


Research Article

Quantitative Detection of Financial Fraud Based on Deep Learning with Combination of E-Commerce Big Data

Jian Liu ¹, Xin Gu,^{1,2} and Chao Shang^{2,3}

¹School of Business, Sichuan University, Chengdu 610064, China

²Chengdu Soft Innovation Intelligence Association, Chengdu 610023, China

³Institute of New Structural Economics, Peking University, Beijing 100871, China

Correspondence should be addressed to Jian Liu; jamesliu@stu.scu.edu.cn

Received 15 November 2020; Revised 10 December 2020; Accepted 14 December 2020; Published 24 December 2020

Academic Editor: M. Irfan Uddin

Copyright © 2020 Jian Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

At present, there are more and more frauds in the financial field. The detection and prevention of financial frauds are of great significance for regulating and maintaining a reasonable financial order. Deep learning algorithms are widely used because of their high recognition rate, good robustness, and strong implementation. Therefore, in the context of e-commerce big data, this paper proposes a quantitative detection algorithm for financial fraud based on deep learning. First, the encoders are used to extract the features of the behaviour. At the same time, in order to reduce the computational complexity, the feature extraction is restricted to the space-time volume of the dense trajectory. Second, the neural network model is used to transform features into behavioural visual word representations, and feature fusion is performed using weighted correlation methods to improve feature classification capabilities. Finally, sparse reconstruction errors are used to judge and detect financial fraud. This method builds a deep neural network model with multiple hidden layers, learns the characteristic expression of the data, and fully depicts the rich internal information of the data, thereby improving the accuracy of financial fraud detection. Experimental results show that this method can effectively learn the essential characteristics of the data, and significantly improve the detection rate of fraud detection algorithms.

1. Introduction

With the development of the economy, there are more and more frauds in the financial field. The detection and prevention of financial fraud are of great significance for regulating and maintaining a reasonable financial order [1, 2]. Due to abnormal or unfair transactional nature, the transaction behaviours involved in fraud are different from the general customer and account operation behaviours and exhibit various abnormal characteristics, including abnormal transaction behaviours, abnormal transaction objects, abnormal transaction data, and abnormal capital trends. Anomaly detection refers to describing the user's behavioural characteristics to distinguish between abnormal and abnormal behavioural characteristics that violate normal behaviour [3]. It can be used to monitor the transaction behaviour of multiple customers, employees, and financial

transaction parties. It is very helpful for discovering the internal connection of things hidden in the data. It can effectively track and detect frauds [4].

In the study of financial fraud, the neural network algorithm has been widely used due to its high recognition rate, good robustness, and strong implementation [5, 6]. Hu et al. [7] used radial basis function neural network to detect credit card fraud. Fisch et al. [8] proposed a data mining system that uses historical transaction data to build a neural network model to detect fraud. Merchant et al. [9] proposed a neural network model based on merchant credit and ROC analysis. Kolalikhormuji et al. [10] proposed a credit card fraud detection system for cascade neural network recognition. That is, the gating network is used to aggregate the confidence values of the three parallel artificial neural network classifiers, and the weight of the gating network is trained by the imperialist competition algorithm to obtain

the optimal. Experiments show that this algorithm can obtain very high recognition rate and reliability. Liu et al. [11] used an evolutionary approach to construct a neural network model to identify fraud. Hu et al. [12] combined Bayesian and neural networks to build a detection model. They found that the simple Bayesian network takes less time to train in building the classifier, and the accuracy is higher. Nevertheless, the effect on the new data is not good, so they joined the neural network algorithm, and the effect has been improved very well. Mosbach et al. [13] used a Bayesian algorithm to construct an experimental system of bank antifraud model, to identify whether a credit card user has committed fraud. Credit card transaction records are also time-specific, and their records have corresponding relevance. Therefore, you can use association rules to find out customer transaction rules and then use the decision tree for anomaly detection. Ordyan et al. [14] used decision trees to analyse the differences between normal and abnormal transactions and then used the differences to identify fraud. Menon et al. [15] introduced a new theory of cost-sensitive machine learning into the decision tree and he found that this decision tree algorithm not only has traditional performance such as accuracy, recall rate is higher than existing algorithms but also has a good expressiveness to the newly defined cost sensitivity in the field of credit card fraud. Roldán-García et al. [16] proposed a multicore support vector machine and introduced user configuration information instead of pure transaction information. Mhatre et al. [17] used Hidden Markov Model to detect fraud. Jana et al. [18] used a new idea of fuzzy logic to detect fraud detection. The method first calculates the initial credit of each transaction through the first-order Sugeno fuzzy model. If the transaction is found to be suspicious, the back credit is calculated using the previous doubt score and applied by Bayesian fuzzy inference. Lin et al. [19] constructed a fraud detection system by combining neural networks and association analysis. Cao et al. [20] introduced a new algorithm based on the root-searching fast hierarchical clustering algorithm in the research on the detection algorithm of merchant category codes and carried out classification experiments on the entire merchant category codes, and achieved good results. Ryan et al. [4] used deep belief networks to extract behavioural features and support vector machines (SVM) to detect fraudulent behaviours. Liu et al. [21] used the spatial-temporal convolutional neural network to extract fraudulent behaviour features for abnormal behaviour detection and localization. Peng et al. [22] used image saliency information and multiscale optical flow histograms as low-level features, and then the deep learning network PCAnet is used to extract more effective features from these underlying features for abnormal behaviour detection.

Through the analysis of the above literature, most of the existing abnormal behaviour detection uses artificial features, but artificial features have high computational complexity, and it is difficult to select and design an effective behaviour feature in complex scenarios. Therefore, this paper proposes a quantitative fraud detection algorithm based on deep learning of e-commerce big data. By

establishing a model for e-commerce big data feature learning, mining the financial fraud behaviour characteristics of e-commerce big data, and inputting the features into the abnormal behaviour detection model, it can effectively, quickly, and accurately identify financial fraud behaviour, quantify fraud risk levels, and do well in advance relevant prevention work to avoid unnecessary losses caused by financial fraud.

2. Financial Fraud

Fraud is a subject of widespread concern in the current society. The definition of financial fraud is diverse [23]. One argument is that financial fraud is the act of using loopholes in the rules of financial products to obtain illicit benefits. It can also be said that any unjust act in the financial market that seeks to gain its own interests and causes losses to others or organizations is financial fraud.

Faced with thousands of financial products and their derivatives in the financial market, the means of financial fraud are diversified. According to the financial products involved, there are loan fraud, deposit fraud, bill fraud, bankcard fraud, securities fraud, and insurance fraud. From the source of fraud, it can be divided into external fraud and internal fraud. According to fraud, it can be divided into the following three types:

- (1) Using the banking transaction system to carry out illegal intrusion or illegal operations, to seek illegitimate benefits. Bankcard fraud, identity theft, and a large number of internal violations of the bank are all of this type. Typical is business process fraud; that is, fraudulent acts that use loopholes in business processes to gain profits, as shown in Figure 1.
- (2) Providing false promises or false credit guarantee information for fraud. Most investment and financing fraud fall into this category. Fraudsters often use high-profit investments as bait to obtain a means of financial fraud input by investors.
- (3) Concealing important information and fraudulently creating information asymmetry. In the securities market, a large amount of insider trading, deliberately hiding its risks during the promotion of derivative products, and using various means to manipulate the securities market with a view to arbitrage are all types of fraud.

