Proceedings of the 7th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/2025.AST27335

Credit Card Fraud Detection Based on Machine Learning Algorithms

Lixin Zhang

Industrail and Commercial Bank of China, Beijing, China
zhanglixin@finrisklab.dev

Abstract. With the popularization of electronic payments and consumer credit, credit cards have become the core tool for global financial transactions, with transaction volumes exceeding $50 trillion by 2024. But at the same time as transaction growth, fraudulent methods have shown intelligent and covert characteristics, with frequent occurrences of theft, identity fraud, and false transactions. According to industry reports, global credit card fraud losses will exceed $35 billion in 2024, and fraud patterns will continue to evolve with the iteration of payment technology. To overcome the bottleneck of existing algorithms, this paper proposes a Transformer model that integrates LSTM and Kernel Extreme Learning Machine (KELM) optimization. In the study, correlation analysis was first conducted on various indicators of the data, and then multiple comparative models were used for testing to compare their effectiveness. The experimental results showed that the LSTM-KELM Transformer performed the best in all evaluation metrics, with accuracy, recall, precision, and F1 values reaching 0.999, and AUC reaching 1, significantly better than other models. Among other models, Random Forest, Adaboost, CatBoost, and BP neural network have similar performance, with Accuracy, Recall, and Precision mostly around 0.992, slightly lower F1 values, and AUC between 0.998-0.999, belonging to the suboptimal level; The various indicators of logistic regression are slightly weaker than the above model, about 0.99; The performance of Decision Tree and Gradient Boosting Tree (GBDT) is relatively poor, especially with GBDT's Accuracy, Recall and other indicators only reaching 0.985, and Decision Tree's AUC also only reaching 0.985. Overall, the LSTM-KELM Transformer outperforms other models in terms of classification accuracy, recognition ability for positive and negative samples, comprehensive performance, and ability to distinguish positive and negative classes, demonstrating stronger classification advantages and contributing to the correct classification of fraud detection. This has important practical significance for improving the efficiency of credit card fraud detection and reducing financial losses.

Keywords: Fraud detection, LSTM, Nuclear Extreme Learning Machine, Transformer

1. Introduction

With the popularization of electronic payments and consumer credit, credit cards have become the core tool for global financial transactions, and the global credit card transaction scale will exceed $50 trillion by 2024. But with the growth of transactions, fraudulent methods have also shown

DOI: 10.54254/2755-2721/2025.AST27335

intelligent and covert characteristics, with frequent occurrences of theft, identity fraud, and false transactions. According to industry reports, global credit card fraud losses will exceed $35 billion in 2024, and fraud patterns will continue to evolve with the iteration of payment technology. The current fraud detection faces two core challenges: one is extreme data imbalance, with fraudulent transactions accounting for less than 0.1%, and traditional methods are prone to missed detections due to sample bias; Secondly, transactions have strong temporal characteristics, and the abnormal features of a single transaction need to be identified by combining historical transaction sequences. The requirement for detection delay in real-time payment scenarios further increases the technical difficulty [3]. Traditional detection systems based on manual rules, such as setting fixed transaction amount thresholds and restricting cross regional transactions, are difficult to adapt to dynamically changing fraud patterns, gradually exposing defects such as high missed detection rates and poor flexibility. There is an urgent need for more intelligent detection technologies to break through bottlenecks.

Machine learning algorithms, with their data-driven feature learning and adaptive capabilities, have become the core technology for solving the problem of credit card fraud detection. In the field of traditional machine learning, logistic regression is often used for preliminary screening of trading risks due to its strong interpretability; Integrated algorithms such as Random Forest and XGBoost can handle nonlinear features, capture abnormal associations in transactions through feature importance analysis, and alleviate data imbalance problems by combining sampling techniques such as SMOTE and ADASYN. With the development of deep learning, CNN can extract local key features from transaction data, while LSTM can model the long short-term dependencies of transaction sequences, further improving detection accuracy in temporal scenarios [5]. However, existing algorithms still have limitations: traditional machine learning relies on artificial feature engineering, making it difficult to mine implicit temporal correlations in transaction data; LSTM is prone to gradient vanishing when processing long sequence transactions, and its ability to capture global transaction context is insufficient; CNN lacks the ability to model dynamic changes in transaction sequences and cannot fully meet the detection requirements in complex scenarios.

