Credit Card Anomaly Detection Based on Multi-Dimensionally Optimized Autoencoder

Qin-jie Zhou
School of Electronic
Information and Electrical
Engineering
Yangtze University
Jingzhou, China

Abstract: To safeguard the financial security of both credit card holders and issuing institutions, this paper proposes an anomaly detection method for credit card transactions based on a multi-dimensionally optimized autoencoder. The data is preprocessed through standardization and stratified sampling to ensure distribution consistency, while isolation forest is employed to filter pure normal samples for model training. To address the limitations of the original autoencoder, optimizations are introduced across several dimensions, including the network architecture, activation functions, and training strategy. Experimental results demonstrate that the proposed model outperforms conventional methods in detection performance, exhibiting enhanced practicality and adaptability for real-world applications.

Keywords: Anomaly Detection；Credit Card Transactions；Autoencoder；Multi-dimensional Optimization；IsolationForest

1. INTRODUCTION

In recent years, the rapid development of China’s economic and financial infrastructure has led to continuous improvements in the credit card payment system [1]. Valued for their convenience, credit cards have gained widespread popularity among consumers and have become a significant component of daily transactions. However, alongside the expansion of the credit card market and its growing integration into everyday life, the prevalence of credit card fraud has also increased [1]. Numerous fraudulent incidents have resulted in substantial financial losses and reputational damage to banks and related institutions. Consequently, the prevention and detection of credit card fraud remain critical areas of research.

Traditional methods for detecting anomalous credit card transactions often rely on manually defined thresholds set by bank staff, combined with data analysis and domain-specific financial knowledge. These approaches not only consume substantial human and material resources but also exhibit limited detection efficiency, struggling to address increasingly sophisticated and evolving fraudulent tactics. With advances in artificial intelligence, machine learning and deep learning techniques have progressively become primary tools in this research domain. A key challenge in credit card anomaly detection arises from the extreme class imbalance—fraudulent cases are vastly outnumbered by legitimate transactions—which complicates the learning process and underscores the need for improved detection accuracy[3].

Most current studies employ supervised learning for credit card anomaly detection. However, due to the scarcity of labeled data and the limited size of annotated datasets, it is often difficult to adequately capture genuine fraudulent patterns [4]. Furthermore, the pronounced class imbalance in transaction data exacerbates the challenges of anomaly detection. Therefore, developing an effective unsupervised learning-based detection framework is of considerable importance.

To address these issues, this paper proposes a novel credit card anomaly detection method based on a multi-dimensionally optimized auto-encoder. The approach systematically enhances the network architecture, optimizes activation functions, and refines training strategies to improve the model's ability to learn normal transaction patterns. It also incorporates data preprocessing and sample selection mechanisms to boost detection robustness and accuracy under highly imbalanced data conditions. Experimental results demonstrate that the proposed method outperforms conventional detection models across multiple evaluation metrics, indicating strong practical value and application potential.

2. RESEARCH FRAMEWORK

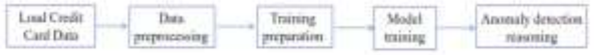

Figure 1 Methodological framework of the study

The study of credit card anomaly detection primarily involves data preprocessing, model training, performance evaluation, and outlier detection. The main framework is illustrated in Figure 1. Initially, credit card data is loaded and preprocessed. Subsequently, preparatory steps are taken before training, followed by the model training phase. Once model training is completed, anomaly detection inference is finally conducted.

2.1 Autoencoder algorithm

Autoencoder (AE) is a classical unsupervised neural network architecture designed to learn efficient representations of input data by reconstructing the input itself as the learning objective [5].

It consists of two core components: an encoder and a decoder. The encoder compresses the input data into a low-dimensional latent representation through a series of neural network layers, thereby performing feature extraction and information compression. The decoder then reconstructs the high-dimensional output from this low-dimensional code via another series of neural network layers, aiming to approximate the original input data based on the encoded representation. The overall architecture is illustrated in Figure2.

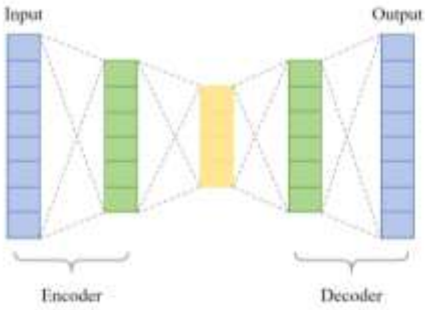

Figure 2 Autoencoder structure

Typically, an autoencoder is trained to capture the patterns of normal data and can accurately reconstruct such data. However, when anomalous data is fed into the model—whose patterns differ from those learned during training—the reconstruction error increases significantly due to the model's inability to reproduce it faithfully. Therefore, by setting an appropriate threshold for the reconstruction error, input data yielding an error above this threshold can be classified as anomalous. In essence, the core principle of autoencoder-based anomaly detection lies in learning the representation of normal data and leveraging the reconstruction error to discriminate between normal and abnormal samples, making it an effective unsupervised approach for anomaly detection.

