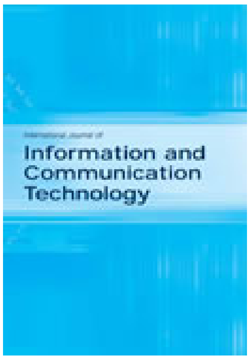


International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

Financial fraud recognition based on deep learning and textual feature

Weiyi Chen

Article History:

Received:	27 October 2024
Last revised:	22 November 2024
Accepted:	22 November 2024
Published online:	02 January 2025

Financial fraud recognition based on deep learning and textual feature

Weiyi Chen

Monitoring and Audit Department of the Financial Shared Center,
National Energy Group Qinghai Electric Power Co., Ltd.,
Xining, 8180000, China
Email: 13975386663@163.com

Abstract: Financial fraud refers to the egregious breach of trust that uses improper means to distort accounting information, which negatively affects the company's operation. Intending to the issues of ignoring text features in existing research, a financial fraud recognition method based on deep learning and text features is designed. The method starts with preprocessing financial indicators and uses BiLSTM to extract sentiment features from the text of the management discussion and analysis (MD&A) chapter. Then the parallel residual network is used to select the financial indicator variables and textual sentiment variables in depth, the selected two variables are inputted into the dual-channel CNN for feature extraction, and feature enhancement and fusion are carried out using multi-head attention, and the recognition results are outputted through softmax. The experimental results show that the proposed model achieves better financial fraud identification, with the accuracy and AUC reaching 91.35% and 98.52%, respectively.

Keywords: financial fraud identification; deep learning; DL; text feature; BiLSTM; CNN.

Reference to this paper should be made as follows: Chen, W. (2024) 'Financial fraud recognition based on deep learning and textual feature', *Int. J. Information and Communication Technology*, Vol. 25, No. 12, pp.1–15.

Biographical notes: Weiyi Chen received her Bachelor's degree from University of Hunan University of Technology Nanhu College. She is the Head of the Monitoring and Audit Department of the Financial Shared Center of Qinghai Electric Power Co., Ltd. under the National Energy Group. Her research interests include financial statement analysis, ratio analysis, cost-benefit analysis, and risk assessment.

1 Introduction

Since the emergence of capital markets, financial fraud has occurred in various countries, which not only infringes on the interests of investors, but also undermines the role of the stock market in regulating the allocation of resources (Du, 2021). After more than 30 years of development of China's securities market, the momentum is still strong, and the number of listed companies is increasing rapidly, but the matching regulatory mechanism and standardised system are not yet mature enough, resulting in numerous incidents of financial fraud. Traditional financial fraud identification relies heavily on the auditor's empirical judgement, which is clearly inefficient and costly in the face of the current large market size and increasingly 'sophisticated' fraud techniques (Hilal et al., 2022). Therefore, it is of great practical significance for both regulators and investors to construct efficient and accurate financial fraud identification models.

Early financial fraud identification is mainly based on combining indicator variables to establish a classical statistical model. Tang and Karim (2019) constructed an triangular model based on the triangular theory by selecting 46 variables from the three factors of pressure, opportunity, and excuses, and tested and proved the ability of the indicators in the model to identify financial fraud. Gong et al. (2022) proposed a method of identifying financial malpractice by combining Benford's law with a logistic model, and after simulation using financial data of listed companies in China, it was found that the logistic model with Benford's factors had a higher accuracy rate.

Financial fraud recognition models based on traditional statistical models rely on preset assumptions, have limited ability to process data with complex and non-linear relationships, and have high computational costs. The machine learning-based financial fraud recognition model does not rely on preset assumptions, and is able to automatically learn patterns from data, enough to improve the recognition accuracy through a large amount of data training. Nami and Shajari (2018) selected 43 indicator variables based on the kinetic theory of financial fraud and used principal component analysis to select the principal components of the indicator variables for financial incomparable identification through random forests with a prediction accuracy of 79.2%. Ali et al. (2022) used textual information from financial reports of listed companies to distinguish between fraudulent and genuine reports, the model utilises a text-based approach and SVM algorithm, and the experimental results show that the model helps to identify financial fraud. Zhao and Bai (2022) developed a systematic text analysis framework for identifying financial malpractice based on the theory of systemic functional linguistics. The framework extracts word-level and document-level features as inputs to SVM and achieves high recognition accuracy. ML-based methods have improved the recognition performance compared with traditional methods, and a few scholars have found the role of text information in identifying financial fraud (Choi et al., 2017), but ML methods have limited ability for natural language processing, and the features of text information have not been fully utilised.

Subsequently, deep learning (DL) has been applied to financial fraud identification due to its powerful feature extraction capabilities. Forough and Momtazi (2022) used BP neural networks (BPNN) and probabilistic neural networks (PNN) on a dataset of 202 publicly traded companies and the results showed that PNN outperforms the BPNN as well as other ML models in terms of identification accuracy. Nakashima et al. (2022) confirmed that the use of management discussion and analysis (MD&A) section in periodic financial reports can unearth some of the financial reporting frauds. Xiuguo and Shengyong (2022) extracted sentiment variables from MD&A text messages to recognise financial fraud using CNN. Wang et al. (2023) utilised a hierarchical attention network (HAN) with a long short-term memory (LSTM) network encoder to extract text features from MD&A, however, the method did not extract the financial fraud text features sufficiently, and the recognition accuracy only reached 88.79%. Benchaji et al. (2021) combined textual content of financial metrics, non-financial metrics, and MD&A, and used LSTM and transformer for financial fraud recognition, and the recognition performance of this study improved significantly.

