refers to the probability of samples predicted as positive among the actual positive samples in the original dataset;

$$\text{Recall }=\frac{TP}{(TP+FN)} \tag{8}$$

$F1_{\text{score}}$, the $F1_{\text{score}}$ is a metric that simultaneously considers precision and recall, achieving a balance between the two measures by finding their optimal trade-off point. It is calculated as the harmonic mean of precision and recall, and is commonly visualized on a precision-recall (P-R) curve;

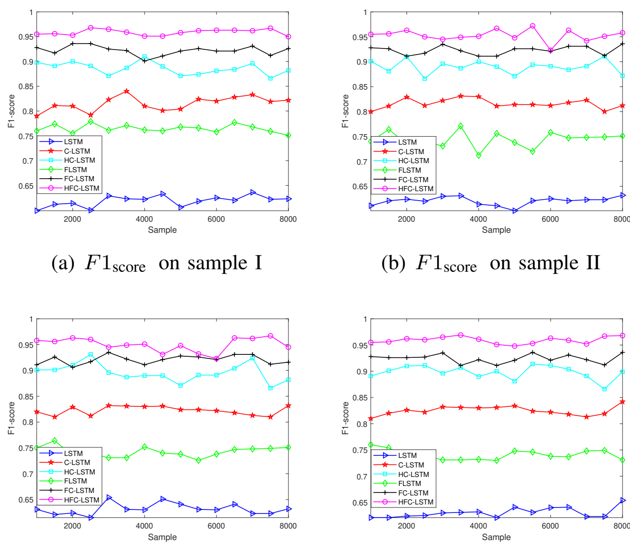
$$F1_{\text{score }}=\frac{2\times\text{Recall}\times\text{ Precission}}{\text{Recall }+\text{ Precission}} \tag{9}$$

C. Experimental results

To validate the effectiveness of the proposed anomaly detection algorithm in this chapter, the JL2023 dataset, which contains real-world industrial production data, was used as the experimental input.

Multiple experiments were conducted to demonstrate the effectiveness of the proposed semi-supervised model. Specifically, the experimental results of six models were compared, including HC-LSTM, C-LSTM, LSTM, HC-LSTM with semi-supervised architecture (HFC-LSTM), C-LSTM with semi-supervised architecture (FC-LSTM), and LSTM with semi-supervised architecture (FLSTM).

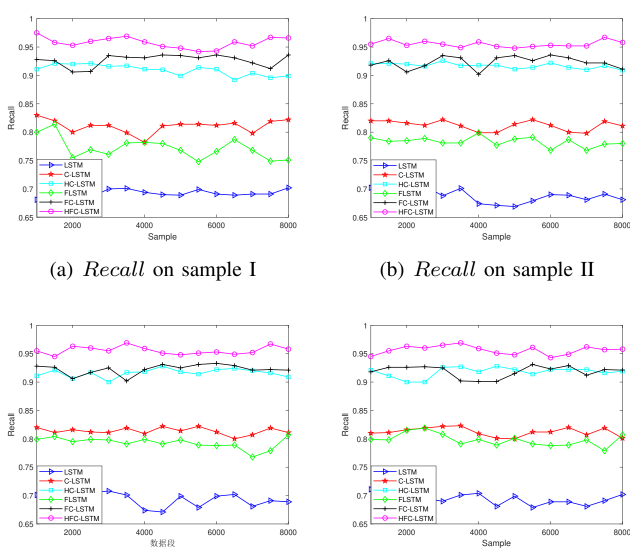
First, 5000 pre-training iterations were performed for each of the six models using the data samples labeled I, II, III, and IV. Then, the three semi-supervised models were trained online with an additional 75,000 data samples, while the other three models performed real-time anomaly detection and recorded the results. The experimental comparison of the six models on different datasets with different anomaly points is shown in Fig. 5 and Fig. 6.



