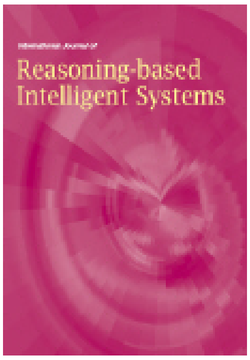


International Journal of Reasoning-based Intelligent Systems

ISSN online: 1755-0564 - ISSN print: 1755-0556
https://www.inderscience.com/ijris

Cross-border e-commerce credit risk early warning model based on deep learning

Yanping Jiang, Peng Du

DOI: 10.1504/IJRIS.2025.10072955

Article History:

Received:	14 June 2025
Last revised:	04 July 2025
Accepted:	14 July 2025
Published online:	27 August 2025

Cross-border e-commerce credit risk early warning model based on deep learning

Yanping Jiang and Peng Du*

Department of Digital Commerce,
Yantai Engineering and Technology College,
Yantai 264000, China
Email: maggie2282021@163.com
Email: rilyxiaojiang@163.com
*Corresponding author

Abstract: Credit risk challenges in cross-border e-commerce are becoming more pronounced, significantly hindering the sector's healthy progression. Existing early warning mechanisms cannot effectively analyse the complex, nonlinear patterns and dynamic changes in cross-border e-commerce transactions, causing low detection accuracy. To this end, this study first introduces an enhanced SMOTE algorithm to address class imbalance in credit hazard data, followed by the selection and standardisation of key hazard indicators. Then one-dimensional CNN is used to deeply mine the features of impact indicators. The particle swarm optimisation (PSO) algorithm is utilised to fine-tune the parameters of the extreme learning machine (ELM), with the optimised ELM subsequently generating the risk warning classifications. The experimental outcome indicates that the forecasting accuracy of the offered model is improved by at least 4.05%, demonstrating significant gains in warning system accuracy.

Keywords: credit risk early warning; deep learning; convolutional neural network; CNN; extreme learning machine; ELM; particle swarm optimisation algorithm; PSO.

Reference to this paper should be made as follows: Jiang, Y. and Du, P. (2025) 'Cross-border e-commerce credit risk early warning model based on deep learning', *Int. J. Reasoning-based Intelligent Systems*, Vol. 17, No. 10, pp.1–9.

Biographical notes: Yanping Jiang received her Master's degree at Nanchang Hangkong University in 2016. She is currently a Lecturer at Yantai Engineering and Technology College. Her research interests include cross-border e-commerce and international trade.

Peng Du received her Master's degree at Harbin Engineering University in 2024. She is currently a Lecturer at Yantai Engineering and Technology College. Her research interests include cross-border e-commerce English, and information assisted instruction.

1 Introduction

In the context of the era of the wave of globalisation and the deep integration of internet technology, as a cutting-edge approach to international trade, cross-border e-commerce is revolutionising global trade structures at an extraordinary rate (Hazarika and Mousavi, 2021). It breaks the restrictions of traditional trade in time and space, enabling enterprises to directly face global consumers, greatly expanding market boundaries, reducing transaction costs and enhancing trade efficiency. Nevertheless, the quick growth of cross-border e-commerce has also brought a series of new challenges, among which the credit risk problem is particularly prominent (Wang et al., 2022). The involvement of trading parties from diverse countries and regions in cross-border e-commerce creates a particularly intricate and changeable transaction environment, exacerbating problems of information asymmetry (Ma et al., 2022). It is challenging for trading partners to verify each other's credit standing before entering into a transaction, and there is a lack of effective supervision and constraint mechanisms in the transaction process, which leads to the frequent occurrence of fraud, default and other credit risk events (Zhou et al., 2022). Beyond imposing considerable economic costs on transacting parties, these credit risks jeopardise the healthy evolution of cross-border e-commerce and weaken market engagement (Zhou, 2023). Therefore, implementing an accurate credit hazard evaluation framework for cross-border e-commerce transactions is practically essential.

Credit risk prediction models constitute the specialised domain within business credit risk evaluation. Previous studies typically employed conventional analytical approaches to construct credit risk early warning models. Lu et al. (2013) constructed an ordered semiparametric probit credit assessment model and demonstrated the significant effectiveness of the model in corporate credit risk early warning through empirical studies. Tsai (2013) proposed a univariate early warning analytical model by using the paired sampling method to extract 79 companies that incurred credit risk and 79 paired healthy companies. Wei and Wang (2024) applied a multivariate linear discriminant model to study the corporate credit crisis, selected 5 indicators as final judgment variables from 22 indicators initially, established a Z-score model, and generated the Z-value as the probability of the occurrence of corporate credit risk. Lin (2022) used a multivariate discriminant analysis model for Turkish manufacturing companies and extracted three indicators that are effective for credit risk early warning capability. Zhou et al. (2021) put forward a predictive model for credit risk assessment that synthesises principal component analysis and logistic regression approaches, and predicted the credit status of 131 listed real estate companies, and the model demonstrated comprehensive prediction accuracy above 80%.