To overcome the bottleneck of existing algorithms, this paper proposes a Transformer model that integrates LSTM and Kernel Extreme Learning Machine (KELM) optimization. The multi head attention mechanism of Transformer has strong global feature association capture ability, which can effectively explore the implicit connections between multi-dimensional transaction data. However, there are two major problems when used alone: first, it is sensitive to noisy data and lacks generalization in small sample fraud scenarios; Secondly, the fully connected classification layer has low training efficiency and is difficult to adapt to real-time detection requirements. To this end, the model introduces LSTM to preprocess the input transaction time series data, utilizing its strong time series modeling ability to extract local time series features, reduce the computational complexity of Transformer, and enhance time series correlation capture; At the same time, KELM is used to replace the fully connected classification layer of traditional Transformers. KELM, with its kernel function mapping characteristics, can solve the problem of classification bias caused by data imbalance and meet the delay requirements of real-time detection. This fusion model not only retains the advantage of Transformer in capturing global transaction associations, but also compensates for the shortcomings of temporal feature extraction, real-time performance, and generalization through optimization of LSTM and KELM, providing a new solution for high-precision and low latency credit card fraud detection.

DOI: 10.54254/2755-2721/2025.AST27335

2. Data source

The data used in this article is the Kaggle open source dataset, which contains 28 principal component variables, namely V1-V28. This article selected 8023 data points, including 492 fraudulent data points labeled as category 1, and 7531 non fraudulent data points. Correlation analysis was conducted on the data, and the results are shown in Figure 1.

According to the ranking chart of credit card fraud correlation, V14 shows the strongest correlation, with an absolute correlation value close to 0.8, far higher than other variables; Following closely behind are variables V3 (0.78), V17 (0.66), V10 (0.64), etc., all of which have correlation values above 0.5, indicating a close potential association with fraud. By comparison, the correlation value of Amount (transaction amount) is only about 0.07, while the correlation values of variables such as V15 and V23 almost approach 0, indicating extremely weak explanatory power for fraud. Overall, high correlation variables such as V14, V3, and V17 have higher analytical priority in credit card fraud identification.

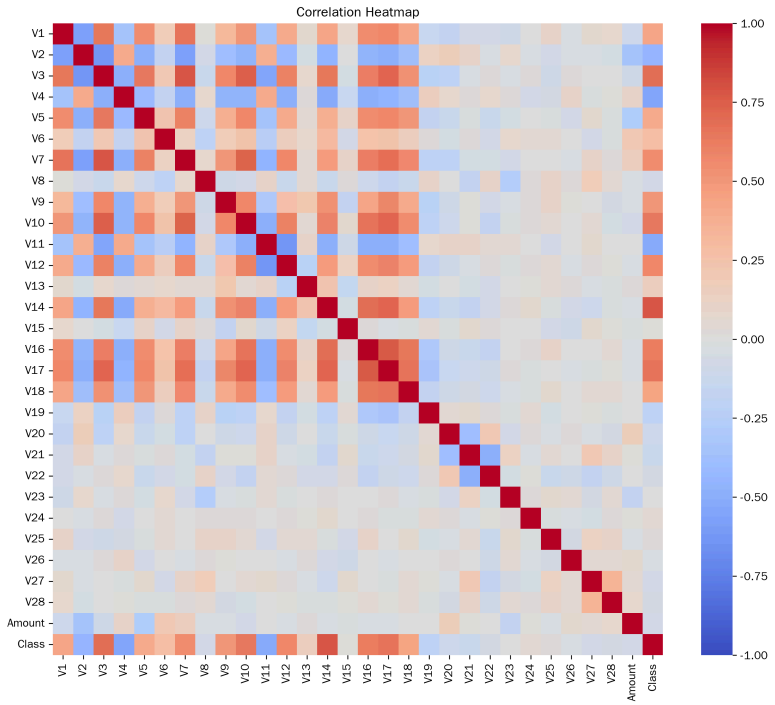

Figure 1. Confusion matrix heatmap

3. Method

3.1. LSTM

LSTM is an improved version of Recurrent Neural Network (RNN), with the core goal of solving the gradient vanishing problem in traditional RNNs when processing long sequences, thereby effectively capturing long-term dependencies in sequence data. Its core design revolves around the cellular state, which is similar to an information conveyor belt that runs through a network, maintaining the stability of key information during sequence transmission without being excessively disturbed by short-term noise. To achieve precise control of cell states, LSTM introduces three gating mechanisms: forget gate, input gate, and output gate. Each gating mechanism determines the proportion of information retained or discarded through a sigmoid activation function, and then

DOI: 10.54254/2755-2721/2025.AST27335

combines with the tanh function to complete information updates [6]. The network structure of LSTM is shown in Figure 2.