Regardless of the specific form of expression, there are two types of fraud commonly used by financial fraudsters: one is “make something out of nothing, it’s fake,” and the other is “the next one is good, the fish is mixed.” The details are shown in Figure 2.

3. Quantitative Detection Algorithm for Financial Fraud Based on Deep Learning

3.1. Stacked Denoising Encoder (SDAE). The denoising encoder DAE is a three-layer neural network used to reconstruct the original data x_i from the noise data \tilde{x}_i . DAE

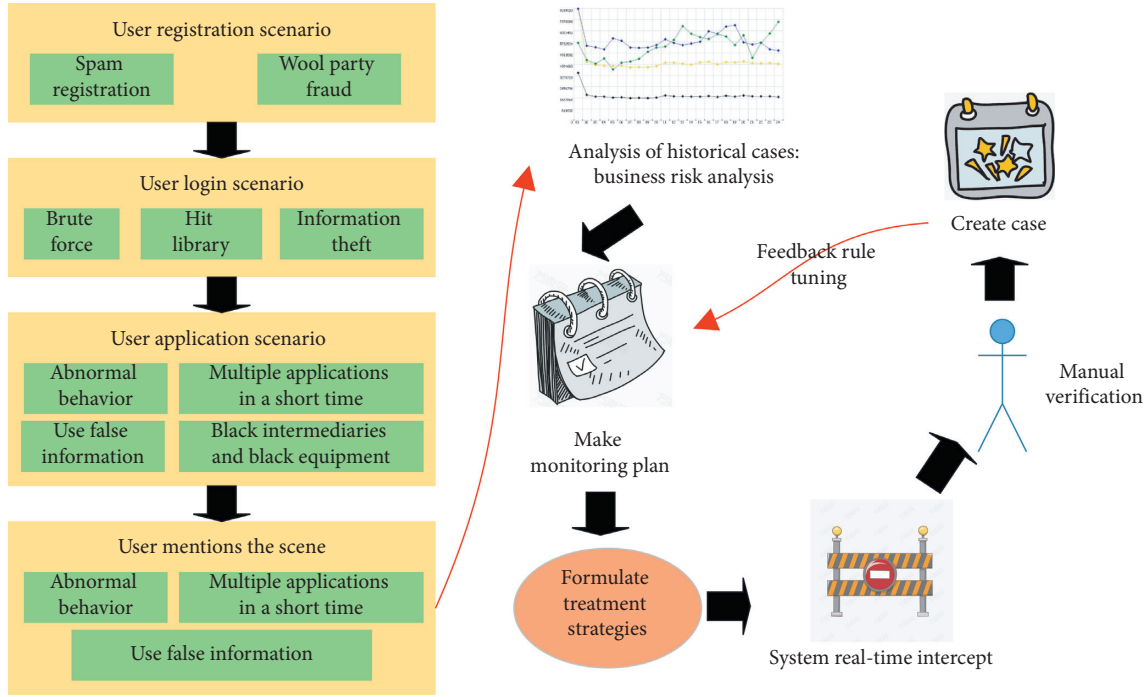


FIGURE 1: Fraud in financial business processes.

contains two parts, namely encoder and decoder. DAE learning is learning two mapping functions $f_e(W, b)$ and $f_d(W', b')$. The variable W and the variable b , respectively, represent the weight matrix and deviation vector of the encoder part. The variable W' and the variable b' , respectively, correspond to the parameters of the decoder. For noise data x_i , the hidden layer output of the encoder is y :

$$y = f_e(\tilde{x}_i) = s(W\tilde{x}_i + b) = \frac{1}{1 + e^{-(W\tilde{x}_i + b)}}. \quad (1)$$

The purpose of the decoder is to reconstruct the original data x_i from the noise data \tilde{x}_i :

$$z_i = f_d(y) = s(W'y + b'). \quad (2)$$

Given a set of training samples $X = \{x_i\}_{i=1}^N$, learn the parameters of DAE by solving the following optimization problem:

$$\min_{W, W', b, b'} \sum_{i=1}^N \|x_i - z_i\|_2^2 + \lambda (\|W\|_F^2 + \|W'\|_F^2) + \beta \sum_{j=1}^K KL(\mu \| \mu'). \quad (3)$$

Among them, the first term represents the reconstruction error, the second term is the weight penalty term, and the third term is the sparsity constraint. Variable λ and variable β are equilibrium parameters. The variable μ is a sparsity parameter, which represents the coefficient level of the hidden layer node. The variable K is the number of nodes in the hidden layer. The variable $\hat{\mu}_j$ is the activation value of the average threshold of the j th node of the hidden layer for all training samples. If the average activation value is greater than 0.5, the formula $\hat{\mu}_j = 1$ is established. Otherwise, the

formula $\hat{\mu}_j = 0$ is established. The third sparsity constraint is as follows:

$$KL(\mu \| \mu') = -\mu \log \hat{\mu}_j + (1 - \mu) \log(1 - \hat{\mu}_j). \quad (4)$$

Using gradient descent method to solve equation (3), the DAE parameters can be determined.

Multiple DAEs are stacked layer by layer to form SDAE, where the output of the lower layer DAE is used as the input of the upper layer DAE. For SDAE training, the DAE in each layer is trained from the lower layer to the higher layer. The trained SDAE can be used to learn an effective feature representation from the input data.

3.2. In-Depth Features of Financial Fraud. We use two SDAEs to extract behavioural features in the 3D volume centered on the trajectory. The size of the 3D volume is $N \times N \times L$. Among them, the value of the length L of the trajectory is 15, and the variable N takes 32. In order to embed structural information, first, divide the cube into a space-time grid of $n_\sigma \times n_\sigma \times n_\tau$, where variable n_σ takes the value 2, and n_τ takes the value 3. Then use SDAE to extract depth features in these grids. Finally, the features of all mesh are combined to obtain the depth features corresponding to the trajectory.

Figure 3 shows the structure of SDAE for extracting behaviour depth features. SDAE includes two parts: encoder and decoder. The number of input layer nodes of the encoder is equal to the number of dimensions of the input data. Then, half by layer reduces the number of nodes in each layer until the “bottleneck” hidden layer. The structure of the decoder is symmetrical to the encoder. The output of the “bottleneck” hidden layer is the depth feature.