Recently, as the nonlinear theory and machine learning algorithms continuously integrating, artificial neural network (ANN) has emerged as a representative of credit risk early warning methods for listed companies. The predominant models in current use are machine learning approaches, particularly ANNs and support vector machines (SVMs) (Feng et al., 2021). Deng et al. (2021) developed a BP neural network-based early warning model for cross-border e-commerce credit hazard assessment. When compared to the multivariate discriminative model, the newly – proposed model exhibited a higher prediction accuracy, specifically reaching 81.9%. Wang and Gao (2023) applied Logistic model and SVM model to warn the credit crisis of listed companies of cross-border e-commerce conglomerates, respectively. SVM significantly outperforms logistic regression in corporate credit risk early warning through its maximum margin principle and robust design, especially when dealing with nonlinear relationships, high-dimensional data, and class imbalance. However, its weaknesses in interpretability and long training time need to be compensated for through hybrid models or post-processing techniques. In practical applications, the appropriate algorithm should be selected based on data scale, feature complexity, and business needs. Comparative evaluation showed that the SVM classifier exceeded Logistic Regression's predictive capability in corporate credit risk alert scenarios. Qasem and Nemer (2019) performed dimensionality reduction of credit impact indicators through feature selection, and then constructed a credit risk forecasting approach in light of ELM with a prediction accuracy of 85.7%.

Corporate credit crises arise from multifaceted causes and evolve through continuous, non-discrete processes, yet generate substantial quantifiable data. Deep learning focuses on processing large amounts of data through deep networks with multiple hidden layers. Features are learned through training, rather than being determined artificially. The use of deep learning network can determine the main characteristics of the research object and make judgment through neural network self-learning, which can effectively improve the prediction effect and efficiency and reduce the work difficulty. Therefore, deep learning has certain advantages over traditional neural networks and existing early warning models, and it is expected to obtain a higher accuracy rate. Du and Shu (2023) constructed a deep belief network for predicting credit crises, which can well distinguish normal and credit crisis firms with an overall accuracy of 82.31%. Ouyang et al. (2021) developed an LSTM-based financial distress prediction model for Chinese listed firms, achieving 84.27% classification accuracy. Meng et al. (2024) innovatively applied NLP-based word vectorisation to e-commerce textual data, coupled with CNN's feature extraction capability, to strengthen credit risk prediction reliability. Li et al. (2023) proposed a new credit risk prediction model based on the self-attention mechanism, which can achieve better accuracy as well as AUC value. Xia (2024) developed a hybrid corporate credit risk early warning system combining LSTM and CNN architectures, where the LSTM sub-model captures temporal patterns while the CNN extracts localised features from multi-year risk data, collectively enhancing prediction accuracy.

Cross-border transactions involve multiple countries. It is difficult to obtain comprehensive credit data of the buyer (such as financial information and business conditions), especially in some countries with an imperfect credit information system. Traditional models rely on historical data and are difficult to capture real-time risk changes in cross-border transactions. In addition, most of the existing cross-border e-commerce early warning models adopt shallow architectures with limited feature extraction capabilities, resulting in less than ideal prediction effects. For this reason, this paper proposes a cross-border e-commerce credit hazard early alert model in light of deep learning theory. Firstly, to address the issue of imbalanced distribution of credit risk data, an improved SMOTE oversampling method is proposed, and the K-nearest neighbours (KNN) approach is adopted for noise filtering. The clustering approach is used to categorise the minority samples into several clusters, and the density of each minority sample is calculated. The roulette wheel selection method is used and the minority samples with higher density are selected based on probability to generate new minority samples. Subsequently, through indicator screening and standardisation procedures, the critical credit risk metrics for cross-border e-commerce are prepared for analysis. Deeper feature abstractions are successfully captured from cross-border e-commerce transaction records through 1DCNN's hierarchical convolution and pooling architecture. Subsequently, the credit risk features extracted by 1DCNN are used as inputs to PSO-ELM, which has fast training speed and excellent high-dimensional feature processing capability, to finally derive the credit risk level. The experimental outcome indicates that the forecasting accuracy of the offered model is 96.23%, which significantly improves the prediction accuracy.

2 Relevant technologies

2.1 Convolutional neural network

CNNs constitute specialised deep learning architectures that inherently capture spatial hierarchies in grid-like data through localised connectivity patterns (Liu et al., 2016). Compared with multilayer feedforward neural networks, CNN is characterised by two fundamental properties: localised receptive fields and parameter sharing across spatial positions, and spatial subsampling, and has fewer network parameters. The typical CNN network structure is usually composed of one or more connections of convolutional levels and pooling levels, as well as the final fully linked level.

The input layer of the CNN network is used to receive time series input data $X = [\chi_1, x_2, \ldots, x_n]$ and process it. At each convolutional level, filter kernels perform element-wise multiplication and summation with sliding input vector segments, weights, and bias to extract corresponding input features, and then employs activation functions to introduce nonlinear transformations between levels. The essence of the convolutional layer is the process of extracting features from the original data. It generates multiple feature maps by performing multiple parallel convolution operations on the input, and then transforms each feature through a nonlinear activation function. The convolution operation is as follows.

$$C_1 = f\left(X \otimes w_1 + b_1\right) = \mathrm{Re}\,LU\left(X \otimes w_1 + b_1\right) \tag{1}$$

where C_1 is the output vector of the convolutional level, w_1 is the weight matrix, b_1 is the bias vector, e is the activation function, ⊗ is the symbol for dot product multiplication.

A Flatten level is used between the convolutional layer and the fully linked level to transform multidimensional characteristic pictures into a 1D vector representation. Batch Normalisation accelerates network training and enhances convergence by stabilising layer inputs. The dropout level serves to avert overfitting and enhances the model's capability for generalisation. The fully linked level is the traditional neural network structure. Its essence is to integrate the characteristics produced by previous convolutional levels and set different parameters. It uses the activation function to assign weights to the feature vector and iteratively updates to obtain an optimal weight parameter matrix, which is adopted to explain characteristics captured by the convolutional part of the model.