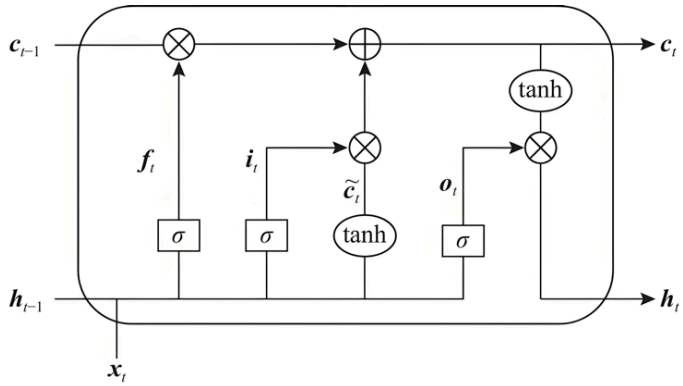

Figure 2. The network structure of LSTM

Specifically, the forget gate is responsible for filtering historical information: it receives the hidden state of the previous moment and the current input, calculates the forget weight through sigmoid, multiplies it with the cell state of the previous moment, and determines which historical information needs to be retained or discarded [7].

3.2. KELM

Kernel Extreme Learning Machine (KELM) is an improved algorithm developed based on Extreme Learning Machine (ELM). Its core is to enhance the model's ability to handle nonlinear problems through kernel techniques, while retaining the advantages of fast training and concise structure of ELM. As a simplified form of a single hidden layer feedforward neural network, ELM's core feature is that the weights and biases from the input layer to the hidden layer are randomly initialized without iterative adjustment, and training is completed only by solving the output weights, greatly improving efficiency. However, ELM has limited fitting ability for nonlinear data, so KELM introduces kernel methods to map low dimensional input space to high-dimensional feature space through kernel functions, constructing a linear model in the high-dimensional space, indirectly capturing nonlinear relationships, and avoiding explicit design of hidden node dimensions and activation functions, simplifying the model structure [8]. The network structure of KELM is shown in Figure 3.

DOI: 10.54254/2755-2721/2025.AST27335

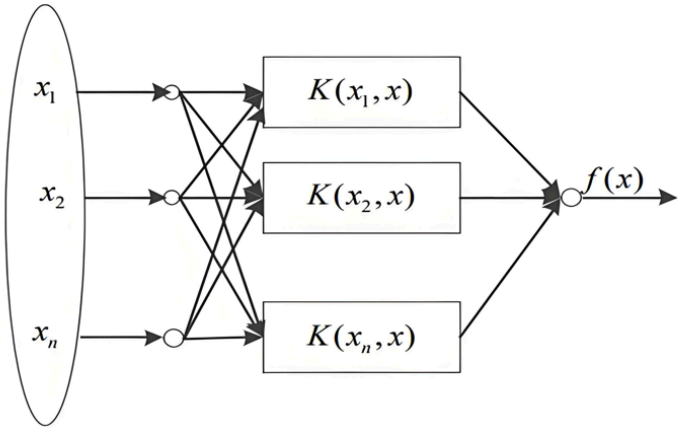

Figure 3. The network structure of KELM

3.3. Transformer

Transformer is a sequence modeling architecture based on self attention mechanism, which completely eliminates the sequence dependency characteristics of recurrent neural networks (RNNs) and greatly improves the efficiency of processing long sequences through parallel computing, becoming a fundamental model in fields such as natural language processing (NLP). Its core architecture consists of an encoder and a decoder, both of which are stacked from multiple identical layers. Each layer ensures training stability through residual connections and layer normalization [9].

The core of the encoder is a multi head self attention mechanism: each input token generates three vectors: Query, Key, and Value. The attention weight is obtained by calculating the similarity between Query and all Keys, and then the Value is weighted and summed to obtain the attention result of the token on all other tokens in the sequence. Multi head is the process of splitting this process into multiple subspaces in parallel, capturing correlations from different dimensions, and finally concatenating the results. The self attention is followed by a feedforward neural network, which is responsible for nonlinear mapping of the attention output. The function of the encoder is to transform the input sequence into a feature matrix containing global contextual information. The decoder consists of three sub layers: one is masked multi head self attention; The second is encoder decoder attention; The third is the same feedforward neural network as the encoder [10].