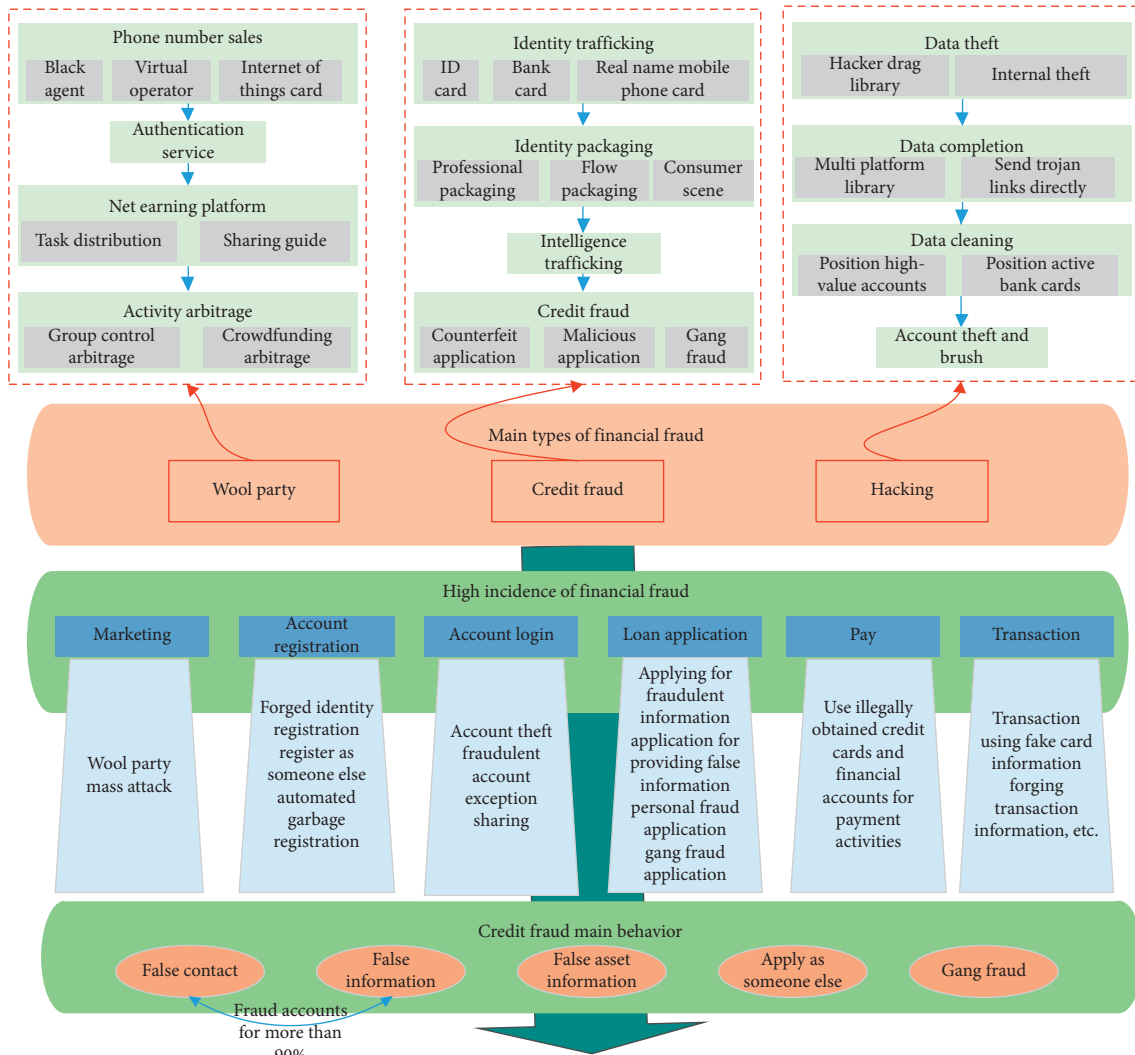


FIGURE 2: The main ways of financial fraud.

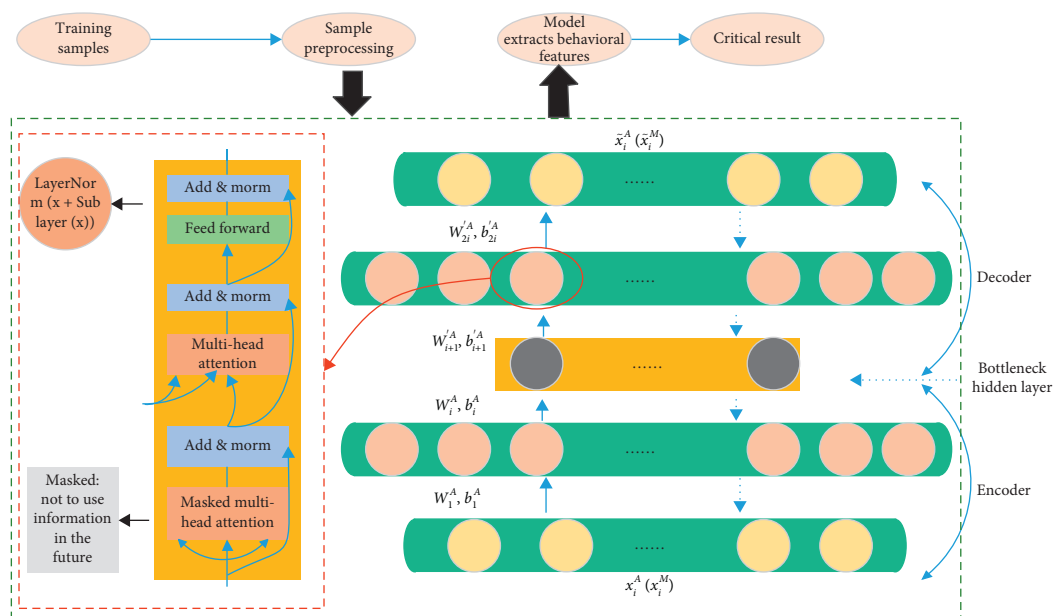


FIGURE 3: Structure of stacked denoising encoder.

3.3. Word Packet Representation of Behaviour Characteristics. For a trajectory, a 1440-dimensional feature vector can be obtained. As the number of targets is different, the number of points of interest and the corresponding number of trajectories are different, and the dimension of behavioural characteristics is different. Therefore, the word packet method is used to express the behaviour characteristics as visual words, to unify the dimensionality of the behaviour.

First, extract the depth features/depth of the trajectory. Then, cluster all trajectory depth features/depths to obtain N_v class centres. Each category corresponds to a visual word. For the test behaviour samples, according to the nearest neighbour principle, each of its trajectories is classified into each category. Therefore, the frequency of occurrence of each visual word in the sample can be obtained, and these frequencies constitute the visual word representation of the sample.

After many experiments, the number of appearance visual words and sports visual words were 370 and 430, respectively. Therefore, 370-dimensional and 430-dimensional visual word vectors represent the appearance depth feature and the motion depth feature, respectively.