2.2 Extreme learning machine

Extreme Learning Machine (ELM) is a very popular machine learning method currently. In essence, ELM is a easy individual implicit level feedforward neural network (Wang et al., 2016). Its characteristics are randomly initialising the input weights and biases of the hidden layer neurons, permitting an analytical solution for the output layer weights via least-squares regression, thereby avoiding the complex computation and parameter adjustment process required by the backpropagation algorithm. ELM boasts the merits of rapid learning speed, a straightforward model structure, and robust generalisation capability. ELM is a fast learning algorithm for single hidden layer feedforward neural networks. Compared with classical machine learning algorithms, ELM randomly generates the connection weights between the input layer and the hidden layer and the threshold of the neurons in the hidden layer, without iterative adjustment. It only needs to calculate the output weights, which greatly shortens the training time and has significant advantages when dealing with large-scale data. In many practical application scenarios, ELMs can achieve generalisation capabilities comparable to or even better than those of traditional algorithms while ensuring training speed. This is because of its unique parameter setting method and output weight calculation approach, which enables it to effectively balance the complexity of the model and its ability to fit data during the learning process, avoiding overfitting and thus performing well on new data.

The principle of ELM is based on random feature mapping and least squares. The fundamental concept involves transforming input features into a higher-dimensional representation space, and then use the least squares method to learn the mapped data. In contrast to traditional neural networks, ELM does not require gradient-based parameter optimisation, but can complete the training with a single forward propagation, so it has high efficiency and scalability. The network randomly generates the input layer weight matrix and the implicit level bias. The random mapping ELM model can be expressed as follows.

$$T = \sum_{i=1}^{l} \beta g\left(\omega_i, b_i, x\right) \tag{2}$$

where x is the input variable, T is the output matrix of the ELM target data, $\omega = [\omega_1, \omega_2, \ldots, \omega_i, \ldots, \omega_l]^T$ is the sigmoid function, $b = [b_1, b_2, \ldots, b_i, \ldots, b_l]^T$ is the input layer weight matrix, is the implicit level bias matrix, $\beta = [\beta_1, \beta_2, \ldots, \beta_i, \ldots, \beta_l]^T$ is the output level weight matrix. The ELM objective function is $\min\|\mathrm{H}\beta - T\|^2$, where $\beta \in R^{L\times m}$, H is the output matrix of the implicit level, T is the objective matrix. The most favourable solution of the equation can be derived by knowledge of linear algebra and matrix theory as follows.

$$\beta^* = H^+ T \tag{3}$$

3 Data processing of unbalanced credit hazard in cross-border e-commerce based on improved SMOTE oversampling

3.1 Noise sample filtering

According to the results of the work of existing research, Cross-border e-commerce companies exhibit a small proportion of corporate credit risk cases and a large number of financially healthy firms, leading to an imbalanced dataset. The problem of unbalanced datasets will directly lead to the problem of incomplete and ill-considered construction of the evaluation index system. As a result, the performance of credit risk forecasting will deteriorate considerably. Therefore, in this study, the SMOTE algorithm is improved (ESMOTE) by using the roulette wheel selection operator. ESMOTE performs data augmentation on the corporate credit risk dataset to solve the data imbalance problem.

Aiming at the traditional SMOTE oversampling method (Chawla et al., 2002) which neglects the intra-class imbalance and does not fully consider the nearest-neighbour characteristics of samples of the same and different classes in the sample space, this paper proposes the ESMOTE oversampling method. Noise filtering was performed using the KNN algorithm prior to the data balancing process. Second, clustering methods are used to categorise the minority samples into a number of clusters and calculate the density of each minority sample. The denser samples contain richer information features compared to the less dense samples. Therefore to determine the quality of each minority class sample in its clustering, a roulette wheel selection method is used and a new minority class sample is generated by selecting the denser minority class sample based on probability.

Noisy samples usually refer to mislabelled or abnormal samples in the dataset, which may be caused by data collection errors, human mislabelling, data entry errors, and so on, and may adversely affect the model's performance. Within data mining research, identifying and processing noisy samples is a very important task to ensure the accuracy and reliability of model training. In view of the above issues, this paper uses KNN noise filtering method in light of Euclidean distance to solve the problem of noise in data samples. In machine learning and data pre-processing, the core purpose of using the KNN algorithm for noise processing is to identify and correct abnormal samples through local neighbourhood analysis, thereby optimising the rationality of sample distribution and enhancing the generalisation ability of subsequent models. The key problems it addresses include noise data interference, class boundary blurring, and uneven sample density, etc. The algorithm first uses the KNN approach to compute the distance between all sample points and its k nearest neighbours. It then measures the consistency of the labels based on these distances. A sample is considered to be noise-free if most of its k nearest neighbours have the same labels. Otherwise, they are labelled as noise samples. In this way, this method successfully identifies and removes noisy labels from the training data, consequently boosting the model's accuracy and stability.

3.2 *Sample selection based on roulette wheel selection operator*

In the traditional SMOTE algorithm, not only the imbalance between different categories of samples affects the model's ultimate classification accuracy, but also the imbalance within a certain category of samples affects the classification outcome. Therefore, in this article, the optimisation is improved: the problem is solved using clustering. K-means++ (Vardakas and Likas, 2024) was chosen for the clustering task due to its advantages in determining the initial centres. Noise filtering is first performed using the KNN algorithm, followed by clustering operations on the denoised minority class samples by the K-means++ algorithm.

