systems

MDPI

Article

Credit Card Fraud Detection Based on Unsupervised Attentional Anomaly Detection Network

Shanshan Jiang [1], Ruiting Dong [1], Jie Wang [1] and Min Xia [2,*]

[1] School of Management Science and Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China; jss@nuist.edu.cn (S.J.); 202252990035@nuist.edu.cn (R.D.); 202212420009@nuist.edu.cn (J.W.)

[2] Jiangsu Key Laboratory of Big Data Analysis Technology, Nanjing University of Information Science and Technology, Nanjing 210044, China

* Correspondence: xiamin@nuist.edu.cn

Abstract: In recent years, with the rapid development of Internet technology, the number of credit card users has increased significantly. Subsequently, credit card fraud has caused a large amount of economic losses to individual users and related financial enterprises. At present, traditional machine learning methods (such as SVM, random forest, Markov model, etc.) have been widely studied in credit card fraud detection, but these methods are often have difficulty in demonstrating their effectiveness when faced with unknown attack patterns. In this paper, a new Unsupervised Attentional Anomaly Detection Network-based Credit Card Fraud Detection framework (UAAD-FDNet) is proposed. Among them, fraudulent transactions are regarded as abnormal samples, and autoencoders with Feature Attention and GANs are used to effectively separate them from massive transaction data. Extensive experimental results on Kaggle Credit Card Fraud Detection Dataset and IEEE-CIS Fraud Detection Dataset demonstrate that the proposed method outperforms existing fraud detection methods.

Keywords: fraud detection; anomaly detection; unsupervised learning; autoencoders; GANs

Citation: Jiang, S.; Wang, J.; Dong, R.; Xia, M. Credit Card Fraud Detection Based on Unsupervised Attentional Anomaly Detection Network. *Systems* **2023**, *11*, 305. https://doi.org/10.3390/systems11060305

Academic Editors: Wendong Yang, Jinpei Liu and Jianzhou Wang

Received: 7 May 2023
Revised: 2 June 2023
Accepted: 12 June 2023
Published: 13 June 2023

1. Introduction

In recent years, the development of technologies such as big data and artificial intelligence has fully upgraded payment methods. The popularity of credit card mobile payment has given fraudsters greater opportunities for credit card fraud, through methods such as credit card cashing, counterfeit card fraud, payment fraud, etc. [1,2]. At present, the means of fraud are characterized by high technology, concealment, and cross-regional crimes. The implementation process of the case is more concealed, the techniques are constantly being renewed, and the risk of fraud is gradually moving forward to the business application process [3]. This has produced a large number of non-performing assets and had a certain degree of impact on the stability of the economic market [4]. Therefore, the prevention and detection of credit card fraud has always been a research area that has attracted significant attention.

With the advent of the information age, a large number of researchers and scholars have conducted research on credit card fraud detection, including improving detection efficiency by overcoming the unbalanced characteristics of credit card transactions and treating fraudulent transactions as abnormal outliers. In recent years, the application of machine learning [5–8] to credit card fraud detection has overcome many of the shortcomings of traditional fraud detection methods and formed a related research field. Many scholars at home and abroad have conducted a lot of research and analysis on credit card fraud data and cases by using machine learning and data mining methods [9–11]. However, due to the high imbalance of credit card fraud transaction data (that is, the number of fraud samples is much smaller than the number of transaction samples), how effective algorithms can be used to further improve the accuracy of credit card fraud detection is still an urgent problem

to be solved. This work aims to survey existing credit card fraud detection methods and propose a new credit card fraud detection network based on the unsupervised attentional anomaly detection paradigm. The network follows the training paradigm of Generative Adversarial Network (GAN) [12] and mainly consists of a generator and a discriminator. The generator generates samples as close as possible to the real data through self-supervised learning [13–15], and effectively encodes the high-level feature representation (hidden vector) of the normal transaction data distribution, and the discriminator detects the forged ones of the generator as much as possible. We used data samples to form an adversarial training mode. In the generator, we propose a channel-wise feature attention, which enables the network to better reconstruct more realistic pseudo-samples during the training phase, which helps to learn the hidden layer feature representation of normal transactions. In order to effectively supervise the training process of the model, this paper also proposes a hybrid weighted loss function. In the test phase, the hidden vector and the distance between the reconstructed sample and the input sample in the feature space are calculated to determine whether the current transaction sample is fraudulent. In the experimental part, we compared the proposed method with the existing machine learning methods and deep learning methods on Kaggle Credit Card Fraud Detection Dataset and IEEE-CIS Fraud Detection Dataset to prove its advancement.