3.4. Feature Fusion Based on Weighted Correlation. In order to improve the classification ability of features, feature fusion method based on weighted correlation is used to combine appearance depth features and motion depth features to form 800-dimensional feature vectors. For the convenience of presentation, the appearance depth feature and the motion depth feature are denoted by y^1 and y^2 , respectively, and then the fused feature y is as follows:

$$y = (w_1 y^1, w_2 y^2), \quad (5)$$

where w_i is the weighting coefficient, and $(w_1)^2 + (w_2)^2 = 1$. We determine the weighting coefficient based on intraclass consistency and interclass separability.

Intraclass consistency: it is generally desired that samples in the same category are as close as possible in the feature space. Nevertheless, usually, the sample features in the same category will have a larger variance. Therefore, there is no need to require all samples in the same category to be close to each other. One trade-off is to ensure that the samples within the same neighbour in the same class are as close as possible. Let $y_i = (w_1 y_i^1, w_2 y_i^2)$ and $y_j = (w_1 y_j^1, w_2 y_j^2)$ represent i th and j th samples, respectively, then the intraclass consistency is defined as follows:

$$S_c = \sum_{i=1}^N \sum_{j \in N_k^+(F_i)} \frac{\langle y_i, y_j \rangle}{\|y_i\| \|y_j\|} = \sum_{i=1}^N \sum_{j \in N^{1/k}(F_i)} \frac{\sum_{k=1}^2 w_k^2 y_i^k y_j^k}{\sqrt{\sum_{k=1}^2 w_k^2 (y_i^k)^2} \sqrt{\sum_{k=1}^2 w_k^2 (y_j^k)^2}}. \quad (6)$$

In the formula, the variable $N_k^+(F_i)$ represents the index set of K nearest neighbour samples of sample F_i that belong to the same class as F_i .

Separability between classes: it is required that the features have good distinguishability; that is, two samples of different classes are as far away as possible in the feature

space. However, there are many such sample pairs. In order to reduce the amount of calculation, only the sample pairs near the interface of the feature space are considered. Therefore, the separability between classes is defined as follows:

$$S_b = \sum_{i=1}^N \sum_{j \in N_k^-(F_i)} \frac{\langle y_i, y_j \rangle}{\|y_i\| \|y_j\|} = \sum_{i=1}^N \sum_{j \in N^{1/k}(F_i)} \frac{\sum_{k=1}^2 w_k^2 y_i^k y_j^k}{\sqrt{\sum_{k=1}^2 w_k^2 (y_i^k)^2} \sqrt{\sum_{k=1}^2 w_k^2 (y_j^k)^2}}. \quad (7)$$

In the formula, the variable $N_k^-(F_i)$ represents the index set of K nearest neighbour samples of sample F_i that are different from F_i . The fused features should have good intraclass consistency and interclass separability. Therefore, the weighting coefficient is determined by solving the following optimization problem:

$$\max\{(S_c - S_b) + \lambda_s \|w\|\}, \quad s.t. \ w_k > 0, \|w\| = 1. \quad (8)$$

In the formula, the variable λ_s is the control parameter. Use a gradient descent method to solve equation (8), namely:

$$w_k(t+1) = w_k(t) + \eta \frac{\partial L}{\partial w_k} \Big|_{w_k=w_k(t)}. \quad (9)$$

In the formula, the variable t is the number of iterations, the variable η is the iteration step, and the formula $L = (S_c - S_b) + \lambda_s \|w\|$ is the objective function.

$$\frac{\partial L}{\partial w_k} = \sum_{i=1}^N \left(\sum_{j \in N_k^+(x_i)} \frac{\partial h_{ij}(w)}{\partial w_k} - \sum_{j \in N_k^-(x_i)} \frac{\partial h_{ij}(w)}{\partial w_k} \right) + 2\lambda_s w_k. \quad (10)$$

3.5. Fraud Detection Based on Sparse Reconstruction. After obtaining the characteristics of the behaviour, sparse reconstruction is used to detect fraud. The basic idea is that

any behaviour can be represented by a sparse linear combination of normal training samples. For normal behaviour, the sparse reconstruction error is small, while the abnormal behaviour sparse reconstruction error is relatively large. Therefore, we can detect fraud based on reconstruction errors.

There are class C normal behaviours, and the above feature vector represents each behaviour. The variable $D = \{D_1, D_2, D_3, \dots, D_C\}$ represents a sparse dictionary, where D_i is a subdictionary composed of K behaviours of type i , which can be expressed as follows:

$$y = Da. \quad (11)$$

The variable $a = \{a_1, a_2, a_3, \dots, a_C\}^T$ is a sparse coding vector.

Given a dictionary D , the test sample y can be expressed by formula (11). Among them, the sparse code a can be obtained by solving the following formula:

$$a^* = \min_a \|y - Da\|_2 + \lambda \|a\|. \quad (12)$$

Once the optimal sparse coding a^* is obtained, the sparse reconstruction cost can be calculated:

$$S(y, a^*, D) = \|y - Da^*\|_2 + \lambda \|a^*\|. \quad (13)$$

For normal behaviour, the cost of sparse reconstruction is smaller, while the abnormal behaviour is more expensive. So if

$$S(y, a^*, D) > \varepsilon. \quad (14)$$

Then y is fraud. In the formula, ε is a present threshold.

As shown in Figure 4, the fraud detection method includes two stages. The first is the training phase of the model, as shown in Figure 4(a). Model training includes feature learning model and classification model training. Among them, the data preprocessing standardizes and normalizes the input data and completes the conversion of data types. The feature learning model training takes preprocessed unlabelled data as input to train the feature learning model. The trained feature learning model can be directly used to learn the features of the data and transform the representation of the sample in the original space into a new feature space. The training and testing of classification models should use labelled training sets and test sets for supervised training and testing.

The second stage is the detection stage of financial fraud, shown in Figure 4(b). In this stage, the input data is first preprocessed, then the feature learning model obtained in the first stage of training is applied to learn the features of the data to be classified, and finally, the learned data features are used as input to the classification model to classify the data.

4. Results and Discussion

4.1. Model Evaluation Index. In the fraud detection algorithm, it is very important to find a good classifier evaluation index. On the one hand, a good evaluation index fully indicates the classifier's ability to solve problems and can show

the effect on users more comprehensively. On the other hand, the selected classifier evaluation index is also conducive to developers to optimize the classifier model. The main evaluation criteria of this article are accuracy rate, recall rate, and FM. Table 1 is a confusion matrix, which introduces common concepts such as TP, FP, FN, TN. Where "0" means normal behaviour, and "1" means fraudulent behaviour. "Y" indicates that the number of detection stages is classified as normal behaviour, and "N" indicates that the number of detection stages is classified as fraud.