After clustering, the minority class samples form k small clusters. Within each small cluster, samples with higher density compared to low-density samples often contain more valuable information. To obtain this information, Euclidean distances are employed to derive density estimates for minority class examples by locating their proximal neighbours in clustered regions. First, all minority samples begin with a local density value of 0 and represented by $D(i)$, that is $D(i) = 0$. After the clustering operation, k clusters are obtained. If the nearest neighbour of sample x_j is x_i, the density of sample is x_l, increased by 1, that is $D(i) = D(i) + 1$. This process is repeated cyclically, and finally, the density of all minority class data samples is calculated, which is relative to each cluster internally. According to the density value, the roulette selection operator is used to select appropriate samples from each cluster for the next sampling.

The roulette selection algorithm is a method for selecting objects with different probability distributions. The basic idea of this operator is to normalise the samples according to their local density and convert them into a probability distribution. Samples are then selected in a roulette-like manner according to the probability distribution. Samples with higher density are assigned a higher probability and are therefore more likely to be selected. Using the partial density of the minority category samples as the fitness, the selection probability is calculated according to equation (4).

$$P\left(x_{mini}\right) = \frac{D\left(x_{mini}\right)}{\sum_{j=1}^{n_{min}} D\left(x_{minj}\right)} \tag{4}$$

where x_{mij} and x_{mini} are the j^{th} and i^{th} minority class samples in a cluster, respectively. Because high-density samples contain more informative data compared to low-density samples, the roulette selection operator is used to select the high-density sample data.

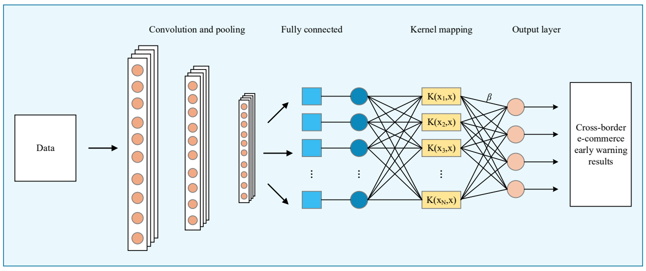
Finally, the ESMOTE algorithm is used to interpolate the samples selected by the roulette selection operator. This process is repeated until the number of minority class samples reaches a balanced state with the majority class samples.

4 Cross-border e-commerce credit hazard early warning model based on deep learning theory

4.1 Selection of influence indicators for cross-border e-commerce credit hazard

For existing early warning models adopt shallow architecture with limited feature extraction capability, which leads to the predictive effect of early warning method for credit hazard in cross-border e-commerce is not satisfactory enough. For this reason, this study proposes a deep learning-based early warning model for credit hazard evaluation, as implied in Figure 1. First, key credit hazard metrics for cross-border e-commerce are identified and normalised. Subsequently, one-dimensional CNN (1DCNN) is used as a feature extractor to deeply mine the features of credit risk impact indicators. At last, the PSO approach is brought in to optimise the parameters of the ELM, creating the optimised extreme learning machine (OELM). Then, the credit hazard warning categories are generated as outputs through the optimised ELM classifier.

Figure 1 The suggested cross-border e-commerce credit hazard early warning model (see online version for colours)



1DCNN demonstrates a keen insight into the deeper features of the data, and is able to dig deeper into the intrinsic data characteristics of credit data in order to reduce redundancy. This is complemented by the instantaneous training feature of OELM, which enables it to perform well in large-scale data classification tasks. The uniqueness of the 1DCNN- OELM model is that the 1DCNN deep feature extraction capability and the efficient training mechanism of the OELM are delicately combined, resulting in a vital improvement in the precision and generalisation performance of the model. Figure 1 illustrates the workflow adopted in this article.

During the construction of credit hazard assessment parameters for cross-border online retail, attention should be paid to the logical relationship between the indicators. The choice of indicators should not only manifest the key characteristics of the financial and operating conditions of listed enterprises, but also reflect the internal relationship between the selected indicators. Each dimension indicator is composed of multiple sub-indicators, which should be independent of each other and interrelated to form dimension indicators. The construction of a statistical indicator system should have certain selection significance and form a comprehensive evaluation system. This paper draws on the credit risk early warning impact indicators selected in existing studies. From the dimensions of transaction subject credit, transaction behaviour, goods and logistics, international environment and market, and platform evaluation and public opinion, n first-level indicators and m second-level indicators are selected to construct a credit risk pre-warning model for cross-border online transactions.

For the goal of avoiding the interference of human factors, this study adopts XGBoost model for the initial screening of risk indicator factors. By scoring the importance of the characteristics of each influencing factor, the preliminary factors affecting the credit risk of enterprises were identified, and a reasonable index system was constructed to form the empirical foundation for later model formulation and prediction. The nature of each evaluation indicator is different and of different orders of magnitude. Failure to process the raw data directly affects the results of the data analysis, neglecting the role of other, equally important but smaller indicators. Therefore, this article employs the min-max standardisation approach to standardise the credit hazard early warning indicator data. A linear transformation is performed on the original indicators, mapping the values to the range [0, 1], as shown in the following equation, where x^* is the standardised value, and x is the initial value.

$$x^* = \frac{x - min}{max - min} \tag{5}$$

The data flow of the 1DCNN model constructed in this paper passes through a series of levels, including two convolutional levels, one pooling level, followed by three rounds of repeated convolution-pooling process, and finally two fully connected layers to obtain the model output.

4.2 ID CNN-based feature extraction for impact indicators

1D CNN is a variant of CNN, and its advantage lies in being able to extract core features from one-dimensional data. In 1D CNN, convolution and pooling operations are applied to one-dimensional data, rendering it especially appropriate for time-series data analysis. Similarly, CNN mainly uses two-dimensional convolution and pooling operations, and is often used for feature extraction of two-dimensional data such as images. The commonality between the two is that they both use convolution and pooling operations, which can effectively extract local features and integrate them in the fully linked level to generate the ultimate output. Given the univariate nature of the extracted cross-border e-commerce credit hazard data, this study employs a one-dimensional 1D CNN to learn deep feature representations of the risk indicators.