The main contributions of this paper lie in the following aspects:

- Reframe the problem of credit card fraud detection as anomaly detection of fraudulent transactions, and propose a new credit card Fraud Detection framework based on Unsupervised Attentional Anomaly Detection Network (UAAD-FDNet).
- A channel-wise feature attention is proposed. This module enables the network to effectively capture the interdependence between feature channels to better learn how to reconstruct normal transaction samples.
- A hybrid weighted loss function is proposed to enable the model to learn an effective encoding method for hidden vectors and reconstruct samples as realistically as possible. In the test phase, fraudulent transactions are identified by calculating the hidden vectors and the characteristic distance between the reconstructed samples and the input samples.
- Experimental results on Kaggle Credit Card Fraud Detection Dataset and IEEE-CIS Fraud Detection Dataset show that our method outperforms existing machine learning-based and deep learning-based fraud detection methods.

2. Background and Related Work

As a product integrating financial business and Internet technology, credit card has attracted much attention since its appearance. With the increasingly prominent problem of credit card fraud, a large number of scholars have carried out in-depth research on this problem in recent years.

Before the rise of deep learning methods, traditional machine learning methods (such as Support Vector Machine (SVM), random forest, Markov model, etc.) were often used to solve the problem of credit card fraud detection. In 2006, Chen et al. [16] proposed an effective fraud detection system based on SVM and genetic algorithm for the unevenly distributed few-sample data. In 2012, Khan et al. [17] used Hidden Markov Model (HMM) to make fraudulent transactions easier to detect by reducing the complexity of the algorithm. In 2015, Zareapoor et al. [18] proposed a new bagging ensemble classifier based on the decision tree algorithm. On a real credit card transaction dataset, this method can achieve real-time reasoning and effectively solve the problem of category imbalance of credit card transaction data. In 2018, Yee et al. [19] studied the performance of a variety of Bayesian classifiers on credit card fraud detection tasks. All classifiers can achieve good detection accuracy on the dataset processed by Principal Components Analysis (PCA). Although the above machine learning methods can solve the problem of financial fraud to a certain extent, given the complexity of credit card transaction behavior in the real world, how to

quickly and accurately extract typical features from limited transaction data is a topic that still needs further exploration.

In recent years, deep learning technology has demonstrated strong capabilities in many fields [20–26], the development of deep learning has significantly promoted the reform of the financial industry, and, at the same time, brought new ideas to the research of financial fraud detection. In 2016, Fu et al. [27] proposed a CNN-based fraud detection network, which learns the intrinsic patterns of fraudulent behavior from labeled data to identify whether there is fraudulent behavior in each transaction sample. In 2018, Chouiekh and Haj [28] explored the performance of Deep Convolutional Neural Network (DCNN) and some traditional machine learning methods on fraud events. The experimental results on mobile communication networks showed that DCNN is significantly better than other methods. However, due to the serious data imbalance in the financial fraud data in the real world, this brings severe challenges to the above methods. Saia et al. [29] proposed analyzing this issue and the heterogeneity of credit card data in the frequency domain to obtain a more stable information representation for fraud detection. In 2019, Fiore et al. [30] aimed at the class imbalance problem of financial fraud transaction data, and used GAN to generate minority class samples to train more effective classifiers. Saia and Carta [31] conducted research on proactive fraud detection based on Fourier transform and Wavelet transform. In 2022, Esenogho et al. [32] used hybrid data resampling technology and integrated learning to conduct robust credit card fraud detection based on LSTM [33] and AdaBoost [34]. In this paper, we reformulate the task of credit card fraud detection as an anomaly detection task, training autoencoder, and GAN in a clever way to avoid adverse effects of data imbalance on the model.

3. Methodology

In this paper, the credit card fraud detection problem is treated as an anomaly detection problem. Fraudulent transaction data are used as an abnormal sample in the transaction system. We use an unsupervised attentional anomaly detection network (including autoencoder with Feature Attention and GAN) to separate it from normal transaction data to complete the purpose of fraud detection. In the following, we will introduce the proposed credit card fraud detection algorithm in detail.