3.4. LSTM-KELM-transformer

The Transformer classification algorithm optimized by LSTM and KELM is a hybrid architecture that combines the advantages of three models, aiming to improve the feature capture ability and classification accuracy of sequence classification tasks. The core idea is to first preprocess the input sequence using LSTM. You use its gating mechanism to extract local temporal dependencies, and then input the preprocessed results into the Transformer encoder. Through a multi head self attention mechanism, you model the global context association to compensate for the shortcomings of LSTM in capturing long-range dependencies; At the same time, KELM is used to replace the traditional linear classification layer of Transformer, and the high-dimensional features are mapped to the reproducing kernel Hilbert space using kernel functions. Nonlinear classification is achieved by solving the regularized least squares problem, which not only retains the efficient and anti

DOI: 10.54254/2755-2721/2025.AST27335

overfitting characteristics of KELM training, but also avoids the fitting limitations of traditional softmax layers on complex feature distributions.

4. Result

In terms of experimental parameter settings, the proportion of the training set to the dataset is 0.7, the number of input channels is the feature dimension, the maximum position encoding is 256 times 2, the number of heads in the self attention mechanism is 4, the number of key channels per head is 32, and the total number of key channels is 128. The model includes a sequence input layer, a positional embedding layer, an addition layer, two self attention layers, a 6-unit LSTM layer, a ReLU activation layer, a 0.01 dropout layer, an index layer, a fully connected layer corresponding to the number of categories, a softmax layer, and a classification layer.In the selection of comparative models, this article uses decision trees, random forests, Adaboost, Gradient Boosting Tree, CatBoost, Logistic regression, and BP neural network. The results of the comparative experiment are shown in Table 1. Output a comparison bar chart for each indicator, as shown in Figure 4.

Table 1. Physical properties of Al-9.6at%Sc alloy

Model	Accuracy	Recall	Precision	F1	AUC
Decision tree	0.988	0.988	0.988	0.988	0.985
Random Forest	0.992	0.992	0.992	0.991	0.999
Adaboost	0.992	0.992	0.992	0.992	0.998
Gradient Boosting Tree (GBDT)	0.985	0.985	0.985	0.985	0.998
CatBoost	0.992	0.992	0.992	0.991	0.999
Logistic regression	0.99	0.99	0.99	0.99	0.998
BP neural network	0.992	0.992	0.992	0.992	0.998
LSTM-KELM-Transformer	0.999	0.999	0.999	0.999	1

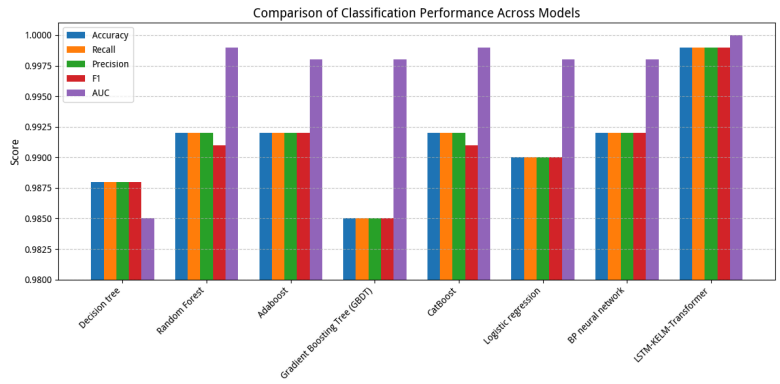

Figure 4. The comparison bar chart for each indicator

From the experimental results, LSTM-KELM Transformer performs the best in all evaluation metrics, with accuracy, recall, precision, and F1 values reaching 0.999, and AUC reaching 1, significantly better than other models. Among other models, Random Forest, Adaboost, CatBoost, and BP neural network have similar performance, with Accuracy, Recall, and Precision mostly around 0.992, slightly lower F1 values, and AUC between 0.998-0.999, belonging to the suboptimal level; The various indicators of logistic regression are slightly weaker than the above model, about

DOI: 10.54254/2755-2721/2025.AST27335

0.99; The performance of Decision Tree and Gradient Boosting Tree (GBDT) is relatively poor, especially with GBDT's Accuracy, Recall and other indicators only reaching 0.985, and Decision Tree's AUC also only reaching 0.985. Overall, LSTM-KELM Transformer outperforms other models in classification accuracy, recognition ability of positive and negative samples, comprehensive performance, and ability to distinguish positive and negative classes, demonstrating stronger classification advantages and contributing to the correct classification of fraud detection.

Output the prediction confusion matrix of the LSTM-KELM Transformer test set, as shown in Figure 5. From the confusion matrix, it can be seen that 3 fraudulent instances were detected, while the other 5613 instances were predicted correctly.