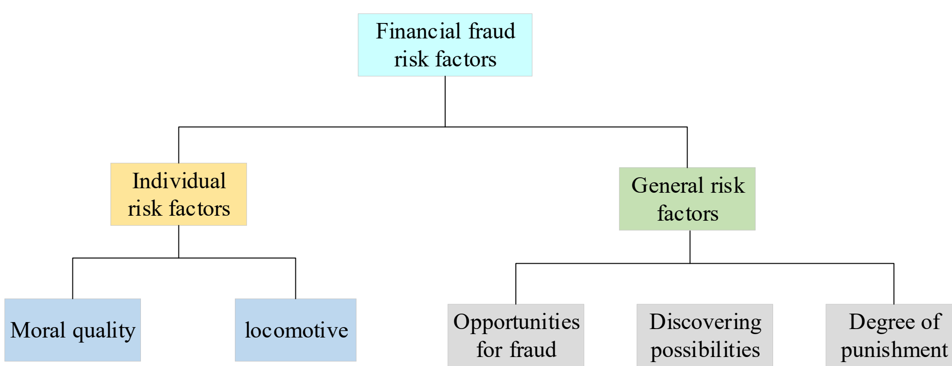
This paper fused the financial index features of company's financial statements and MD&A text sentiment features to construct a financial fraud recognition method relied on DL. The financial indicators were first screened using Pearson's correlation coefficient, and a two-sample test of means was performed to eliminate redundant indicators. Secondly, the Bert model is utilised for MD&A text word vector embedding, where the word vectors are used as input sequences to BiLSTM, and textual sentiment variables are extracted and fused into the model. Then, the parallel residual network is used to select the financial indicator variables and textual sentiment variables in depth, and the two variables are fed into the 1D CNN and text CNN for feature extraction, respectively, and the important features are focused on using the multihead attention mechanism (MHA) to splice to obtain the fusion variables of the financial indicator variables and the textual sentiment variables. The financial fraud recognition results are output through softmax. The experimental outcome indicates that the accuracy and AUC of the suggested method are improved by 0.69%–9.79% compared with the comparison method, which has a better effect of financial fraud identification.

2 Relevant theoretical foundations

2.1 Financial fraud risk factor theory

The fraud risk factor theory is an extension and refinement of the traditional Iceberg, Triangle and GONE theories, as shown in Figure 1. The theory categorises opportunity, exposure as general risk factors and greed, need as individual risk factors (Aghghaleh and Mohamed, 2014).

Figure 1 Fraud risk factor theory (see online version for colours)



The theory has gradually made a distinction between internal and external, overall and individual, in analysing the drivers of fraud, and has become more specific and comprehensive in describing fraud factors. The theory is based on the development of the fraud GONE theory of fraud motivation theory, for the analysis of fraud motivation gradually have internal and external, overall and individual distinction, due to the existence of fraud risk factors arising from the existence of the fraud risk factor, according to whether it can be controlled by the external environment, the fraud risk factor of the fraud GONE theory to improve the analysis of the fraud factors, can be more comprehensive. It can be seen that the factors that lead to financial fraud are complex, and when it comes to fraud prevention, only one of them needs to be destroyed in order to stop the occurrence of financial fraud, which provides ideas for identifying financial fraud.

2.2 Convolutional neural network

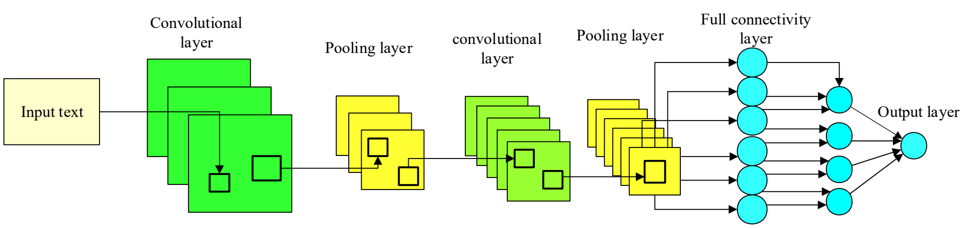
CNN is a specially structured feed-forward neural network capable of automatically extracting important features such as word vectors and syntactic structures in text, which is useful for tasks such as text categorisation and sentiment analysis (Li et al., 2021). CNN consists of an input level, a convolutional level, a pooling level and an output level as shown in Figure 2.