3.1. Proposed Model

The credit card Fraud Detection Network based on Unsupervised Attentional Anomaly Detection (UAAD-FDNet) proposed in this paper is mainly composed of a generator G and a discriminator D, as shown in Figure 1.

G consists of an autoencoder (including a pair of encoder G_E and decoder G_D) and an encoder E. The autoencoder first uses G_E to encode the input feature $\hat{x} \in \mathbb{R}^C$ into a hidden vector $z \in \mathbb{R}^d$, and then maps it back to the original feature space through G_D to obtain the reconstructed feature $\hat{x} \in \mathbb{R}^C$. During the training process, the abstract representation (advanced representation) of high-dimensional features can be effectively learned through the specific loss function self-encoding. E is used to re-encode the reconstructed feature, and map it to the hidden vector space to obtain $\hat{z} \in \mathbb{R}^c$. z and $\hat{z}$ have the same vector dimension, which enables us to supervise the training process of the network by computing feature distances between them. In terms of network structure, G_E and E have similar parameter structures and have a symmetrical relationship with G_D. Specifically, G_E mainly includes a fully connected layer (FC) and a LeakyReLU activation layer. FC is used to linearly map low-dimensional features to high-dimensional hidden space, and LeakyReLU is used to perform non-linear transformation on features to enhance the feature expression ability of the model. The same goes for E. G_D is the reverse process of G_E. They have a symmetrical network structure. The only difference is that the output layer features of G_D need to be activated by Tanh to normalize their feature values to $[-1,1]$ ($\hat{x}$ and x are in the same vector space). D contains only one encoder, which is used to extract the abstract features of x and $\hat{x}$, respectively, and judge whether the input feature is true or not through

the Softmax classification layer. The above basically form our credit card fraud detection framework. Considering that the fully connected layer is prone to overfitting, the dropout layer should be selectively added to our network according to the feature dimension of the input data.

In addition, in order to make G better learn the reconstruction method of $\hat{x}$, this paper also proposes a channel-wise feature attention, and its structure diagram is shown in Figure 2. For a set of transaction sample features $u \in \mathbb{R}^C$, we first group its feature channels to obtain two sets of sub-features $u_g \in \mathbb{R}^{N \times d}$ and $u_l \in \mathbb{R}^{d \times N}$ (N represents the number of groups, and d represents the feature dimension of each group). Then, feature computation with self-attention [35,36] is performed on two sets of sub-features separately to fully capture the global and local correlations among feature channels. Self-attention is widely used in the field of Natural Language Processing (NLP), which can effectively capture the dependencies between arbitrary features by computing the spatial distance of pairs of features. Specifically, it first performs linear transformation on the input feature $q \in \mathbb{R}^{k \times n}$ to obtain Query ($Q \in \mathbb{R}^{k \times n}$), Key ($K \in \mathbb{R}^{k \times n}$), and Value ($V \in \mathbb{R}^{k \times n}$). Then, the correlation between Q and K is calculated to obtain $S \in \mathbb{R}^{k \times k}$, each row of which passes through Softmax to represent the degree of correlation between paired features. Finally, S and V are fused to complete the output representation $P \in \mathbb{R}^{k \times n}$. The mathematical expression of the above process is as follows

$$p = Softmax(Q \circ K^T) \circ V, \tag{1}$$

where $\circ$ means matrix multiplication. In this paper, k and n are denoted as N, d, d, and N in the two branches, respectively. Finally, element-wise fusion is performed on the output features of the two branches to obtain the output of the feature attention module. We introduce this module in the skip connection of the generator. On the one hand, it can preserve the fine-grained features of the original sample, and on the other hand, it can effectively filter the redundant features in the transaction sample.

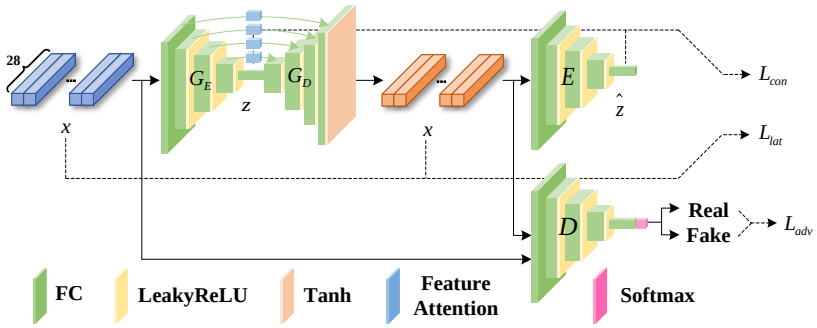


Figure 1. Schematic diagram of the framework structure of credit card fraud detection based on unsupervised attentional anomaly detection.