Based on the above four classifications, the following two concepts can be further extended to evaluate the performance of the classifier:

$$\text{Precision} = \frac{TP}{TP + FP}, \quad (15)$$

$$\text{Recall} = \frac{TP}{TP + FN}. \quad (16)$$

Precision reflects accuracy. Recall reflects the recall rate. In general, the accuracy rate is to determine how accurate the detected classification is, and the recall rate indicates how many items with the correct classification are detected. Accuracy and recall indicators are sometimes opposite. The measure of model performance that combines accuracy and recall into a single value is the F metric. The F measure uses the harmonic mean to integrate the accuracy and recall rates. The specific formula is shown in the following formula:

$$F = \frac{(q^2 + 1) \text{Precision} * \text{Recall}}{q^2 (\text{Precision} + \text{Recall})}. \quad (17)$$

When the parameter $q = 1$ is established, it is the most common FM. The formula is as follows:

$$\text{FM} = \frac{2 \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (18)$$

Because the F metric turns the model's performance index into a single value, it provides a convenient way to compare the model to good or bad. However, this requires assuming that the accuracy and recall rates have the same weight. A better practice is to combine F measures with other measures.

In order to weigh the quality of the model, this paper also adds the ROC curve and AUC index to measure the overall credibility of the classifier. ROC represents the relationship between TPR and FPR in the classification confusion matrix. Therefore, the abscissa of the ROC curve represents the probability of a negative instance being regarded as a positive instance, and the ordinate represents the probability of a positive instance being regarded as a positive instance. In the ROC graph, TPR increases with the increase of FPR, and the faster the increase is, the more prominent the curve is, and the better the classification performance of the response model is. The value of AUC is the size of the area under the ROC curve. The larger the AUC is, the better the classifier performance is.

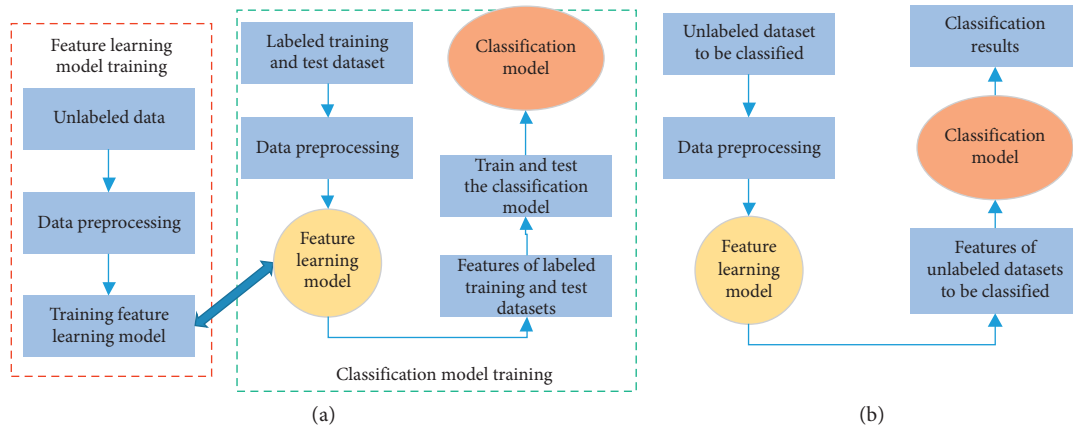


FIGURE 4: Detection process of financial fraud based on feature learning. (a) Model training stage. (b) Anomaly detection stage.

TABLE 1: Confusion matrix applied to the detection model.

| | | Experiment classification label | |
|------------------------|---|--|--|
| | | 0 | 1 |
| Predict fraud category | Y | TP: Correct and affirmative. A sample of normal behaviour predicted to be normal behaviour. | FP: The error is positive. A sample of normal behaviour predicted to be fraudulent. |
| | N | FN: False negative. A sample of fraudulent behaviour predicted as normal behaviour. | TN: Correctly denied. A sample of fraudulent behaviour predicted to be fraudulent. |

In summary, this paper uses four indicators: accuracy rate, recall rate, FM, and ROC chart to measure the quality of the experimental results of this paper.

4.2. Detection Performance Analysis of the Algorithm. In this paper, based on each sample group in the training set and prediction set, the actual fraud rate and the predicted fraud rate in each group are calculated to evaluate the prediction ability of the model. The actual fraud rate and predicted fraud rate values and the line chart of each group are shown in Tables 2 and 3 and Figure 5, respectively.

Through the analysis of the above chart data, we can find that the algorithm proposed in this paper has a good predictive ability regardless of the balanced distribution of data or the extreme distribution of imbalance.

In order to verify the effectiveness of the proposed method, the algorithm in this paper is compared with the following four methods:

- (1) Lin et al. [19] constructed a fraud detection system by combining neural networks and association analysis.
- (2) Ryan et al. [4] used deep belief networks to extract behavioural features and SVMs to detect fraudulent abnormal behaviours.
- (3) Liu et al. [21] used spatial-temporal convolutional neural networks to extract fraudulent behaviour features for abnormal behaviour detection and location.

- (4) Peng et al. [22] used image saliency information and multiscale optical flow histogram as low-level features. Then, the deep learning network PCANet is used to extract more effective features from these underlying features for abnormal behaviour detection.

In the experiment, we first extracted the dense trajectory. Secondly, randomly select 5 million trajectories to train SDA and used the K-means clustering method to get 370 appearance visual words and 430 motion visual words. Then, we randomly selected 800 normal behaviour learning feature fusion parameters to obtain the parameters $w_1 = 0.3, w_2 = 0.7$. Finally, we learned sparse dictionaries with 800 normal behaviours. All abnormal fraud behaviour samples and the remaining 200 normal samples are used as test samples. The test results are shown in Table 4.

From Table 4, we draw a clustered bar chart of the table, as shown in Figure 6. Through a comprehensive comparison, we find that the algorithm proposed in this paper has not only a high accuracy rate but also a high recall rate. While other algorithms have different degrees of defects, these are not suitable for practical engineering applications.

Figure 7 shows the ROC curves of several detection methods. The results show that the algorithm proposed in this paper has better detection accuracy than the other four methods for the detection of financial fraud. Literature [19] adopts the combination of neural network and association analysis for abnormal behaviour detection, resulting in a

TABLE 2: Comparison of actual fraud rate and predicted fraud rate in the training set.

| Sample number | Number of samples | Contains the number of frauds | Actual fraud rate | Predicted fraud rate |
|---------------|-------------------|-------------------------------|-------------------|----------------------|
| 1 | 300 | 122 | 39.665% | 39.495% |
| 2 | 300 | 50 | 15.258% | 15.021% |
| 3 | 300 | 35 | 10.178% | 9.885% |
| 4 | 300 | 32 | 9.156% | 8.882% |
| 5 | 300 | 26 | 7.125% | 6.647% |
| 6 | 300 | 20 | 5.085% | 5.472% |
| 7 | 300 | 20 | 5.731% | 5.658% |
| 8 | 300 | 16 | 3.731% | 3.115% |
| 9 | 300 | 14 | 3.058% | 3.102% |
| 10 | 300 | 12 | 2.389% | 2.178% |

TABLE 3: Comparison of actual fraud rate and predicted fraud rate in the test set.