The main principle is to extract features by convolution operation with different sizes of convolution kernels, and then streamline and integrate the features extracted from the convolution level in the pooling level to get the most effective and informative features, and then send the features obtained from the pooling layer to the classifier to classify them, and finally output the results. The convolution is calculated as bellow:

$$x_j^k = f\left(\sum\nolimits_{i \in R_j} x_i^{k-1} * w_{ij}^k + b_j^k\right) \tag{1}$$

where w_{ij}^k is the weight value of the convolution kernel, b_j^k is the bias value of the j^{th} neuron in the k^{th} level, R_j is the set of feature maps, and f is the activation function

Figure 2 CNN network architecture (see online version for colours)



2.3 Long- and short-term memory network

CNNs have more layers and larger number of parameters, which leads to higher computational complexity of the model, more time and resources are required for the training and inference process, and it is difficult to capture the long-term temporal dependencies of the data sequences. LSTMs are able to efficiently capture and model long-term dependencies through the introduction of a gating mechanism, which is suitable for processing sequential data. LSTM is one of the RNN variants, which introduces the concept of gates for solving the problem that it is difficult to utilise the information over long distances (Zhao et al., 2017). LSTM consists of an input gate, an output gate and an oblivion gate. At time t, the LSTM has three inputs: the input required at the current time, the short-term memory state h_{t-1} at the previous time, and the long-term memory state c_{t-1} at the previous time of the LSTM, and the LSTM has two outputs: the short-term memory state h_t at the current time, and the long-term memory state c_t.

The LSTM first decides which information in state c is not needed through a forgetting gate, then decides which information in the current state needs to be added to state c. Secondly, the initial state c is multiplied by the old state just entered to get the new long-term memory value c. The final decision of what needs to be output is made by passing the state c through the tanh layer and filtering it through the Sigmoid layer, multiplying the values of the two states to determine the information that needs to be output, as shown in the following equation:

$$f = \delta\left(V_f x_t + W_f h_{t-1} + b_f\right) \tag{2}$$

$$i = \delta\left(V_i x_t + W_i h_{t-1} + b_i\right) \tag{3}$$

$$o = \delta\left(V_o x_t + W_o h_{t-1} + b_o\right) \tag{4}$$

$$c = f_t o c_{t-1} + i_t o \delta_c\left(V_c x_t + W_c h_{t-1} + b_c\right) \tag{5}$$

where b is the bias, V and W are the weights corresponding to each stage, δ is the nonlinear activation Sigmoid function, i and o represent the input and output gates, respectively

3 Pre-processing of financial indicators based on financial reporting data

Financial reports contain more financial indicators and studies have found that some financial indicators are more important than others (Aftabi et al., 2023). This paper collects financial reporting data indicators from the Cathay Pacific database, and based on the financial fraud risk factor theory in Section 2.1 and existing research, six categories of indicators, namely, solvency, operating ability, profitability, cash flow analysis, risk level, and development ability, are used as auxiliary indicators for identifying financial fraud, with a total of n indicators. Due to the large number of indicators, redundant indicators are eliminated. Pearson's correlation coefficient is an important measure of the degree of linear correlation between two variables X and Y. The formula is as follows:

$$\rho = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} \tag{6}$$

where ρ is between –1 and 1. The closer it is to 1, the stronger the correlation between X and Y is. If $0.8 < |\rho| \leq 1$, it means that X and Y are extremely strongly correlated, if $0.6 < |\rho| \leq 0.8$, it means that X and Y are strongly correlated, and if $|\rho| \leq 0.6$, it means that X and Y are weakly correlated. $Cov(X, Y)$ is the covariance of X and Y. σ_X and σ_Y are the standard deviations of X and Y, respectively.

According to the above formula the Pearson correlation coefficient between the indicators can be obtained, in this paper, one of the indicators of $|\rho| \geq 0.8$ is eliminated, and the remaining indicators m, $m < n$. For the remaining indicator variables, a two-sample test of means was conducted by treating them as coming from two different aggregates, financial and non-financial fraud. Let any characteristic variable to be tested be X, with means μ_1 and μ_2 in different aggregates, with overall variances δ_1^2 and δ_2^2, and with original and alternative hypotheses H_0: $\mu_1 = \mu_2$, H_1: $\mu_1 \neq \mu_2$. X obeys a normal distribution $\bar{X}_1 \sim N(\mu_1, \sigma_1^2 / n_1), \bar{X}_2 \sim N(\mu_2, \sigma_2^2 / n_2)$ in both aggregates, then $\bar{X}_1 - \bar{X}_2 \sim N(\mu_1 - \mu_2,\ (\sigma_1^2 / n_1) + (\sigma_2^2 / n_2)),$ the location of the variance of the aggregates needs to be estimated in terms of the sample variance, as follows.

$$s_1^2 = \frac{\sum_{i=1}^{n} \left(X_{1i} - \bar{X}_1 \right)^2}{n_1 - 1}, s_2^2 = \frac{\sum_{i=1}^{n} \left(X_{2i} - \bar{X}_2 \right)^2}{n_2 - 1} \tag{7}$$

When the original hypothesis holds, i.e., $\mu_1 = \mu_2$, the following equation is available according to Welch-Satterwaite's T-est.

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \tag{8}$$

Given the significance level $\alpha = 0.05$, the two-tail test is adopted. When the t-value of the test is less than 0.05, the null hypothesis is rejected and the mean values of X_1 and X_2 in the two populations are considered to be significantly different, and the characteristic variable X is retained. Otherwise, the X variable is removed. After the above feature screening, q variable features are finally obtained, and Z-SCORE is used to standardise the data, where μ and σ represent the sample mean and variance respectively.

$$X' = \frac{X - \mu}{\sigma} \tag{9}$$

4 MD&A text emotion extraction based on Bi-LSTM

For MD&A text data, bi-directional bi-LSTM is more mature in sentiment analysis, with higher efficiency and accuracy (Xu et al., 2019), so in this paper, we firstly train word vectors with practical BERT model, and then utilise bi-LSTM to extract MD&A text sentiment, as shown in Figure 3.