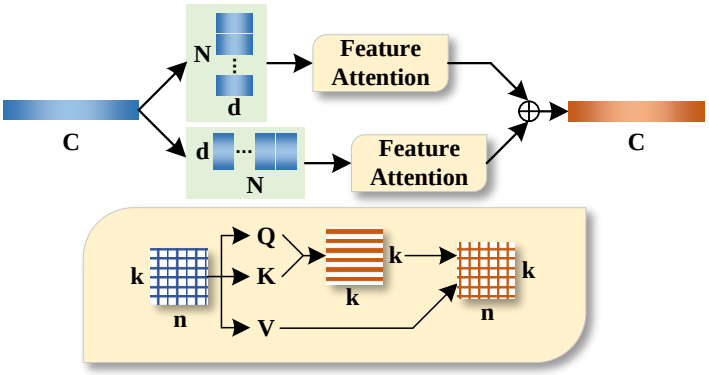


Figure 2. Schematic diagram of the internal structure of channel-wise feature attention.

3.2. Model Training

In this paper, the credit card fraud detection problem is reformulated as an anomaly detection problem since, in real transaction systems, compared with normal transactions, fraudulent transaction records often contain abnormal data values, so this definition is a relatively straightforward way. Utilising credit card transaction data, our goal is to use the proposed anomaly detection framework to perform unsupervised adversarial learning on one of the categories of data (e.g., normal transaction data), and make the model parameters highly fit this category. In the testing phase, when the model encounters transaction data that have not been seen in the training phase (e.g., fraudulent transaction data), the generator G's before and after hidden vector representations z and $\hat{z}$ tend to produce large feature differences. By setting a threshold α, the difference value will be transformed into a class value, 0 (normal transaction data) or 1 (fraud transaction data). We follow a hypothesis that a model with high bias towards normal transaction samples is difficult to generate fraudulent transaction samples. As we only use normal transaction data during the training phase, our model parameters only fit normal transaction data and are not suitable for fraudulent transaction data, resulting in the aforementioned differences.

Specifically, in the data preprocessing stage, the credit card fraud detection dataset T is first split into a training set T_r and a test set T_e. Among them, $T_r = \{x_1, x_2, \cdots, x_m\}$ contains m normal transaction data samples, $T_e = \{\tilde{x_1}, \tilde{x_2}, \cdots, \tilde{x_n}\}$ contains n normal transaction data samples and fraudulent transaction data samples, and its label $y_i \in \{0, 1\}, i = 1, 2, \cdots, n$. In the experiments of this paper, the amount of fraudulent transaction data in Kaggle Credit Card Fraud Detection Dataset and IEEE-CIS Fraud Detection Dataset only accounts for a very small part of the entire dataset, which means that the problem of fraud detection also needs to consider the imbalance of data samples.

In the training phase, the model only performs parameter learning on T_r to train a set of model parameters that are highly biased to the distribution of normal transaction data. In order to make the proposed model better learn the biased network parameters, we propose a hybrid weighted loss function $L_{adv} + \lambda_2 L_{con} + \lambda_3 L_{lat}$ (λ_1, λ_2, λ_3 denote hyperparameters).

- Adversarial Loss L_{adv}: In our framework, the goal of the adversarial loss is to make the samples generated by G as close as possible to the distribution of real normal transaction data samples, so that D cannot accurately distinguish generated samples from real samples. In other words, the adversarial loss is an objective function for adversarial training by maximizing the misjudgment rate of D for generated samples while minimizing the misjudgment rate of G. Its mathematical expression is as follows:

$$L_{adv} = E_{x \sim p_x}[logD(x)] + E_{x \sim p_x}[log(1 - D(\hat{x}))]. \tag{2}$$