| Sample number | Number of samples | Contains the number of frauds | Actual fraud rate | Predicted fraud rate |
|---------------|-------------------|-------------------------------|-------------------|----------------------|
| 1 | 300 | 116 | 37.632% | 35.946% |
| 2 | 300 | 47 | 14.246% | 14.268% |
| 3 | 300 | 45 | 13.596% | 12.587% |
| 4 | 300 | 35 | 10.175% | 10.158% |
| 5 | 300 | 30 | 8.457% | 8.328% |
| 6 | 300 | 20 | 5.082% | 4.389% |
| 7 | 300 | 23 | 6.125% | 5.821% |
| 8 | 300 | 14 | 3.059% | 2.985% |
| 9 | 300 | 12 | 2.379% | 3.109% |
| 10 | 300 | 10 | 1.695% | 1.479% |

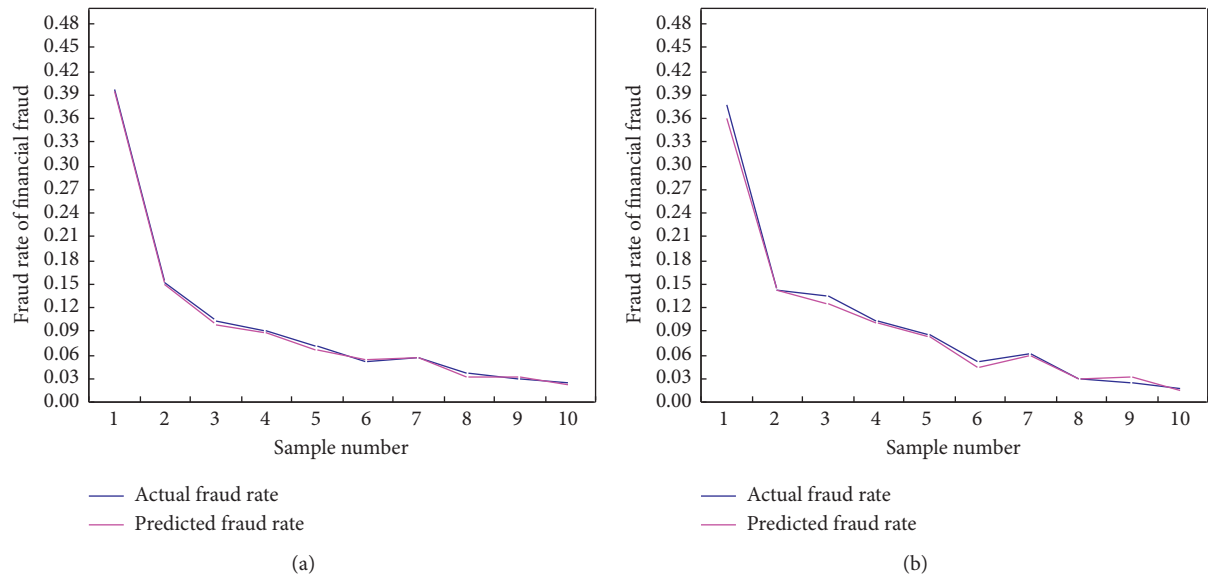


FIGURE 5: Grouped predictions for training and test sets. (a) Training. (b) Testing.

lower abnormal detection rate than other algorithms. Literature [4] uses deep confidence networks to extract behavioural features and uses SVM to detect fraudulent behaviours. However, SVM takes a long time to find the optimal parameters. Literatures [21, 22] use space-time convolutional neural networks and PCANet networks to extract behavioural features, which can describe behaviours well, so their detection rate exceeds 90%. However, the detection rate of these two algorithms is lower than the algorithm proposed in this paper, because these two algorithms require a large number of samples for training deep

TABLE 4: Test results of different algorithms.

| Algorithm | Precision | Recall | FM | AUC |
|-----------------|-----------|---------|---------|---------|
| Literature [19] | 82.853% | 72.581% | 70.831% | 91.021% |
| Literature [4] | 93.581% | 89.367% | 88.213% | 96.123% |
| Literature [21] | 91.952% | 85.987% | 75.902% | 95.368% |
| Literature [22] | 93.598% | 90.002% | 89.291% | 96.225% |
| This paper | 97.582% | 93.691% | 90.781% | 99.687% |

learning networks, and the behaviour database samples are relatively small. At the same time, in order to balance the computational cost, these two algorithms extract behaviours

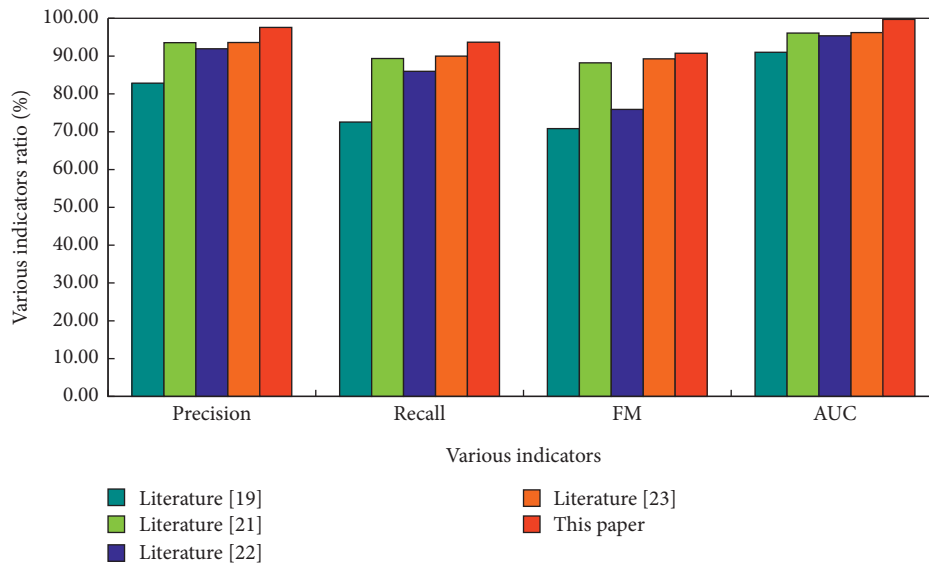


FIGURE 6: Cluster graph of test results of different algorithms.

that usually indicate a downsampling strategy, which leads to loss of information. Due to the rich motion information near the dense trajectory, the method proposed in this paper uses the powerful learning ability of the stacked denoising encoder to extract effective behavioural features. At the same time, the deep network does not directly extract the entire behaviour features, but only extracts the characteristics of the sampling points in the behaviour area. Moreover, the number of these sampling points is sufficient to train the deep network, so a large number of samples are not required to train the deep network, which solves the shortage and impact of training samples on deep learning, so the fraud detection rate of the algorithm proposed in this paper is higher than several other methods.

4.3. Time Performance Analysis of the Algorithm. In order to verify the superiority of the fraud detection model of the algorithm proposed in this paper, this paper mainly conducts two experiments. The first experiment is to compare the time efficiency of the algorithm in this paper and the other four comparison algorithms. The second experiment is to analyse the acceleration ratio of the cluster, that is, to compare the time efficiency of different classifier detection with different numbers of nodes.