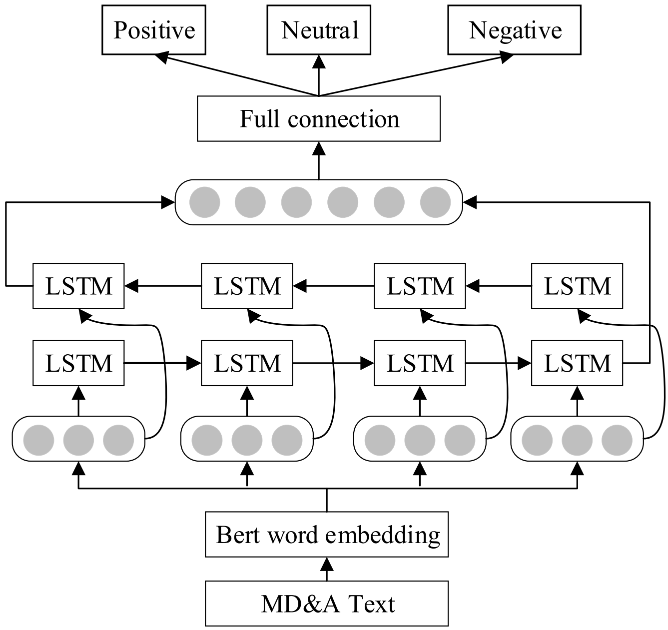
First, the MD&A text sequence is input into the BERT model, and the word vector is output after being constructed by bidirectional transformer. Then, the word vector is fused according to the pre-determined word boundaries, and the fusion strategy is mainly based on the 'sum average of all words', the specific formula is as follows.

$$W_i = \frac{\left(x_1 + x_2 + \cdots + x_n \right)}{n} \tag{10}$$

where x_i denotes each of the word vectors that make up the word vector, and W_i denotes the word vector representation obtained by summing and averaging all the word vectors.

Then the word vectors output from BERT are used as the input sequences of BiLSTM and transferred to the bidirectional recurrent neural network for further learning of the word vectors. The fused word vector trained by BERT is denoted as $W^i = \{W_1, W_2, \ldots, W_n\}$, where W^i denotes the vector matrix of the i^{th} utterance, W_i denotes the vector representation of the i^{th} word, and n denotes the maximum sentence length.

Figure 3 Text sentiment extraction model diagram



Then the sequence of word vectors $\{W_1, W_2, \ldots, W_n\}$ is input to the forward LSTM to get its hidden output $\{\overrightarrow{h_1}, \overrightarrow{h_2}, \ldots, \overrightarrow{h_n}\}$, given the input $\{W_n, W_{n-1}, \ldots, W_1\}$ of the backward LSTM, its hidden output $\{\overleftarrow{h_n}, \overleftarrow{h_{n-1}}, \ldots, \overleftarrow{h_1}\}$ is obtained. The output $\{h_1, h_2, \ldots, h_n\}$ of the Bi-LSTM is obtained by splicing $\{\overrightarrow{h_1}, \overrightarrow{h_2}, \ldots, \overrightarrow{h_n}\}$ and $\{\overleftarrow{h_n}, \overleftarrow{h_{n-1}}, \ldots, \overleftarrow{h_1}\}$. The output is passed into the full join, and the sentiment categorisation of MD&A text is obtained by Softmax.

After obtaining the sentiment categorisation of all stock review texts, the sentiment categorisation is combined by company and year. Let the number of MD&A texts with positive sentiment of a company in year t be P and the number of MD&A texts with negative sentiment be N, then the MD&A text sentiment of the company in that year is shown in equation (11).

$$s = \left(\frac{P - N}{P + N}, t \right) \tag{11}$$

Therefore, the MD&A text emotion vector s contains two elements. If the value of $(P - N) / (P + N)$ is larger, it indicates that the company's emotion in the year is more positive, and vice versa, it is more negative. Moreover, when the positive or negative emotion accounts for a large proportion, the value will be closer to 1 or –1, which can better reflect the overall trend of feeling. Overall, the affective tendency of the text is related to $(P - N) / (P + N)$. When close to –1, the overall affective tendency is negative, and when close to 1, the overall affective tendency is positive. t is the time information of MD&A text, and retaining the time information helps the model to make a judgement on the credibility of emotion.

5 Financial fraud recognition based on DL and textual features

5.1 Variable selection based on parallel residual networks

After preprocessing the financial report data and MD&A text, this paper designs a financial fraud recognition model based on DL and text features, as shown in Figure 4. The model contains a parallel residual network layer, a two-channel CNN feature extraction layer, and a recognition output layer. Firstly, the parallel residual network is used to select the financial indicator variables and MD&A text sentiment variables in depth, and then the selected indicator variables and sentiment variables are inputted into the dual-channel CNN for feature extraction, and feature enhancement and fusion are carried out by utilising multi-attention, and finally the results of financial fraud recognition are outputted by softmax.