The proposed 1D CNN architecture processes data through multiple sequential layers, including two convolutional levels, one pooling level, followed by three rounds of repeated convolution-pooling processes, and finally, the model output is obtained through two fully linked levels. The overall structure includes 6 convolutional levels, 3 pooling levels, and 2 fully inked levels This design aims to effectively capture the abstract features in credit risk monitoring parameters for global e-commerce platforms through multi-level convolution and pooling operations.

In the convolution operation of 1D CNN, the input data performs a point-by-point dot product with the convolution kernel, which means that the convolution kernel is applied to different positions of the input data, and the results of each position are summed to generate the feature map. This operation process allows the network to effectively identify regional dependencies and discriminative characteristics within the data. The one-dimensional convolution is as follows.

$$S(t) = (X * W)(t) = \sum_{a=1}^{m} X(a \cdot stride) \cdot W\left(\frac{t-a}{stride} + b\right) + b \quad (6)$$

where X is the input sequence, W is the convolution kernel, t is the current position of the convolution function, a is the position index of the convolution kernel, which is controlled by the stride to determine the sliding interval of the convolution kernel on the input sequence, m is the size of the convolution kernel, indicating the window size of the convolution kernel on the input sequence, b is the bias term, $S(t)$ represents the result after convolution.

In order to reduce the data size and the number of parameters, a pooling layer is usually introduced after the convolution layer. The pooling level effectively reduces the dimensionality and the number of parameters of the feature data through compression. This strategy helps to maintain the key information in credit while achieving effective dimensionality reduction of credit data, thereby providing a more efficient feature representation in the subsequent network structure. There are two common types of pooling levels, and the equation for the maximum pooling level is as follows.

$$z_{max} = \max 1_{k=0}^{k_{size}-1} C_{(i*stride)+k} \quad (7)$$

where z_{max} is the output value after pooling, c is the input feature of the pooling level, k_{size} is the size of the pooling window, stride is the sliding step length of the window, and i is the position of the pooling window on the feature.

After the convolution level and pooling level the extracted risk features are integrated through the fully linked level. Given that the cross-border e-commerce credit risks addressed are categorised into five types: high risk, medium risk, low risk, dishonest, and severely dishonest. Five output neurons are implemented in the classification layer to correspond with 5 risk categories

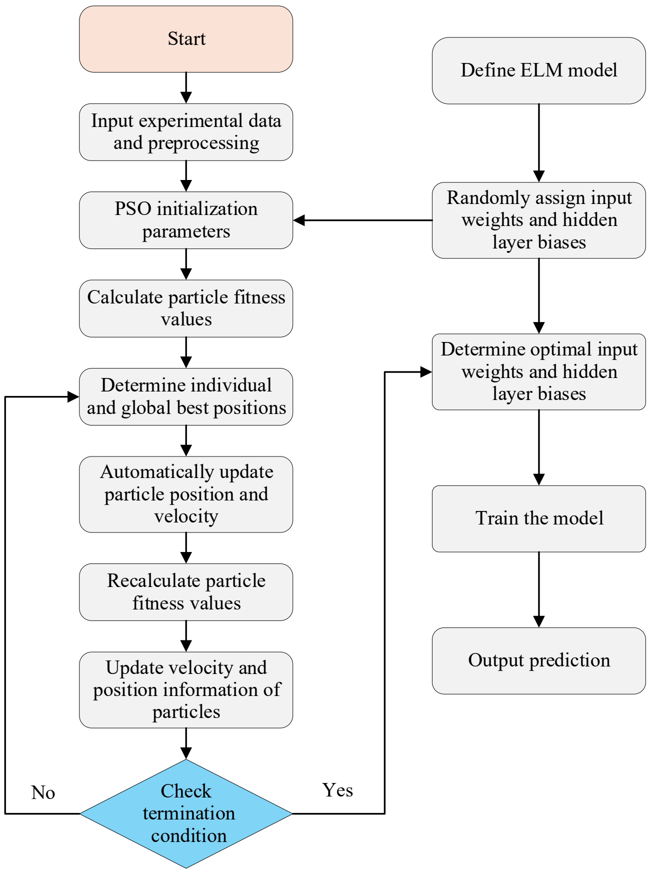
4.3 *Cross-border e-commerce credit hazard prediction based on improved extreme learning machine*

After 1DCNN characteristic extraction, the extracted features are given to the improved ELM for credit risk classification. ELM, as a training technique for single-implicit level feed-forward neural networks, can directly obtain the weight values of the outgoing levels of the network through a one-step parsing method, thus realising a high learning efficiency. When the training instance is $D = \{(x_n, y_n), n = 1, 2..., N\}$, its regression function is as follows.

$$\hat{y} = f(x) = h(x)\beta = H\beta \quad (8)$$

Let x be the input vector; the network's predicted output is $\hat{y} = f(x)$, the random features of the implicit level are mapped to $h(x) = H$, and the connection weights among the implicit level and the output level are denoted as β, and its solution is as bellow according to the theory of generalised inverse matrices.

Figure 2 The entire flow of the PSO-ELM (see online version for colours)



$$\beta = H^T \left(\frac{I}{C} + HH^T\right)^{-1} y \quad (9)$$

where I is the identity matrix, C is the penalty coefficient, y is the output objective vector.