The specific workflow is illustrated in Figure 3.

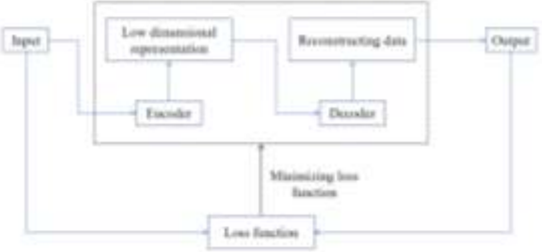

Figure 3 Workflow of the Autoencoder-based detection framework

First, the input credit card data is encoded into a low-dimensional representation via the encoder, and then reconstructed by the decoder. The mean absolute error between the input and the reconstructed data is computed as the loss function. The structural parameters of the Autoencoder are then iteratively optimized by minimizing this loss, thereby yielding an optimally trained model.

2.2 Data preprocessing

This study employs the publicly available "Credit Card Fraud Detection" dataset from Kaggle, the distribution of which is illustrated in Figure 4. As shown in the pie chart, normal transaction samples comprise 99.8273% (284,315 instances), whereas fraudulent samples constitute only 0.1727% (492 instances), resulting in a pronounced class-imbalance problem[6].

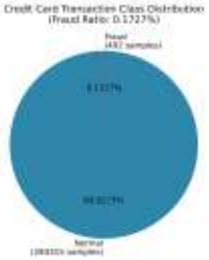

Figure 4 Data sample distribution

To enhance compatibility with the Autoencoder network, the data undergo preprocessing prior to training. The preprocessing workflow is illustrated in Figure 5.

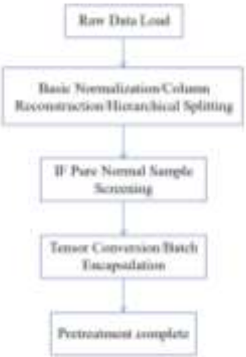
Raw Data Load
Basic Normalization/Column Reconstruction/Hierarchical Splitting
IF Pure Normal Sample Screening
Tensor Conversion/Batch Encapsulation
Pretreatment complete

Figure 5 Workflow of Data Preprocessing

First, the dataset is imported by loading the "creditcard.csv" file, which comprises numerous credit card transaction records. Each record contains multiple features along with the target variable "Class" (where "Class = 1" denotes a fraudulent transaction and "Class = 0" indicates a normal transaction). Upon successful data loading, basic preprocessing is performed on the dataset. The two features "Time" and "Amount" are normalized using StandardScaler to rescale them to a distribution with zero mean and unit variance. The data distributions before and after normalization are displayed in Figure 6. Subsequently, the original "Time" and "Amount" columns are removed, retaining only the standardized features, and the target variable "Class" is moved to the end of the dataset to clearly separate features from labels. The data are then split into training and test sets at an 8:2 ratio using stratified sampling (stratify = y) to ensure that the proportion of fraudulent samples remains consistent with the original distribution and to avoid sampling bias.

To address the core challenge of extreme class imbalance, the isolated forest algorithm is applied to the training samples to filter out a set of pure normal transaction samples. This provides the unsupervised Autoencoder with training data that reflect the feature distribution of normal transactions only. Finally, the selected normal samples are converted into tensor format and packaged for subsequent model training.

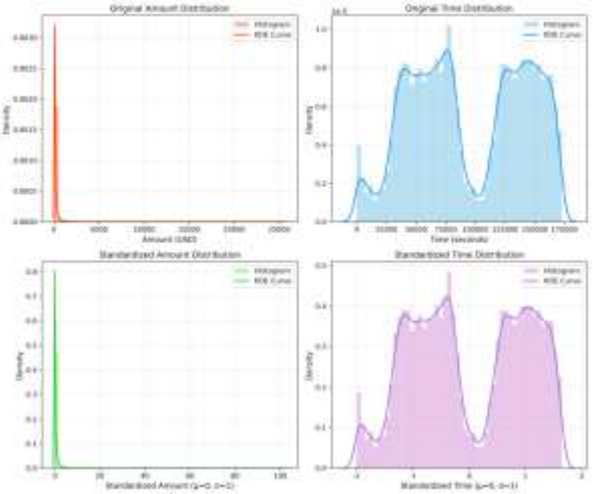

Figure6 Data distribution before and after normalization

2.3 Autoencoder Structure Optimization

The Optimized Autoencoder proposed in this study builds upon the original Autoencoder (Original Autoencoder) architecture by enhancing both its encoder and decoder components to improve detection accuracy for credit card anomaly data.