In the study of financial fraud, due to the high complexity and strong fitting ability of the DL model, it is easy to overfitting on the financial fraud dataset. Meanwhile, if the number of network layers is too deep, the gradient disappears easily. Therefore, to improve the recognition efficiency, this paper adds the residual structure (Wang et al., 2021a) on the basis of the deep neural network, so that it constitutes a residual block, and multiple residual blocks are connected together to form the parallel residual network module of the model in this paper.

Residual network is a DL model mainly used to solve the gradient vanishing and gradient explosion problems encountered by deep neural networks during the training process. Compared with traditional neural networks, residual networks enable the network to perform feature extraction and optimisation more efficiently by introducing constant connections. The residual block uses a short circuit to add elements of the input to the output of the fully connected level. If the input of the current residual block is x_0, the computation process of the residual block and its output x_{out} are shown below:

$$x_i = \begin{cases} f\left(x_{i-1}v_i + b_i\right), & 1 < i < n \\ x_{i-1}v_i + b_i, & \text{else} \end{cases} \tag{12}$$

$$x_{out} = f\left(x_0 + x_n\right) \tag{13}$$

where x_i is the hidden output of the i^{th} neural network layer in the residual block, v_i and b_i are the weight and bias of the i^{th} neural network layer, respectively, and $f(\times)$ is the activation function ReLU used in the residual block. If the required optimal solution is $H(x) = x$, then the residual mapping is $F(x) = H(x) - x$. When the network reaches the optimal state, $F(x)$ is infinitely close to 0. If we continue to deepen the network, it will remain in the optimal state, thus avoiding performance degradation and enabling the model to learn the relationship between variables better.

5.2 Feature extraction based on two-channel CNN and multi-head attention mechanism

After variable selection for financial indicator variables and MD&A text sentiment variables, the two variables were fed into 1D CNN and text CNN for feature extraction as follows:

1 Financial indicator variable feature extraction. After the residual network, the financial indicator variable X is firstly convolved by 1D CNN, and the convolution feature F_a can be denoted as $F_a = X * L + b$, where L is the residual convolution kernel, b is the bias, and * denotes the convolution operation. After that, pooling operation is performed on the convolutional feature $F_P = pool(F_a)$. The pooled feature can be denoted as F_c. At the same time, the nonlinear feature $F_c =$ LeakyReLu(F_P) can be obtained as F_z by Leaky ReLU.

$$F_z = \text{LeakyReLU}\left(F_N V + \alpha\right) \tag{14}$$

where V and α are the weights and biases of the fully connected layer.

2 MD&A text sentiment variable feature extraction. The text emotion variables are inputted into TextCNN, and the MD&A text word vector $\{W_1, W_2, \ldots, W_n\}$ is shifted according to a certain length by the filter of convolution kernel to obtain the deep local information of the text data, and the output feature vector of convolution layer is obtained.

$$Z = f\left(G \cdot W_i + E\right) \tag{15}$$

where h represents the convolution kernel size, G denotes the weight matrix of the output convolution kernel, E denotes the bias parameter, f is the activation function, and the vector consisting of Z is the feature vector extracted from the convolutional layer.

Then use the MHA to focus on the key parts of the financial indicator variable features and textual sentiment features to obtain different levels of hidden information. MHA can learn multiple features in parallel by using multiple attention matrices (Wang et al., 2021b), each of which is called a head, and the model in this chapter uses a self-attention mechanism for each attention head. It allows the model to learn different aspects of the input data in parallel in different representation subspaces, thus increasing the model's flexibility and ability to capture more complex feature relationships. The computational flow of the multi-head attention mechanism is shown as follows:

$$S_{Att} = soft\max\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{16}$$

$$head_i = S_{Att}\left(QU_i^Q, KU_i^K, VU_i^V\right) \tag{17}$$

$$MHA(Q, K, V) = Concat\left(head_1, head_2, \ldots, head_n\right)U^a \tag{18}$$

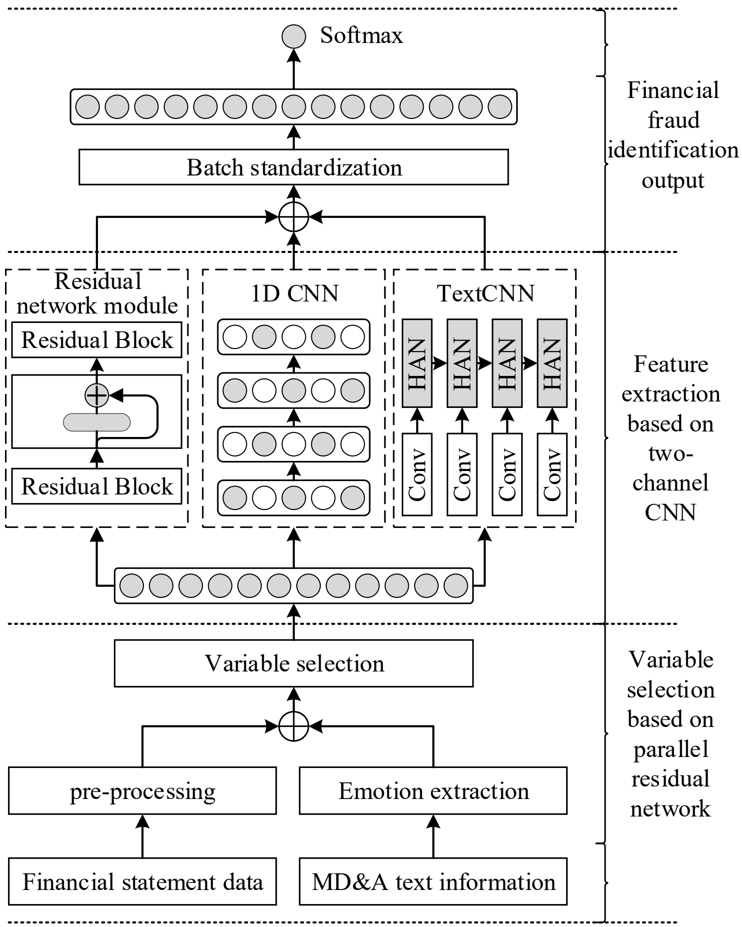
where U_i^Q, U_i^K, U_i^V are weight matrices used to linearly transform Q, K, and V. $\sqrt{d_k}$ is a scaling factor, the size of which is generally set to the dimension of the feature to prevent the gradient from disappearing when the dot product result is too large when calculating the attention score, and U^a is a parameter matrix used to linearly transform the results of the MHA.