(a) $F1_{\text{score}}$ on sample I (b) $F1_{\text{score}}$ on sample II

(c) $F1_{\text{score}}$ on sample III (d) $F1_{\text{score}}$ on sample IV

Fig. 5. Comparison of $F1_{\text{score}}$ among different algorithm models



(a) *Recall* on sample I (b) *Recall* on sample II

(c) *Recall* on sample III (d) *Recall* on sample IV

Fig. 6. Comparison of *Recall* among different algorithm models

The average scores of various algorithms on samples labeled as I, II, III, and IV are shown in TableII:

Comparison results of different algorithms on datasets with varying anomaly ratios are shown in Figures 4.5 and 4.6. The use of the semi-supervised architectures in the three groups of models, FLSTM vs. LSTM, FC-LSTM vs. C-LSTM, and HFC-LSTM vs. HC-LSTM, resulted in significant improvements in both the F1 score and recall compared to the control groups of the original algorithms without semi-supervised learning. The average improvements in F1 score and recall compared to the three control groups were 12.2% and 9.3%, 10.3% and 11.5%, and 4.1% and 6.4%, respectively. These results demonstrate the effectiveness of the semi-supervised anomaly detection method proposed in this paper.

By comparing the results of HC-LSTM, C-LSTM, and LSTM in Fig. 5 and Fig. 6, it was found that the use of HC-LSTM resulted in significant improvements in both $F1_{\text{score}}$ and recall. Specifically, the $F1_{\text{score}}$ of HC-LSTM improved by 9.5% compared to C-LSTM and by 28.9% compared to LSTM. Moreover, by comparing the results of HFC-LSTM, FC-LSTM, and FLSTM, it can also be found that the use of HFC-LSTM also resulted in significant improvements in both $F1_{\text{score}}$ and recall. Specifically, the $F1_{\text{score}}$ of HFC-LSTM improved by 3.5% compared to FC-LSTM and by 28.1% compared to FLSTM. These results demonstrate the effectiveness of the proposed HFC-LSTM with a sub-attention mechanism in extracting spatial features from the data.

The experiments on datasets with different numbers of anomalies, as shown in Fig. 5 and Fig. 6 and Table II, indicate that the proposed HFC-LSTM semi-supervised anomaly detection method exhibits stable performance and robustness across datasets with different anomaly ratios.

Furthermore, it was observed that the use of the semi-supervised HC-LSTM and C-LSTM resulted in less improvement in $F1_{\text{score}}$ and *Recall* compared to LSTM. This is because LSTM alone cannot effectively extract sufficient features from the time series data, which makes pre-training difficult. However, the further updated WFDA algorithm helped the model fit more effectively with new data, greatly improving the accuracy of the anomaly detection algorithm. On the other hand, although HC-LSTM and C-LSTM can extract temporal and spatial features, the limited pre-training data makes the model only relatively fit and unable to accurately measure the performance of the anomaly detection algorithm.

In order to further investigate whether the HFC-LSTM semi-supervised model can maintain good performance compared to other algorithms when the pre-training set is sufficient, we gradually increased the proportion of the pre-training set in stages, including 5%, 10%, 20%, and 30%. The experimental results are shown in Fig. 7: Observing Figure 4.7, it can be seen that as the proportion of the pre-training set increases, HC-LSTM, C-LSTM, and LSTM all achieve varying degrees of improvement in F1 score and recall.

Observing Fig. 7, it can be seen that as the proportion of pre-training data increases, the scores of each algorithm show significant improvements. The F1 scores and recall rates of LSTM, C-LSTM, and HC-LSTM on 5% pre-training data are 0.638, 0.721, and 0.807, respectively. On 30% pre-training data, the scores are 0.701, 0.831, and 0.873, respectively, with improvements of 6.3%, 11%, and 5.8%. FLSTM, FC-LSTM,