Because the input weight w_j and the implicit bias b_j will directly affect the prediction accuracy and stability of ELM, this paper considers combining PSO and ELM to realise the optimisation of w_j and b_j, in order to achieve better stability and accuracy. The specific steps of PSO algorithm optimising ELM are indicated in Figure 2.

1 Initialise the maximum number of iterations Max_iteration, inertia weight, population size M, select appropriate acceleration constants c_1 and c_2, and

initialise the random location of the particles. In this paper, it is the input w_j and b_j as the velocity and location of the particles.

2 Calculate the fitness of each particle to ascertain the single extreme value as well as the group extreme value. In the optimisation algorithm, it is necessary to calculate the w_j and b_j of a single particle in the population M. ELM can calculate the output weight matrix. Here, particle fitness evaluation employs training sample root mean square error to determine individual particle performance, with the optimisation objective being the discovery of the population-wide optimal extreme value. Let the fitness of the particle swarm algorithm be $F = fitness$. In the same iteration, compare F and the single extreme value first. If $F > P_i$ (single extreme value), then replace P_i with F, otherwise keep the status quo. Then compare F with P_g (global extreme value) and if $F > P_i$, replace P_i with F, otherwise maintain the current status. The fitness function is shown in equation (10), where n is the amount of measurements, and d_i quantifies the variation of measurements relative to the central tendency.

$$fitness = RMSE = \sqrt{\frac{\sum d_i}{n-1}} \tag{10}$$

3 The algorithm iteratively updates each particle's position and velocity vectors to converge toward the extremum, while fitness evaluations guide these updates through the objective function. This iterative process persists until one of the following conditions is met: the count of runs reaches or exceeds the predetermined maximum number of iterations; the elapsed running time equals or surpasses the specified longest running time; or the fitness value becomes less than or equal to the predefined limit value. Upon the occurrence of any of these conditions, the process is terminated. Then, the optimal velocity and position are output, which are the optimal w_j and b_j in ELM.

4 After determining the optimal w_j and b_j, they are input into the ELM model for training and prediction. The optimal result is then adopted to compute the output weight matrix, and finally a relatively stable and highly accurate result is obtained.

5 Experimental results and analyses

The dataset used in this paper is from the Airbnb website and contains data from 6,940 cross-border e-commerce business credit samples in five classes: high-risk, medium-risk, low-risk, and default and serious default. The imbalance ratio of low-risk samples and default samples is about 6: 6: 1. In this paper, ESMOTE method is applied to synthesise new samples and the final samples reach equilibrium state. The pytorch framework is used for neural network construction, and python programming language and Jupyter notebook compiler are used for modelling and research. In the experiment the pool size is 2, the bacth size is 128, the epoch is 100, the loss function is Cross-entropy loss, the fully linked level activation operation is ReLU, and the output level activation operation is Softmax.

Figure 3 illustrates the training loss of five data balancing methods, namely SMOTE algorithm, borderline-SMOTE algorithm (Al Majzoub et al., 2020), ADASYN algorithm (Gu et al., 2020), randomised oversampling algorithm (Hayaty et al., 2020), and ESMOTE algorithm. As the number of iterations in each round increases, the loss value gradually decreases until it levels off. The ESMOTE algorithm has the largest slope and the lowest loss value in predicting the change of model training loss. It indicates that this method converges at the fastest speed and is more efficient in training. ESMOTE filters noise samples through KNN and improves the problem of intra-class sample imbalance. Compared with the oversampling method in the comparative experiments of this chapter, it has obvious advantages in dealing with the problem of data balance. The experimental outcome implies that the ESMOTE oversampling approach can synthesise more effective minority class samples and improve the predictive performance of the model.

Figure 3 Training loss scenarios for data balancing methods (see online version for colours)

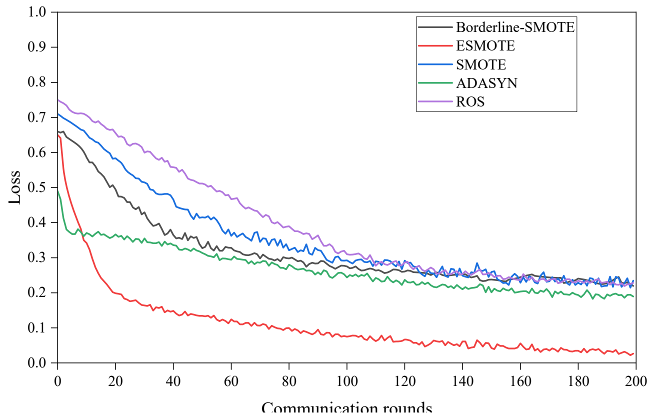
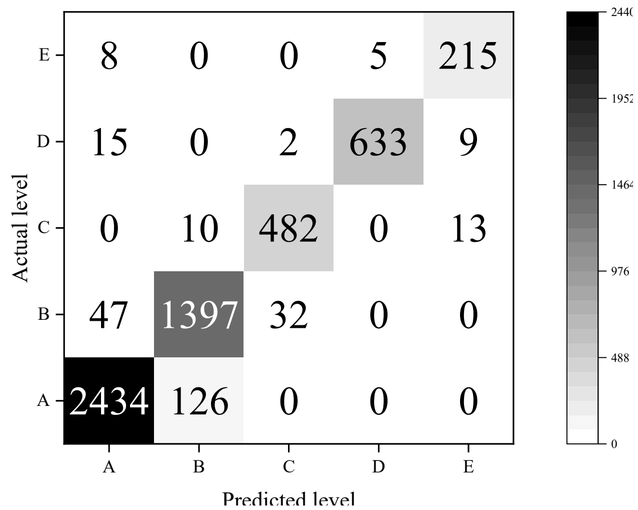