- Context Loss L_{con}: In order to make the samples generated by G closer to the original data distribution in terms of eigenvalues to produce more realistic samples, the context loss is introduced into the training phase of the model. It minimizes the L_1 distance between the generated samples and the original normal transaction data samples in the feature space, so that G can preserve the semantic and structural information of the input features as much as possible when generating samples. Its mathematical definition is as follows:

$$L_{con} = E_{x \sim p_x} \|x - \hat{x}\|_1. \tag{3}$$

- Latent Loss L_{lat}: In addition to the above two loss functions, this paper also introduces a Latent Loss. This function ensures that G can produce similar latent space representations by minimizing the L_2 distance of two latent vectors of G in the feature space. In other words, Latent Loss enables G to learn effective encoding methods for normal transaction data from generated samples. In the testing stage, when encountering never-before-seen fraudulent transaction samples, the encoding method of G may fail, resulting in a large feature difference between z and $\hat{z}$. For such sample data, we can

classify it as an abnormal sample (fraudulent transaction sample). The mathematical expression of Latent Loss is as follows:

$$L_{lat} = E_{x \sim p_x} \| z - \hat{z} \|_2. \tag{4}$$

4. Experiments

4.1. Dataset

4.1.1. Credit Card Fraud Detection Dataset

In this section, the credit card fraud detection dataset on Kaggle is used to verify the effectiveness of the proposed UAAD-FDNet. The dataset collects 284,807 credit card transaction records, which are generated by European cardholders within two days in September 2013. Considering data privacy issues, this dataset only provides transaction data processed by PCA. Each transaction record contains a total of 28 principal component values of V1–V28 and the other two unpreprocessed 'Time' and 'Amount' feature. 'Time' represents the time difference between each transaction and the first transaction in the dataset. 'Amount' indicates the amount of each transaction. The schematic diagram of the data sample is shown in Figure 3. In addition, each transaction also contains a set of 'Class' tags: 0 for normal transaction data, 1 for fraudulent transaction data. Among them, there are only 492 fraudulent transactions, which only account for 0.172% of the entire dataset. Figure 4 shows the significant difference in the number of positive and negative samples in this dataset. Therefore, the category imbalance problem should be considered first. The statistical information of the dataset is shown in Table 1.

Time	V1	V2	V3	V4	V5	...	V26	V27	V28	Amount
0	-1.35981	-0.07278	2.536347	1.378155	-0.33832		-0.18911	0.133558	-0.02105	149.62
0	1.191857	0.266151	0.16648	0.448154	0.060018		0.125895	-0.00898	0.014724	2.69
1	-1.35835	-1.34016	1.773209	0.37978	-0.5032		-0.1391	-0.05535	-0.05975	378.66
1	-0.96627	-0.18523	1.792993	-0.86329	-0.01031		-0.22193	0.062723	0.061458	123.5
2	-1.15823	0.877737	1.548718	0.403034	-0.40719		0.502292	0.219422	0.215153	69.99
		⋮						⋮		
160832	0.008812	0.94412	-0.38981	-0.59405	0.738905		0.094063	0.152648	-0.08589	9.51
160832	-2.45901	2.117867	-1.205	-0.62517	-1.48174		0.513479	-0.46243	-0.01536	9.25
160833	-2.11399	1.748864	-1.95475	0.768964	-0.08916		-0.31459	0.770459	0.100563	248.52
160833	-5.26402	5.795819	-5.58939	-0.25467	-0.18698		-0.26523	-0.14674	0.758428	5.9

Unpreprocessed Processed by PCA

Figure 3. The schematic diagram of the data sample.

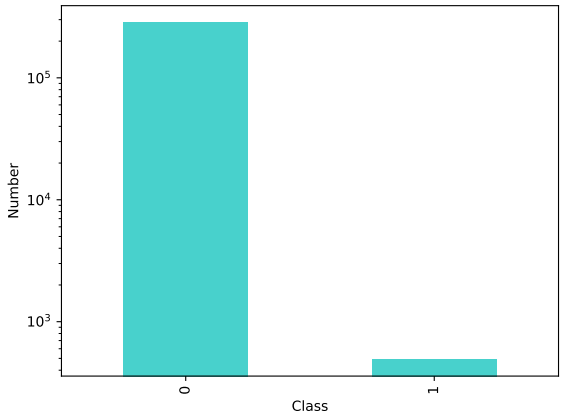


Figure 4. Statistical chart of the number of positive and negative samples in the dataset.

Table 1. Statistical information of credit card fraud detection dataset.