TABLE II
PERFORMANCE OF DIFFERENT ALGORITHMS ON DATASETS WITH DIFFERENT ANOMALY RATIOS

Algorithm	1%(Outlier)		3%(Outlier)		5%(Outlier)		10%(Outlier)	
	Recall	$F1_{\text{score}}$	Recall	$F1_{\text{score}}$	Recall	$F1_{\text{score}}$	Recall	$F1_{\text{score}}$
LSTM	0.693	0.619	0.687	0.621	0.692	0.632	0.692	0.629
C-LSTM	0.811	0.816	0.812	0.817	0.814	0.825	0.813	0.826
HC-LSTM	0.909	0.884	0.917	0.889	0.916	0.896	0.916	0.899
FLSTM	0.771	0.763	0.783	0.744	0.793	0.741	0.792	0.739
FC-LSTM	0.927	0.925	0.923	0.926	0.923	0.922	0.923	0.924
HFC-LSTM	0.958	0.960	0.955	0.955	0.956	0.953	0.957	0.959

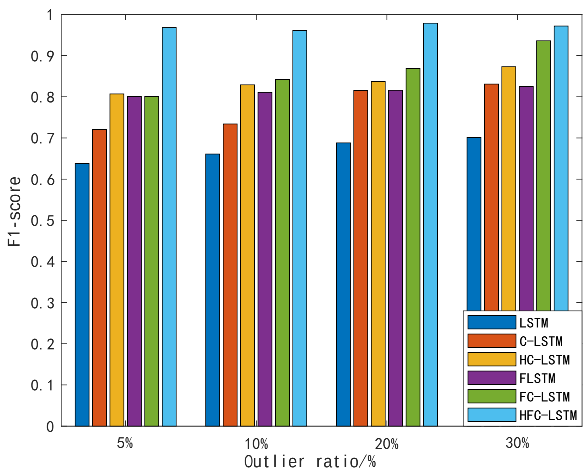
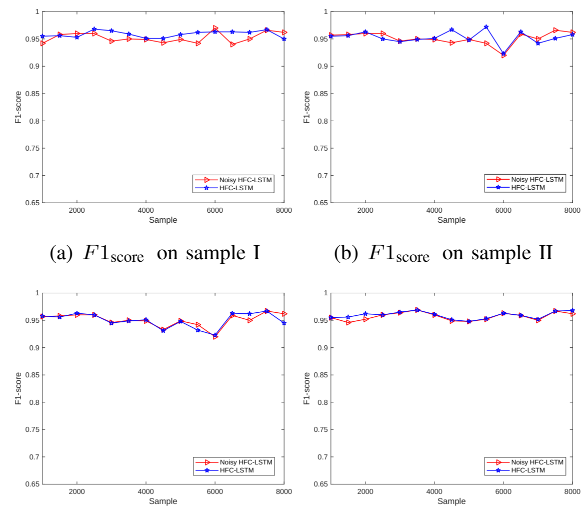


Fig. 7. Scores of algorithms under different proportions of anomalous training sets



(a) $F1_{\text{score}}$ on sample I (b) $F1_{\text{score}}$ on sample II

(c) $F1_{\text{score}}$ on sample III (d) $F1_{\text{score}}$ on sample IV

Fig. 8. HFC-LSTM algorithm's performance score with and without noise environment

and HFC-LSTM on 5% pre-training data are 0.801, 0.801, and 0.968, respectively. On 30% pre-training data, the scores are 0.825, 0.936, and 0.972, respectively, with improvements of 2.9%, 16.9%, and 0.4%. The experimental results demonstrate that the performance of the models improves with the increase of pre-training data, but it can be seen that the proposed semi-supervised HFC-LSTM model still exhibits outstanding superiority. And HFC-LSTM has good performance when the pre-training ratio is low, which is more in line with the actual production environment needs.

Experiments were conducted on different data samples (I, II, III, IV) with varying percentages of abnormal values (1%, 3%, 5%, 10%) without preprocessing to test the noise resistance of the semi-supervised anomaly detection scheme described above. The experimental results are shown in Figure 8.