Figure 4 CNN-OELM confusion matrix for five credit risk categories

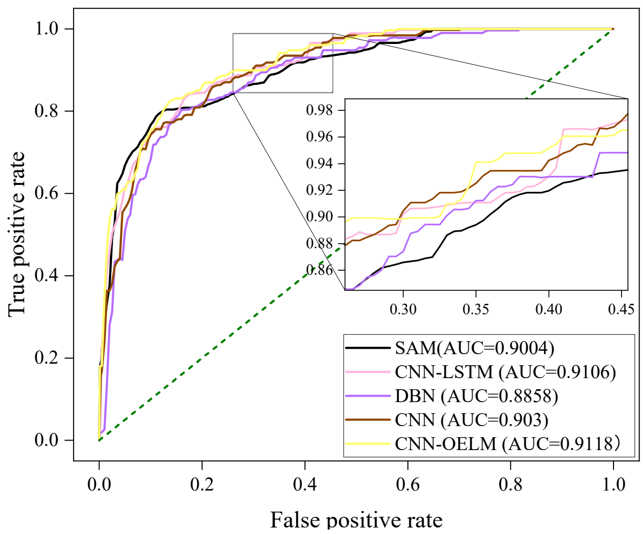


The confusion matrix of the proposed early warning model CNN-OELM for five credit risk categories is shown in Figure 4. The five categories of high risk, medium risk, low risk and default and severe default are denoted as A, B, C, D and E, respectively. The average prediction accuracy of the CNN-OELM model is 95.05%, the recognition rate of grade A enterprises is 95.08%, the recognition rate of grade B enterprises is 94.65%, the recognition rate of grade C enterprises is 95.45%, the recognition rate of grade D enterprises is 96.05%, and the recognition rate of grade E enterprises is 94.30%. The above outcome implies that the model suggested in this article is feasible and effective.

Table 1 Comparison of credit risk early warning accuracy

Model	*DBN*	*CNN*	*SAM*	*CNN-LSTM*	*CNN-OELM*
Accuracy/%	82.31	85.07	87.59	92.18	96.23
Precision/%	81.53	85.36	86.33	89.31	94.37
Recall/%	83.97	84.91	88.47	93.54	97.29
F1/%	82.73	85.13	87.39	91.38	95.81

Figure 5 ROC curves for each model on the test set (see online version for colours)



For the goal of fully validating the prediction performance of the CNN-OELM model, this paper uses the quantitative metrics accuracy, precision, recall, F1, and ROC curves to compare the prediction accuracies of different models, and the benchmark models are selected as DBN (Du and Shu, 2023), CNN (Meng et al., 2024), SAM (Li, et al., 2023), and CNN-LSTM (Xia, 2024). Table 1 shows the outcome of each accuracy for each model, and Figure 5 shows the ROC curves for each model test set. The prediction accuracy of CNN-OELM is 96.23%, which is improved by 13.92%, 11.16%, 8.64%, and 4.05% compared to DBN, CNN, SAM, and CNN-LSTM, respectively. Comparing the combined average F1 of Precision and Recall again, the F1 of CNN-OELM is as high as 97.29%, which is an improvement of at least 3.75% compared to the benchmark model. It can be found that the feature extraction of cross-border e-commerce enterprises' predictive indicators through ID CNN has been utilised to predict the early warning results using PSO-optimised ELM, which can significantly improve the accuracy, precision, recall, and F1 of the early warning. From the ROC curve in Figure 5, it can be seen that the AUC value of DBN is the lowest among all the models, which is 0.8858. According to the criterion of AUC evaluation of the model performance (> 0.85 is excellent), each model obtains a better performance. The AUC value of CNN-OELM reaches 0.9118, which is almost in the leading position in all indicators, and can attain a more accurate early – warning system for credit hazards in the cross-border e-commerce domain.

6 Conclusions

Aiming at the current cross-border e-commerce credit hazard early warning models, most of them adopt shallow architecture with limited feature extraction capability, resulting in low prediction accuracy. This study develops a deep learning-based early alert system for cross-border e-commerce credit hazard, leveraging neural networks' superior feature abstraction and pattern detection capacities to enhance prediction accuracy. Firstly, for the problem of unbalanced data of cross-border e-commerce credit hazard, the improved SMOTE algorithm is proposed to balance the data. The clustering method is used to categorise the minority samples into several clusters and calculate the density of each minority sample. A roulette wheel selection method is used and a new minority class sample is generated by selecting the minority class sample with higher density based on probability. Thereafter, the credit hazard-related impact metrics are selected and undergo standardisation. The PSO algorithm is adopted to enhance the weights and biases of ELM, and the credit hazard warning categories are output through the PSO-ELM classifier. The experimental outcome indicates that the proposed model has an AUC of 0.9118, which is better than the comparison model and has a high prediction accuracy. Although the proposed model has made significant progress, it still has shortcomings and involves multiple hyper-parameter adjustments when using the 1D CNN-OELM model for credit hazard prediction. The hyperparameter's set in this paper are all based on previous studies. However, to optimise these hyperparameters in a reasonable way, it is essential to conduct a large number of experiments and repeatedly test different parameter combinations to further enhance the accuracy of the early warning model.

Declarations

All authors declare that they have no conflicts of interest.

References

Al Majzoub, H., Elgedawy, I., Akaydın, Ö. and Köse Ulukök, M. (2020) 'HCAB-SMOTE: a hybrid clustered affinitive borderline SMOTE approach for imbalanced data binary classification', *Arabian Journal for Science and Engineering*, Vol. 45, No. 4, pp.3205–3222.