To extract the key features at different levels, each attention head uses a different linear layer to linearly transform the features of the financial indicator variables and the features of the textual sentiment variables, and the n attention heads obtain n sets of Q, K, and V matrices, and each set of matrices is then computed by the self-attention mechanism. The features learned by the n attention heads are then concatenated to obtain the fused feature vectors learned *MAtt* within each convolutional channel. Afterwards, the fusion feature vectors extracted from each channel using MHA are spliced to obtain the fusion vector *F*, as shown in equation (19).

$$F = Concat\left(MAtt_1, MAtt_2, ..., MAtt_s\right) \tag{19}$$

where $MAtt_i$ denotes the fusion feature vector obtained from the i^{th} convolutional channel, and *M* is the total number of convolutional channels.

Figure 4 Financial fraud recognition model based on DL and textual features



5.3 Financial fraud identification output

The fusion vectors obtained from the two-channel convolutional layer are fed into the fully connected network, and the softmax function is used to generate the distribution of financial fraud behaviour categories, and the calculation process is shown in equation (20).

$$\hat{y}_i = soft\max\left(U_a F + b_a\right) \tag{20}$$

In this paper, the model uses the cross-loss function to calculate the loss, and introduces regularisation coefficients to avoid overfitting of the model due to complex weights, and the loss function is shown as follows.

$$Loss = -\sum_{i \in T_N} \sum_{c \in C_N} y_i^c \log\left(\hat{y}_i^c\right) + \lambda \,\|\, \theta \,\|^2 \tag{21}$$

where $\hat{y}_i$ is the category of financial fraud predicted by the model, y_i is the true category of financial fraud, C_N is the total number of categories, T_N is the total number of financial indicator variables and textual sentiment variables, λ is the regularisation coefficient, and θ is the model parameter.

6 Experimental results and analyses

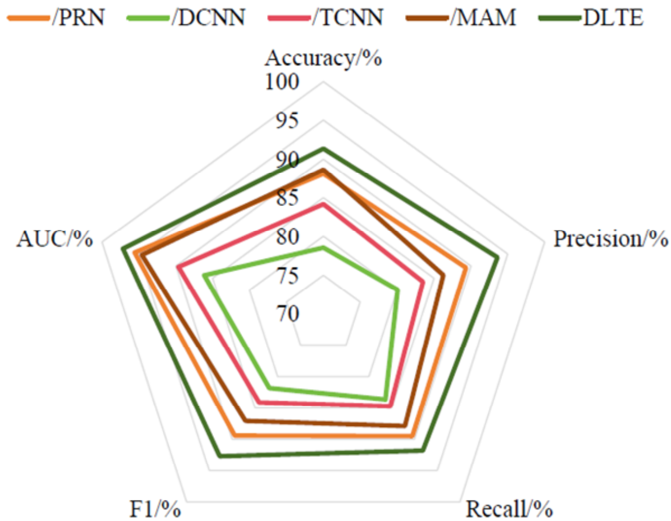
In this paper, the financial annual reports and fraud behaviours of listed companies from 2015 to 2019 are selected as datasets. For the financial annual reports of listed companies, financial report data indicators are collected from the database of GuotaiAn, and then crawled and integrated; for the financial malpractice of listed companies, the information of punishment orders made public by the China Securities Regulatory Commission is crawled using a crawler. After data cleansing and data preprocessing, the dataset available for experimental use contains 13,648 annual reports of listed companies, of which 13094 rows are free of financial malpractice, while 554 rows have financial malpractice. For the fusion of textual sentiment features, the data come from the CMDA database under the China Research Data Service Platform (CNRDS), with a total of 51,291 textual information. In this paper, the ratio of the training set, validation set and test set is 7:2:1, and the parameters of the proposed financial fraud identification model are shown in Table 1.