2.3.1 Activation Function

The Original Autoencoder typically employs the ReLU (Rectified Linear Unit) function as its activation function, whose curve is shown in Figure 7. However, when the input gradient is negative, certain neurons may become permanently inactive, leading to the vanishing gradient problem in deeper networks and hindering effective backpropagation. To accommodate deeper network architectures, this study adopts the LeakyReLU function (with a negative slope of 0.1) as the activation function. As illustrated in the figure, LeakyReLU maintains a small gradient in the negative region, thereby ensuring stable gradient flow through deep layers and enabling the model to learn more complex patterns of normal transaction behavior.

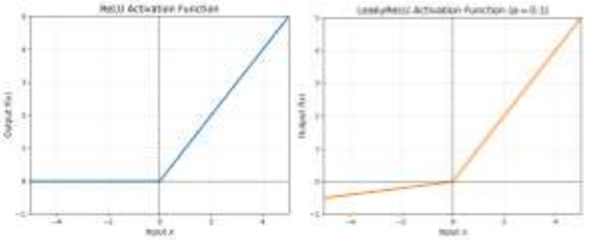

Figure 7 Comparison of ReLU and LeakyReLU activation functions

2.3.2 Encoder Structure Optimization

Building upon the original Autoencoder architecture, the encoder and decoder structures have been refined to enhance model performance. Figure 8 provides a comparative illustration of the Original_Autoencoder and the Optimized_Autoencoder structures.

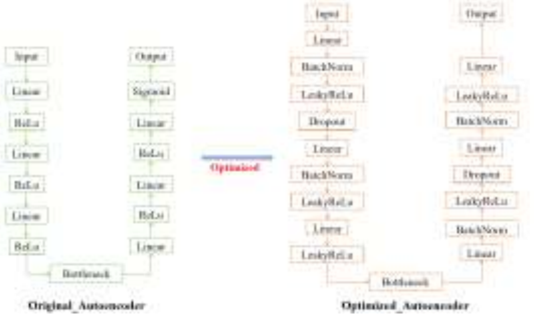

Figure 8 Architectural comparison between the Original Autoencoder and the Optimized Autoencoder

Compared with the Original Autoencoder, the Optimized Autoencoder incorporates several key enhancements. First, the network width and bottleneck dimension are expanded: the encoder structure is adjusted from “30 → 64 → 32 → 16” to “30 → 128 → 64 → 32,” which enhances the model’s capacity to encode high-dimensional features and better preserves the complex feature correlations present in normal transactions. Second, the ReLU activation function is replaced with LeakyReLU (negative slope = 0.1) to alleviate gradient vanishing in the negative region. BatchNorm1d is introduced after each linear layer to mitigate internal covariate shift and improve training stability. Third, Dropout (drop rate = 0.1) is applied in both the encoder and decoder, and the AdamW optimizer (weight decay = 1e-5) is adopted to reduce overfitting to noise in the normal samples. Finally, the Sigmoid activation function at the decoder output layer is removed to accommodate the continuous-value distribution of the normalized input, thereby avoiding distortion in reconstruction-error calculation.

2.4 Loss Function

This study adopts the L1 loss function—i.e., the mean absolute error (MAE)—for both the Optimized Autoencoder and the Original Autoencoder to quantify the discrepancy between the model’s output and the target. In an Autoencoder, the objective is to reconstruct the input with minimal deviation. The L1 loss computes the average absolute difference between the predicted and actual values and demonstrates reduced sensitivity to outliers. Compared with the L2 loss (mean squared error, MSE), the L1 loss exhibits enhanced robustness when handling anomalous data points. Its formulation is as follows:

$$MAE\left(x_i, \bar{x}_i\right) = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}_i\right)^2$$

(2-1)

3. EXPERIMENT AND RESULT ANALYSIS

3.1 Experimental Settings

3.1.1 Experimental Parameter Setting

The experiments were conducted in a Windows 10 (64-bit) environment equipped with an AMD Ryzen 5 processor and 8 GB of RAM. The code was implemented in Python 3.6.1 using VS Code as the development platform.