Chawla, N.V., Bowyer, K.W., Hall, L.O. and Kegelmeyer, W.P. (2002) 'SMOTE: synthetic minority over-sampling technique', *Journal of Artificial Intelligence Research*, Vol. 16, pp.321–357.

Deng, Z., Yan, M. and Xiao, X. (2021) 'An early risk warning of cross-border E-commerce using BP neural network', *Mobile Information Systems*, Vol. 20, No. 1, pp.55–64.

Du, P. and Shu, H. (2023) 'Design and implementation of china financial risk monitoring and early warning system based on deep learning', *IEEE Access*, Vol. 11, pp.78052–78058.

Feng, Q., Chen, H. and Jiang, R. (2021) 'Analysis of early warning of corporate financial risk via deep learning artificial neural network', *Microprocessors and Microsystems*, Vol. 87, pp.10–23.

Gu, X., Angelov, P.P. and Soares, E.A. (2020) 'A self-adaptive synthetic over-sampling technique for imbalanced classification', *International Journal of Intelligent Systems*, Vol. 35, No. 6, pp.923–943.

Hayaty, M., Muthmainah, S. and Ghufran, S.M. (2020) 'Random and synthetic over-sampling approach to resolve data imbalance in classification', *International Journal of Artificial Intelligence Research*, Vol. 4, No. 2, pp. 86–94.

Hazarika, B.B. and Mousavi, R. (2021) 'Review of cross-border e-commerce and directions for future research', *Journal of Global Information Management* (*JGIM*), Vol. 30, No. 2, pp.1–18.

Li, M., Zhang, Z., Lu, M., Jia, X., Liu, R., Zhou, X. and Zhang, Y. (2023) 'Internet financial credit risk assessment with sliding window and attention mechanism LSTM model', *Tehnički Vjesnik*, Vol. 30, No. 1, pp.1–7.

Lin, M. (2022) 'Innovative risk early warning model under data mining approach in risk assessment of internet credit finance', *Computational Economics*, Vol. 59, No. 4, pp.1443–1464.

Liu, M., Shi, J., Li, Z., Li, C., Zhu, J. and Liu, S. (2016) 'Towards better analysis of deep convolutional neural networks', *IEEE Transactions on Visualization and Computer Graphics*, Vol. 23, No. 1, pp.91–100.

Lu, Y-C., Shen, C-H. and Wei, Y-C. (2013) 'Revisiting early warning signals of corporate credit default using linguistic analysis', *Pacific-Basin Finance Journal*, Vol. 24, pp.1–21.

Ma, S., Chai, Y. and Jia, F. (2022) 'Mitigating transaction risk for cross-border e-commerce firms: a multiagent-based simulation study', *International Business Review*, Vol. 31, No. 4, p.10–19.

Meng, B., Sun, J. and Shi, B. (2024) 'A novel URP-CNN model for bond credit risk evaluation of Chinese listed companies', *Expert Systems with Applications*, Vol. 255, pp.124–137.

Ouyang, Z-S., Yang, X-T. and Lai, Y. (2021) 'Systemic financial risk early warning of financial market in China using Attention-LSTM model', *The North American Journal of Economics and Finance*, Vol. 56, pp.13–27.

Qasem, M.H. and Nemer, L. (2019) 'Extreme learning machine for credit risk analysis', *Journal of Intelligent Systems*, Vol. 29, No. 1, pp.640–652.

Tsai, B-H. (2013) 'An early warning system of financial distress using multinomial logit models and a bootstrapping approach', *Emerging Markets Finance and Trade*, Vol. 49, No. 2, pp.43–69.

Vardakas, G. and Likas, A. (2024) 'Global k-means++: an effective relaxation of the global k-means clustering algorithm', *Applied Intelligence*, Vol. 54, No. 19, pp.8876–8888.

Wang, F., Wei, M. and Teo, B.S-X. (2022) 'Research on credit risks and countermeasures of cross-border E-commerce in China', *Frontiers in Business, Economics and Management*, Vol. 6, No. 3, pp.180–183.

Wang, G-G., Lu, M., Dong, Y-Q. and Zhao, X-J. (2016) 'Self-adaptive extreme learning machine', *Neural Computing and Applications*, Vol. 27, pp.291–303.

Wang, L. and Gao, X. (2023) 'Construction of cross-border e-commerce financial risk analysis system based on support vector machine', *Journal of Industrial Integration and Management*, Vol. 8, No. 1, pp.25–38.

Wei, H. and Wang, X. (2024) 'Financial risk management early-warning model for Chinese enterprises', *Journal of Risk and Financial Management*, Vol. 17, No. 7, pp.25–37.

Xia, Z. (2024) 'Early warning of credit risk of internet financial enterprises based on CNN-LSTM Model', *Procedia Computer Science*, Vol. 243, pp.506–513.

Zhou, K. (2023) 'Financial model construction of a cross-border e-commerce platform based on machine learning', *Neural Computing and Applications*, Vol. 35, No. 36, pp.25189–25199.

Zhou, L., Wang, J., Li, F., Xu, Y., Zhao, J. and Su, J. (2022) 'Risk aversion of B2C cross-border e-commerce supply chain', *Sustainability*, Vol. 14, No. 13, pp.80–88.

Zhou, Q., Wang, L., Juan, L., Zhou, S. and Li, L. (2021) 'The study on credit risk warning of regional listed companies in China based on logistic model', *Discrete Dynamics in Nature and Society*, Vol. 14, No. 1, pp.66–72.