Table 1 Model parameter setting

Parameter	*Learning rate*	*Optimiser*	*Batch size*	*Epochs*	*Kernel size*
Value	0.0001	Adam	32	20	2

For the financial fraud identification problem, the commonly used evaluation metrics are accuracy, precision, recall, F1, and the AUC value of the area under the ROC curve, and in this paper, we will use these metrics to analyse the identification performance of the model. The suggested model (DLTE) consists of three main components: parallel residual network module, two-channel CNN-based financial indicator or textual feature extraction, and feature enhancement based on multi-head attention mechanism. To verify the impact of these three components on the recognition performance of the proposed model, ablation experiments are conducted, where '/PRN' is to remove the parallel residual network module, '/DCNN' is to remove the financial indicator feature extraction module and keep only the text feature extraction module The '/TCNN' is to remove the text feature extraction module and keep the financial index feature extraction module, and '/MAM' is to remove the feature enhancement module. The ablation results for the different modules are shown in Figure 5.

Figure 5 Results of ablation experiments for different modules of the proposed model (see online version for colours)



The indicators of /DCNN are the lowest, indicating that the feature extraction of financial indicator variables is crucial to the financial fraud recognition model, the indicators of /TCNN are the second lowest, indicating that the extraction of text emotion features is also indispensable, and the indicators of /PRN and /MAM have decreased less compared to the complete DLTE model, indicating that the variable selection through residual network and the multi-attention mechanism feature enhancement can also affect the recommendation effect. The accuracy of DLTE is 91.35%, which is improved by 3.23%, 12.81%, 7.17%, and 2.71% compared to /PRN, /DCNN, /TCNN, and /MAM, respectively. Therefore, each module in the DLTE is essential and the DLTE that incorporates all modules has the best recognition performance.

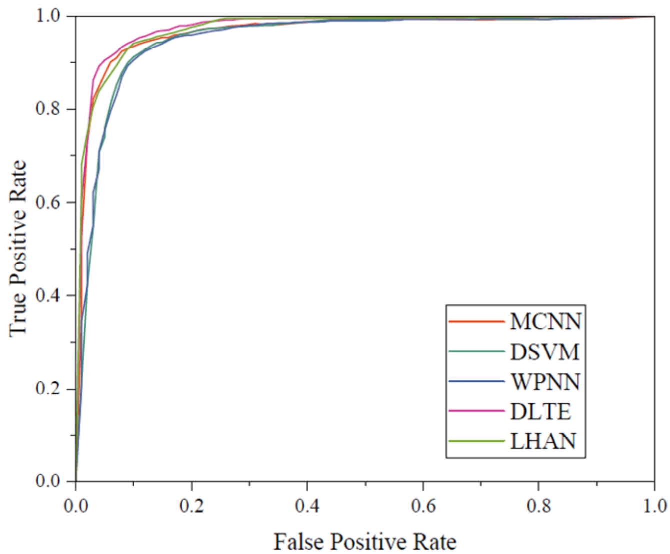
Table 2 Recognition performance comparison results

Model	*Accuracy/%*	*Precision/%*	*Recall/%*	*F1/%*
DSVM	81.56	84.75	85.62	85.18
WPNN	83.14	82.52	83.93	83.22
MCNN	84.52	83.68	84.85	84.26
LHAN	88.79	87.05	89.63	88.32
DLTE	91.35	93.61	91.84	92.72

To further verify the validity of the proposed model DLTE, this paper compares with the models proposed in existing studies. The comparison models are DSVM, WPNN, MCNN and LHAN. The experimental outcome is indicated in Table 2.

From the experimental outcome, it can be seen that the accuracy, precision, recall and F1 of DLTE are better than the other four models. The recognition accuracy of WPNN and MCNN is not much different, and WPNN only considers the financial indicator features as the input of the PNN without considering the emotional features of the text. MCNN extracted the text sentiment features and did not consider the financial indicator features. The recognition accuracy of DSVM is better than WPNN and MCNN, but the F1 value is better than WPNN and MCNN though it is lower than WPNN and MCNN. This is because DSVM considers both financial indicator features and textual sentiment features. The recognition performance of LHAN is second highest, not only the textual features of MD&A are extracted, but also the importance features are highlighted by HAN, but the financial indicator features are not sufficiently taken into account, so the recognition performance of LHAN is worse than that of DLTE.

Figure 6 Comparison of ROC curves for different models (see online version for colours)



Comparison of ROC curves for different models is shown in Figure 6, the larger the AUC value, the better the classification efficiency. The AUC values of DSVM, WPNN, MCNN, LHAN, and DLTE are 95.57%, 95.31%, 96.87%, 97.83%, and 98.52%, respectively, and the AUC values of all five models are greater than 0.5, which are above the null ROC curve, indicating that each model is valid. The AUC value of DLTE is closest to 1, which is improved by 2.95%, 3.21%, 1.65%, and 0.69% compared to DSVM, WPNN, MCNN, and LHAN, respectively, which is significantly better than the other models, and once again illustrates that the recognition effect of DLTE is better than the other four models.

7 Conclusions

Intending to the issue that the existing financial fraud recognition methods ignore the text emotional features, resulting in low recognition accuracy, a company financial fraud recognition method based on DL and text features is proposed. Firstly, the Pearson correlation coefficient is utilised to screen the financial indicators and eliminate the redundant indicators. Secondly, the Bert model is utilised for MD&A text word vector embedding, and the word vectors are used as input sequences for BiLSTM to extract textual sentiment variables. Then the parallel residual network is used to select the financial indicator variables and textual sentiment variables in depth, the two variables are put through two-channel CNN for feature extraction, and the MHA is used to pay attention to the key parts of the two variables, to obtain the hidden information at different levels, and splicing to get the fusion variables of the financial indicator variables and the textual sentiment variables. Finally, the final identification result is obtained by batch normalisation and full connectivity.