3.1.2 Comparative Models

This study compares the proposed Optimized Autoencoder against three representative baseline methods: the Original Autoencoder, Logistic Regression (supervised), and Isolation

Forest (unsupervised). The main parameter configurations for each algorithm are summarized in Table 1.

Table 1 Main parameters of the algorithm

Algorithm	Main parameters
Optimized_Autoencoder	BatchNorm1d, LeakyReLU (0.1), Dropout (0.1),AdamW (lr=1e-3, weight_decay=1e-5), batch_size=256
Original_Autoencoder	ReLU/Sigmoid, Adam (lr=1e-3), batch_size=256
LogisticRegression	solver='liblinear', random_state=42
IsolationForest	n_estimators=500,contamination =0.001727

3.1.3 Evaluation Metrics

The performance of anomaly detection models is commonly assessed using metrics such as ROC-AUC, Precision, Recall, and the F1-Score. These indicators are particularly suitable for evaluating classification models on imbalanced datasets, where higher values correspond to better classification performance[7].

These metrics are derived from the following basic terms of the confusion matrix:

TP (True Positives): The number of correctly identified positive (anomalous) samples.

FP (False Positives): The number of negative (normal) samples incorrectly predicted as positive.

TN (True Negatives): The number of correctly identified negative samples.

FN (False Negatives): The number of positive samples incorrectly predicted as negative.

(1) ROC-AUC

In credit card anomaly detection tasks, the performance of the Autoencoder algorithm is primarily evaluated on the test set.

The True Positive Rate (TPR), also referred to as recall or sensitivity, represents the proportion of actual positive (fraudulent) samples that are correctly identified by the model. It is calculated as follows:

$$TPR = \frac{TP}{TP + FN} \quad (3\text{-}1)$$

(2) False Positive Rate (FPR)

The False Positive Rate (FPR) measures the proportion of actual negative (normal) samples that are incorrectly predicted as positive by the model. It is calculated as follows:

$$FPR = \frac{FP}{FP + TN} \quad (3\text{-}2)$$

The area under the ROC curve (ROC-AUC) is a key metric for evaluating the performance of a classification model. It quantifies the trade-off between the model's true positive rate and false positive rate across different classification thresholds. The ROC-AUC value ranges from 0 to 1, with a value closer to 1 indicating superior model performance. In the context of credit card anomaly detection, the ROC-AUC effectively measures a model's discriminative power. A higher ROC-AUC value signifies that the model is more capable of distinguishing normal transactions from fraudulent ones.

（2）Precision

Precision quantifies the proportion of correctly identified fraudulent samples among all instances predicted as fraud. It is defined as:

$$Precision = \frac{TP}{TP + FP} \quad (3\text{-}3)$$

(3) Recall

Recall measures the proportion of actual fraudulent transactions that are correctly identified by the model among all actual fraud samples. It is defined as:

$$Recall = \frac{TP}{TP + FN} \quad (3\text{-}4)$$

(4) F1-Score

The F1-Score is the harmonic mean of Precision and Recall, providing a single balanced metric that accounts for both false positives and false negatives. It is particularly useful for evaluating performance on imbalanced datasets, such as in credit card fraud detection. The F1-Score is calculated as follows:

$$F1 = \frac{2 \times Precison \times Recall}{Precision + Recall} \quad (3\text{-}5)$$

3.2 Experimental Results and Analysis

3.2.1 Comparative Experiments

In this experiment, the proposed Optimized Autoencoder is compared against three baseline algorithms: the Original Autoencoder (unsupervised), Logistic Regression (supervised), and Isolation Forest (unsupervised). All four models are trained and evaluated on the same test set under identical conditions. The comparative performance results are summarized in Table 2.

Table 2 Comparative performance of the evaluated models

Anomaly detection model	ROC_ AUC	Precision	Recall	F1-Score
Optimized_ Autoencoder	**0.9653**	**0.7619**	**0.8163**	**0.7882**
Original_ Autoencoder	0.9523	0.4800	0.2449	0.3243
Logistic-Regression	0.9575	0.7080	0.8103	0.7557
IsolationForest	0.9583	0.2603	0.5816	0.3596

Figure 9 presents the ROC curves of the four evaluated models. A model's performance is considered superior when its AUC value is closer to 1.