V. Conclusion & Future Work

In this paper, we propose a semi-supervised learning-based anomaly detection method specifically designed for non-standard mechanical production processes. The method incorporates a hybrid C-LSTM pre-training model with a self-attention mechanism and a model updating algorithm based on WFDA. This approach addresses the challenges of acquiring a large annotated dataset for training due to the characteristics of small production scale and sparse samples in non-standard mechanical equipment. To address the issue of insufficient learning caused by a limited training set, we employ the WFDA-based model updating algorithm for semi-supervised learning. Compared to similar algorithms, our semi-supervised approach achieves an $F1_{\text{score}}$ of 0.955 and a $Recall$ of 0.957 on real-world datasets with varying proportions of anomalies. This represents a 4.1% improvement in $F1_{\text{score}}$ and a 6.4% improvement in $Recall$ compared to similar anomaly detection algorithms that do not utilize our proposed method. Furthermore, our approach requires only a small amount of labeled training data, resulting in significant cost savings. It demonstrates practical research value in the field.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China under Grant No. 62220106004 and No. 61972308; in part by the Major Research Plan of the National Natural Science Foundation of China under Grant No. 92267204; in part by the Key Research and Development Program of Shaanxi under Grant No. 2022KXJ-093 and Grant No. 2021ZDLGY07-05, in part by the Innovation Capability Support Program of Shaanxi (Program No. 2023-CX-TD-02); in part by the Construction of the "Scientists+Engineers" Team in Qin Chuangyuan, Shaanxi Province (Program No. 2023KXJ-033).

REFERENCES

[1] Sisinni E, Saifullah A, Han S, et al. Industrial internet of things: Challenges, opportunities, and directions[J]. IEEE transactions on industrial informatics, 2018, 14(11): 4724-4734.

[2] Malik P K, Sharma R, Singh R, et al. Industrial Internet of Things and its applications in industry 4.0: State of the art[J]. Computer Communications, 2021, 166: 125-139.

[3] Li G, Fu Y, Chen D, et al. Deep anomaly detection for CNC machine cutting tool using spindle current signals[J]. Sensors, 2020, 20(17): 4896.

[4] Chen W. Intelligent manufacturing production line data monitoring system for industrial internet of things[J]. Computer communications, 2020, 151: 31-41.

[5] Liu Y, Tong K D, Mao F, et al. Research on digital production technology for traditional manufacturing enterprises based on industrial Internet of Things in 5G era[J]. The International Journal of Advanced Manufacturing Technology, 2020, 107: 1101-1114.

[6] Sain S R. The nature of statistical learning theory[J]. 1996.

[7] Breiman L. Random forests[J]. Machine learning, 2001, 45: 5-32.

[8] Shahzad S, Ilyas M, Lali M, et al. Sperm Abnormality Detection Using Sequential Deep Neural Network[J]. Mathematics, 2023, 11(3): 515.

[9] Zhang Q. Financial data anomaly detection method based on decision tree and random forest algorithm[J]. Journal of Mathematics, 2022, 2022.

[10] Rumelhart D E, Hinton G E, Williams R J. Learning representations by back-propagating errors[J]. nature, 1986, 323(6088): 533-536.

[11] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural computation, 1997, 9(8): 1735-1780.

[12] Greff K, Srivastava R K, Koutník J, et al. LSTM: A search space odyssey[J]. IEEE transactions on neural networks and learning systems, 2016, 28(10): 2222-2232.

[13] Wang Z, Liu N, Chen C, et al. Adaptive self-attention LSTM for RUL prediction of lithium-ion batteries[J]. Information Sciences, 2023.

[14] Glielmo A, Husic B E, Rodriguez A, et al. Unsupervised learning methods for molecular simulation data[J]. Chemical Reviews, 2021, 121(16): 9722-9758.

[15] Verma V, Kawaguchi K, Lamb A, et al. Interpolation consistency training for semi-supervised learning[J]. Neural Networks, 2022, 145: 90-106.

[16] Yuan Z, Chen B, Liu J, et al. Anomaly detection based on weighted fuzzy-rough density[J]. Applied Soft Computing, 2023: 109995.

[17] Glas A S, Lijmer J G, Prins M H, et al. The diagnostic odds ratio: a single indicator of test performance[J]. Journal of clinical epidemiology, 2003, 56(11): 1129-1135.

[18] Yacouby R, Axman D. Probabilistic extension of precision, recall, and f1 score for more thorough evaluation of classification models[C]//Proceedings of the first workshop on evaluation and comparison of NLP systems. 2020: 79-91.