In this experiment, we selected different amounts of data. Among them, the system selects four nodes for parallel computing. The experimental results are shown in Figure 8. It can be seen from Figure 8 that when the amount of data is small, the algorithm proposed in this paper has the highest time efficiency. This is because the algorithm proposed in this paper uses the deep network not to directly extract the entire behaviour features, but only extracts the features of the sampling points in the behaviour area, and the number of these sampling points is sufficient to train the deep network. Therefore, a large number of samples are not required to train the deep network. With the increase of data

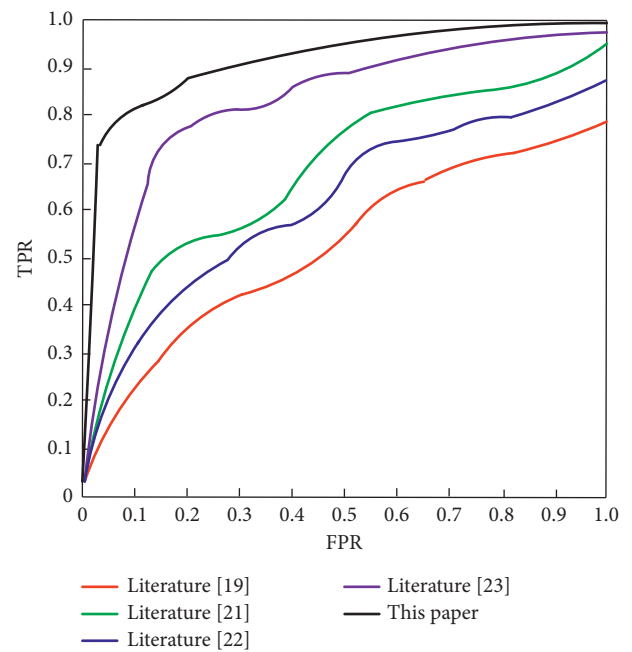


FIGURE 7: ROC curves of different algorithms.

volume, other algorithms will show greater fluctuations in time.

We continue to increase the number of nodes to detect the time efficiency of different numbers of nodes to complete fraud detection tasks. For the number of nodes, select 2 nodes, 4 nodes, 6 nodes, 8 nodes, 10 nodes, 12 nodes, 14 nodes, 16 nodes, 18 nodes, and 20 nodes for statistics. The experimental results are shown in Figure 9. From Figure 9, we can find that the running time of the algorithm proposed in this paper decreases with the increase of the number of nodes, but there is no big change in the comparison algorithm. It can be seen that the increase of nodes can improve the classification efficiency of the financial fraud detection

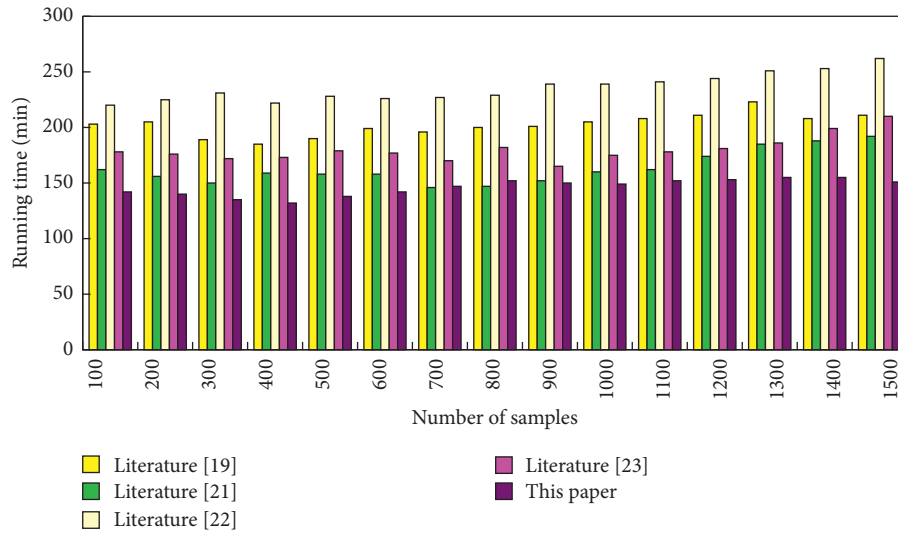


FIGURE 8: Time performance comparison of different algorithms.

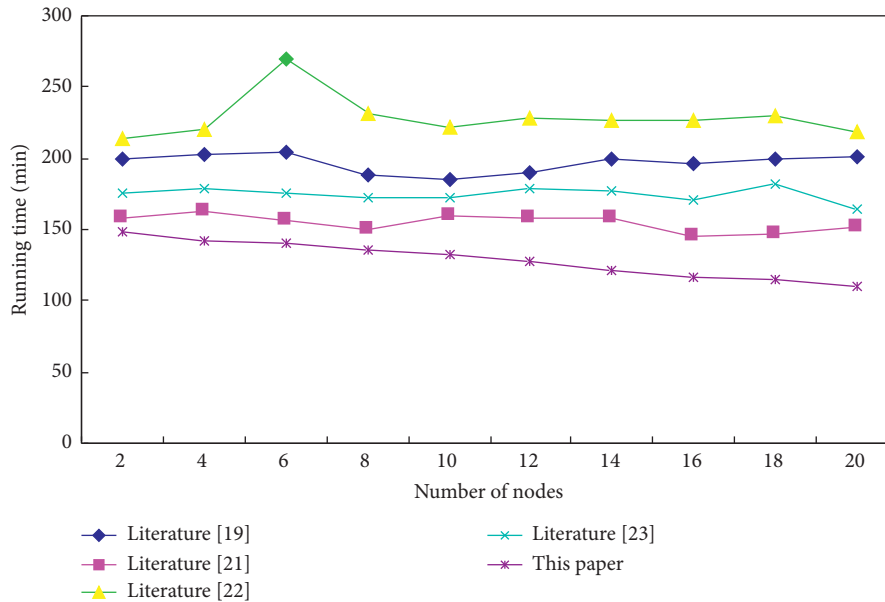


FIGURE 9: Performance comparison of different algorithms at different node numbers.

algorithm proposed in this paper. This shows that when we are faced with a larger and larger amount of data, we can simply increase the number of machines to improve the execution efficiency of fraud detection algorithms.