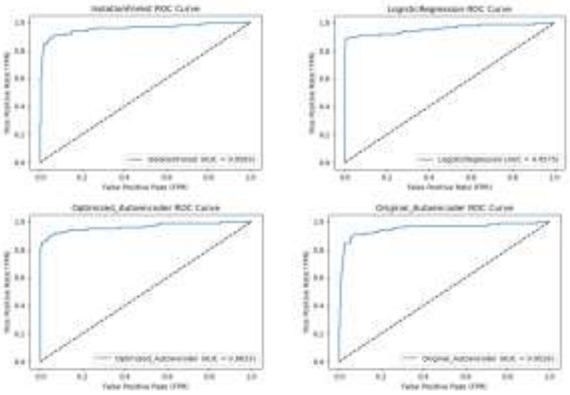

Figure 9 ROC curves of the evaluated models

3.2.2 Analysis of Results

In the context of extreme class imbalance characteristic of credit card anomaly detection, this study compares four algorithms: the Original Autoencoder, the Optimized Autoencoder, Logistic Regression (supervised), and Isolation Forest (unsupervised). The experimental outcomes, summarized in Table 2 and illustrated in Figure 9, demonstrate that the Optimized Autoencoder holds notable advantages in terms of global discriminative ability, precision, fraud detection rate, and overall performance.

Global Discriminative Ability: The Optimized Autoencoder achieves an ROC-AUC of 0.9653, outperforming all other methods. This indicates its superior capacity to distinguish between the feature distributions of normal and fraudulent transactions.

Precision: With a precision score of 0.7619, the Optimized Autoencoder shows an improvement of 58.7% over the Original Autoencoder (0.4800) and 191.9% over Isolation Forest (0.2603). This means that 76.19% of the samples flagged as fraudulent by the model are indeed fraudulent, thereby effectively reducing misjudgment costs and mitigating negative impacts on user experience in credit card risk control.

Recall: The recall (fraud detection rate) of the Optimized Autoencoder reaches 0.8163, which is 233.3% higher than that of the Original Autoencoder (0.2449) and slightly exceeds the supervised baseline Logistic Regression model (0.8103). This result reflects a successful containment of fraudulent transaction risks at a low level.

F1-Score: The Optimized Autoencoder attains an F1-score of 0.7882, representing an optimal balance between the “cost of false alarms” and the “risk of missed detection.” This value is 2.43 times that of the Original Autoencoder (0.3243), 4.3% higher than Logistic Regression (0.7557), and 119.2% greater than Isolation Forest (0.3596).

These results confirm that the Optimized Autoencoder not only overcomes the inherent accuracy limitations typical of unsupervised models but also surpasses the performance of the supervised baseline. Consequently, it exhibits higher practical value and application potential for real-world credit card anomaly detection.

4. SUMMARY

This study addresses the challenge of severe class imbalance in credit-card anomaly detection by proposing a multi-dimensional optimization strategy for Autoencoders. Through comparative experiments, the effectiveness and superiority of the optimized model are validated, demonstrating its ability to identify a greater number of genuine fraudulent transactions. In future work, the Autoencoder framework could be further enhanced through integration with graph neural networks and attention mechanisms to capture complex relational patterns in transaction data. Such advancements would help advance the adoption of unsupervised anomaly detection techniques in the field of financial risk control.

5. REFERENCES

[1] Guo T, Li G Y, Yuan D. Credit card anomaly detection based on confidence and neural network [J]. Computer Engineering, 2008, (15): 205-207 225.

[2] Xu T P, Luo Y S. Credit card fraud detection model based on ensemble learning [J]. Information Systems Engineering, 2024, (01): 129-132.

[3] Liu R X, Xu H Z. WGAN-BiLSTM credit card fraud detection method based on attention mechanism optimization [J]. Modern Electronic Technology, 2024, 47 (10): 73-78. DOI: 10.16652/J.issn.1004-373x. 2024.10.01 4.

[4] Cheng J H, Pang M L. Anomaly detection of credit card approval based on dynamic integrated selection algorithm [J]. Journal of Hefei University (Comprehensive Edition), 2023, 40 (05): 86-94.

[5] Luo Y P, He A L, Yu K M, et al. Structural damage early warning method based on convolutional neural network and self-encoder [J/OL]. Steel Structure (Chinese and English), 1-12 [2026-01-05].

[6] Tao M Z, Xiong X X, Chen J W, et al. Credit card fraud detection based on denoising diffusion probability model [J/OL]. Computer Science and Exploration, 1-20 [2026-01-05].

[7] Li Y T, Yang W J. Credit card fraud detection model based on deep learning [J/OL]. Journal of Tianjin University of Technology, 1-9 [2026-01-05].