The research in this paper solves, to a certain extent, the problems of weak learning ability of machine learning algorithms and single neural networks as well as the limitations of feature selection in existing research, and provides ideas for the application of artificial intelligence in economic issues. However, the method has some shortcomings, future work is intended to be carried out in two ways.

1. Segmentation of financial fraud samples. Currently, most studies, including this paper, simply divide the sample into two categories: fraud samples and non-fraud samples. In the future, we can consider further dividing the fraud samples based on the severity of financial fraud, the intensity of corporate penalties, and other information, so as to further deepen the study of fraud samples.

2. The connection between abnormal data and time of fraudulent enterprises is analysed. Some listed companies have persisted in financial fraud for many years or several consecutive years, and in the future, we can consider introducing the time factor into the identification model and try to mine the association between abnormal indicators and time.

References

Aftabi, S.Z., Ahmadi, A. and Farzi, S. (2023) ‘Fraud detection in financial statements using data mining and GAN models’, *Expert Systems with Applications*, Vol. 227, p.120144.

Aghghaleh, S.F. and Mohamed, Z.M. (2014) ‘Fraud risk factors of fraud triangle and the likelihood of fraud occurrence: evidence from Malaysia’, *Information Management and Business Review*, Vol. 6, No. 1, pp.1–7.

Ali, A., Abd Razak, S., Othman, S.H. et al. (2022) ‘Financial fraud detection based on machine learning: a systematic literature review’, *Applied Sciences*, Vol. 12, No. 19, p.9637.

Benchaji, I., Douzi, S., El Ouahidi, B. et al. (2021) ‘Enhanced credit card fraud detection based on attention mechanism and LSTM deep model’, *Journal of Big Data*, Vol. 8, pp.1–21.

Choi, S., Lee, J. and Kwon, O. (2017) ‘Financial fraud detection using text mining analysis against municipal cybercriminality’, *Journal of Intelligence and Information Systems*, Vol. 23, No. 3, pp.119–138.

Du, M. (2021) ‘Corporate governance: five-factor theory-based financial fraud identification’, *Journal of Chinese Governance*, Vol. 6, No. 1, pp.1–19.

Forough, J. and Momtazi, S. (2022) ‘Sequential credit card fraud detection: a joint deep neural network and probabilistic graphical model approach’, *Expert Systems*, Vol. 39, No. 1, p.e12795.

Gong, Y., Li, J., Xu, Z. et al. (2022) ‘Detecting financial fraud using two types of Benford factors: evidence from China’, *Procedia Computer Science*, Vol. 214, pp.656–663.

Hilal, W., Gadsden, S.A. and Yawney, J. (2022) ‘Financial fraud: a review of anomaly detection techniques and recent advances’, *Expert Systems with Applications*, Vol. 193, p.116429.

Li, Z., Liu, F., Yang, W. et al. (2021) ‘A survey of convolutional neural networks: analysis, applications, and prospects’, *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 33, No. 12, pp.6999–7019.

Nakashima, M., Hirose, Y. and Hirai, H. (2022) 'Fraud detection by focusing on readability of MD&A disclosure: evidence from Japan', *Journal of Forensic and Investigative Accounting*, Vol. 14, No. 2, pp.276–298.

Nami, S. and Shajari, M. (2018) 'Cost-sensitive payment card fraud detection based on dynamic random forest and k-nearest neighbors', *Expert Systems with Applications*, Vol. 110, pp.381–392.

Tang, J. and Karim, K.E. (2019) 'Financial fraud detection and big data analytics-implications on auditors' use of fraud brainstorming session', *Managerial Auditing Journal*, Vol. 34, No. 3, pp.324–337.

Wang, G., Ma, J. and Chen, G. (2023) 'Attentive statement fraud detection: distinguishing multimodal financial data with fine-grained attention', *Decision Support Systems*, Vol. 167, p.113913.

Wang, R., Chencho, C., An, S. et al. (2021a) 'Deep residual network framework for structural health monitoring', *Structural Health Monitoring*, Vol. 20, No. 4, pp.1443–1461.

Wang, X., Tang, M., Yang, T. et al. (2021b) 'A novel network with multiple attention mechanisms for aspect-level sentiment analysis', *Knowledge-Based Systems*, Vol. 227, p.107196.

Xiuguo, W. and Shengyong, D. (2022) 'An analysis on financial statement fraud detection for Chinese listed companies using deep learning', *IEEE Access*, Vol. 10, pp.22516–22532.

Xu, G., Meng, Y., Qiu, X. et al. (2019) 'Sentiment analysis of comment texts based on BiLSTM', *IEEE Access*, Vol. 7, pp.51522–51532.

Zhao, Z. and Bai, T. (2022) 'Financial fraud detection and prediction in listed companies using SMOTE and machine learning algorithms', *Entropy*, Vol. 24, No. 8, p.1157.

Zhao, Z., Chen, W., Wu, X. et al. (2017) 'LSTM network: a deep learning approach for short-term traffic forecast', *IET Intelligent Transport Systems*, Vol. 11, No. 2, pp.68–75.