5. Conclusion

Fraud is very common in the field of financial services. Large database management systems are basic system software widely used by financial institutions. The use of data mining in large database systems is an advanced technical means for detecting financial fraud. It is an effective method to detect financial fraud by mining and analysing the data in a large amount of processing business data and finding the corresponding rules, rules and conclusions, and then combining manual analysis. In

this paper, the encoder is first used to extract the appearance and motion features of the behaviour, and in order to reduce the computational complexity, the feature extraction is constrained to the space-time volume of the dense trajectory. Secondly, the deep learning model is used to transform the features into behavioural visual word representations, and the feature fusion is performed using the weighted correlation method to improve the classification ability of the features. Finally, the sparse reconstruction error is used to judge the abnormality of fraud. The results show that the algorithm proposed in this paper can effectively learn the essential characteristics of the data and has a higher detection rate and lower computational complexity. Although this article has achieved good experimental results, the results are still in the experimental stage. Our next research plan is to apply the

algorithm to the actual environment. By obtaining data in the actual environment, we can further optimize our algorithm.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no known conflicts of interest or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the innovation spark project of Sichuan University: Research on the evolution and value realization mechanism of knowledge advantage in knowledge chain (project no.: 2019hhs-18); Research on the formation, maintenance and transformation of knowledge advantage in knowledge chain to competitive advantage (project no.: 71971146); Chengdu soft science research project “giving full play to the role of Chengdu’s scientific and technological innovation” and promoting the collaborative innovation of “five regions” (project no.: 2019-rk00-00182-zf).

References

- [1] Y. Gao, J. B. Kim, D. Tsang et al., “Go before the whistle blows: an empirical analysis of director turnover and financial fraud,” *Review of Accounting Studies*, vol. 22, no. 1, pp. 1–41, 2017.
- [2] D. Burnes, C. R. Henderson, C. Sheppard, R. Zhao, K. Pilemer, and M. S. Lachs, “Prevalence of financial fraud and scams among older adults in the United States: a systematic review and meta-analysis,” *American Journal of Public Health*, vol. 107, no. 8, pp. e13–e21, 2017.
- [3] Y. Pillemer, T. Sugawara, Y. Ohkusa et al., “Severe abnormal behavior incidence after administration of neuraminidase inhibitors using the national database of medical claims,” *Journal of Infection and Chemotherapy*, vol. 24, no. 3, pp. 177–181, 2018.
- [4] R. R. Darby, A. Horn, F. Cushman, and M. D. Fox, “Lesion network localization of criminal behavior,” *Proceedings of the National Academy of Sciences*, vol. 115, no. 3, pp. 601–606, 2018.
- [5] L. Bramslw, G. Naithani, A. Hafez, T. Barker, N. H. Pontoppidan, and T. Virtanen, “Improving competing voices segregation for hearing impaired listeners using a low-latency deep neural network algorithm,” *The Journal of the Acoustical Society of America*, vol. 144, no. 1, pp. 172–185, 2018.
- [6] J. Zhang, W. Chen, M. Gao et al., “Intelligent adaptive coherent optical receiver based on convolutional neural network and clustering algorithm,” *Optics Express*, vol. 26, no. 14, pp. 18684–18698, 2018.
- [7] O. Hu, J. Chen, P. Gao et al., “Fusion of near-infrared and fluorescence spectroscopy for untargeted fraud detection of Chinese tea seed oil using chemometric methods,” *Journal of the Science of Food and Agriculture*, vol. 99, no. 5, pp. 2285–2291, 2019.
- [8] J. E. Fisch, J. B. Gelbach, and J. Klick, “The logic and limits of event studies in securities fraud litigation,” *Texas Law Review*, vol. 96, no. 3, pp. 553–621, 2018.
- [9] M. L. Merchant, M. E. Brier, M. S. Slaughter, J. B. Klein, and K. R. McLeish, “Biomarker enhanced risk prediction for development of AKI after cardiac surgery,” *BMC Nephrology*, vol. 19, no. 1, pp. 102–111, 2018.
- [10] M. K. Khormuji, M. Bazrafkan, M. Sharifian, S. J. Mirabedini, and A. Harounabadi, “Credit card fraud detection with a cascade artificial neural network and imperialist competitive algorithm,” *International Journal of Computer Applications*, vol. 96, no. 25, pp. 1–9, 2014.
- [11] A. Liu, X. Deng, L. Ren, and Y. Liu, “An inverse power generation mechanism based fruit fly algorithm for function optimization,” *Journal of Systems Science and Complexity*, vol. 32, no. 2, pp. 634–656, 2019.
- [12] Y. Hu, G. Zhou, C. Zhang et al., “Identify compounds’ target against alzheimer’s disease based on in-silico approach,” *Current Alzheimer Research*, vol. 16, no. 3, pp. 193–208, 2019.
- [13] S. Mosbach, A. Braumann, P. L. W. Man, C. A. Kastner, G. P. E. Brownbridge, and M. Kraft, “Iterative improvement of Bayesian parameter estimates for an engine model by means of experimental design,” *Combustion and Flame*, vol. 159, no. 3, pp. 1303–1313, 2012.
- [14] N. E. Ordyan, S. G. Pivina, Y. O. Fedotova, and V. V. Rakitskaya, “Formation of an anxious-depressive state in an experimental model of post-traumatic stress disorder in prenatally stressed female rats,” *Neuroscience and Behavioral Physiology*, vol. 43, no. 6, pp. 712–717, 2013.
- [15] N. M. Menon, “Information spillover and semi-collaborative networks in insurer fraud detection,” *MIS Quarterly*, vol. 42, no. 2, pp. 407–426, 2018.
- [16] M. D. M. Roldán-García, J. García-Nieto, and J. F. Aldana-Montes, “Enhancing semantic consistency in anti-fraud rule-based expert systems,” *Expert Systems with Applications*, vol. 90, pp. 332–343, 2017.
- [17] G. Mhatre, O. Almeida, D. Mhatre, and P. Joshi, “Credit card fraud detection using Hidden Markov model,” *International Journal of Computer Science & Information Technologies*, vol. 5, no. 1, pp. 37–48, 2014.
- [18] K. Jana, E. Samková, and L. Hasoňová, “Food fraud detection by Czech agricultural and food inspection authority in retail market,” *British Food Journal*, vol. 120, no. 4, pp. 930–938, 2018.
- [19] B. J. Lin, X. Li, and W. L. Yu, “Binary neutron stars gravitational wave detection based on wavelet packet analysis and convolutional neural networks,” *Frontiers of Physics*, vol. 15, no. 1, pp. 1–8, 2020.
- [20] W. Cao, J. Yuan, Z. He, Z. Zhang, and Z. He, “Fast deep neural networks with knowledge guided training and predicted regions of interests for real-time video object detection,” *IEEE Access*, vol. 6, pp. 8990–8999, 2018.
- [21] Q. Liu, B. Wang, and Y. Zhu, “Short-term traffic speed forecasting based on attention convolutional neural network for arterials,” *Computer-Aided Civil and Infrastructure Engineering*, vol. 33, no. 11, pp. 999–1016, 2018.
- [22] C. Peng, M. Liu, X. P. Yuan, L. X. Zhang, and J. F. Man, “A new method for abnormal behavior propagation in networked software,” *Journal of Internet Technology*, vol. 19, no. 2, pp. 489–498, 2018.
- [23] V. Botes and A. Saadeh, “Exploring evidence to develop a nomenclature for forensic accounting,” *Pacific Accounting Review*, vol. 30, no. 2, pp. 135–154, 2018.