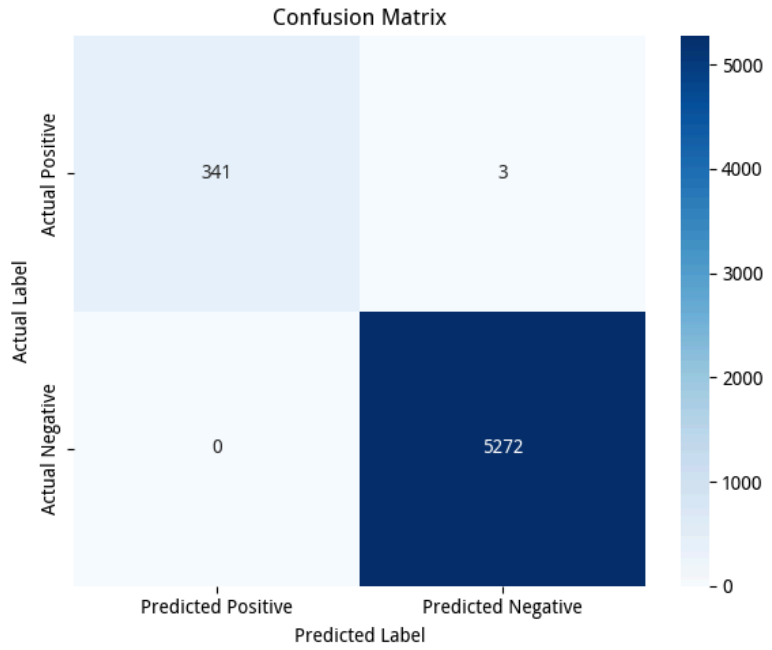

Figure 5. The prediction confusion matrix of the LSTM-KELM Transformer test set

5. Conclusion

Among other models, Random Forest, Adaboost, CatBoost, and BP neural networks have similar performance, with Accuracy, Recall, and Precision mostly around 0.992, slightly lower F1 values, and AUC in the range of 0.998-0.999, belonging to the suboptimal level; The various indicators of logistic regression are slightly inferior to the above model, about 0.99; The performance of Decision tree and Gradient Boosting Tree (GBDT) is relatively weak, especially with GBDT's Accuracy, Recall and other indicators only 0.985, and Decision tree's AUC only 0.985. Overall, LSTM-KELM Transformer surpasses other models in classification accuracy, positive and negative sample recognition ability, comprehensive performance, and positive and negative class discrimination ability, demonstrating stronger classification advantages and providing strong support for accurate classification of credit card fraud detection. This model not only effectively improves the technical level of fraud detection, but also helps reduce the fraud losses of financial institutions and users. It is of great practical significance to maintain the security of global credit card transactions and stabilize the order of financial markets.

DOI: 10.54254/2755-2721/2025.AST27335

References

[1] Bello, Oluwabusayo Adijat, et al. "Analysing the impact of advanced analytics on fraud detection: a machine learning perspective." European Journal of Computer Science and Information Technology 11.6 (2023): 103-126.
[2] Manoharan, Geetha, et al. "Machine learning-based real-time fraud detection in financial transactions." 2024 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI). IEEE, 2024.
[3] Achary, Rathnakar, and Chetan J. Shelke. "Fraud detection in banking transactions using machine learning." 2023 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE). IEEE, 2023.
[4] Pranto, Tahmid Hasan, et al. "Blockchain and machine learning for fraud detection: A privacy-preserving and adaptive incentive based approach." Ieee Access 10 (2022): 87115-87134.
[5] Singh, Aditi, et al. "Design and implementation of different machine learning algorithms for credit card fraud detection." 2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME). IEEE, 2022.
[6] Schneider, Moritz, and Rolf Brühl. "Disentangling the black box around CEO and financial information-based accounting fraud detection: machine learning-based evidence from publicly listed US firms." Journal of Business Economics 93.9 (2023): 1591-1628.
[7] Verma, Jyoti. "Application of machine learning for fraud detection–a decision support system in the insurance sector." Big data analytics in the insurance market. Emerald Publishing Limited, 2022. 251-262.
[8] Chhabra, Raunak, Shailza Goswami, and Ranjeet Kumar Ranjan. "A voting ensemble machine learning based credit card fraud detection using highly imbalance data." Multimedia Tools and Applications 83.18 (2024): 54729-54753.
[9] Enjyo T, Ikeuchi K, Kanai M, Maruyama T. Diffusion welding of aluminum to titanium. Trans. JWRI, 1977, 6(1): 123-130.
[10] Karunachandra, Bryan, et al. "On the benefits of machine learning classification in cashback fraud detection." Procedia Computer Science 216 (2023): 364-369.