Item	Value
Total Number of Transactions	284,807
Number of Fraudulent Transactions	492
Percentage of Fraudulent Transactions	0.172%
Number of Transaction Data Columns	31
PCA Principal Components Feature Quantity	28
Number of Labels	1

Since the data in the 'Time' and 'Amount' columns have not been preprocessed before model training, we first perform data standardization on them. The mathematical expression of this step is as follows:

$$x_i^{'} = \frac{x_i - \mu}{\sigma} \tag{5}$$

Among them, x_i represents the original eigenvalue, $x_i^{'}$ represents the standardized eigenvalue, μ represents the mean, and σ represents the standard deviation. Through this step, except for the 'Class' column, each transaction record in this dataset contains 30 standardized feature values. Next, we split the dataset into training and testing sets. In view of the serious class imbalance problem in this dataset, the construction of training set and test set needs to be treated carefully. In this paper, the proposed UAAD-FDNet only relies on normal transaction samples for training during the training phase, which is highly consistent with the category imbalance characteristics of the Kaggle credit card fraud detection dataset. In other words, thanks to the unique training method, our method cleverly avoids the disadvantage of data imbalance. In this experiment, we divided 284,315 normal transaction samples into a ratio of 8:2, of which 227,452 normal transaction samples were used as the training set, and the remaining 56,863 normal samples and 492 fraudulent transaction samples constituted the test set. The statistical information of the training set and test set is shown in Table 2.

Table 2. Statistical information of training set and test set.

Split	Training		Test		Total
Class	Normal	Fraud	Normal	Fraud	Both
Number	227,452	0	56,863	492	284,807

4.1.2. IEEE-CIS Fraud Detection Dataset

In addition, the IEEE-CIS fraud detection dataset is used as a generalization dataset to fully validate the robustness of the algorithm proposed in this paper. The IEEE-CIS fraud detection dataset consists of four files, train transaction, train identity, test transaction, and test identity, which contain 394, 41, 393, and 41 columns of features, respectively. Transaction and identity are joined by TransactionID. According to Najadat et al. [37]'s settings, we first concatenate transaction and identity based on TransactionID to obtain 590,540 transaction samples with feature dimension of 433, and then remove unimportant transaction date. Given that 378 features contain a large number of null values, which may have a negative impact on the learning of the model during the training stage, these features are also removed from the experiment in this paper. Similar to the Credit Card Fraud Detection Dataset, this dataset also has a huge difference in terms of the size of positive and negative samples, with fraud samples accounting for about 3.5% of the training set. Therefore, it is necessary to develop an advanced algorithm that can effectively solve the problem of data imbalance for credit card fraud detection. Different from existing supervised learning-based methods, this paper utilizes an ingenious way to avoid the negative impact of data imbalance based on the anomaly detection framework.

4.2. Experimental Setup

All experiments in this article are implemented based on the Pytorch framework, and the model is trained and tested on a GeForce RTX 2080Ti GPU. Adam is used as an optimizer for training to perform gradient updates of model parameters. The initial learning rate is set to 0.001 on Kaggle credit card fraud detection dataset, while 0.01 on IEEE-CIS Fraud Detection Dataset. The batch size is set to 256. The model is trained on two datasets until convergence, with a maximum number of epochs of 100.

In our proposed model, for G_E, the parameters of its five FC layers are denoted as $W_1 \in \mathbb{R}^{(30\times64)}$, $W_2 \in \mathbb{R}^{(64\times128)}$, $W_3 \in \mathbb{R}^{(128\times256)}$, $W_4 \in \mathbb{R}^{(256\times512)}$ and $W_5 \in \mathbb{R}^{(512\times1024)}$, respectively. E and D have similar network structure parameters, while G_D is the opposite. Therefore, the hidden vector $z, \hat{z} \in \mathbb{R}^{1024}$.

In addition, in order to effectively evaluate the effectiveness of the fraud detection framework proposed in this paper, we adopt Precision (PR), Recall (RC), F1-score ($F1$), and Area under the receiver operating characteristic curve (AUC) as the model evaluation indicators for this experiment. The mathematical formulas are as follows

$$PR = \frac{TP}{TP + FP}, \tag{6}$$

$$RC = \frac{TP}{TP + FN}, \tag{7}$$

$$F1 = \frac{2 \times PR \times RC}{PR + RC}, \tag{8}$$

$$Sensitivity = RC, \tag{9}$$

$$Specificity = \frac{TN}{TN + FP}, \tag{10}$$

where TP represents True Positive. FP represents False Positive. TN represents True Negative. Furthermore, FN represents False Negative. AUC is the area under the curve composed of 1-Specificity and Sensitivity in the horizontal and vertical coordinates. From Equations (9) and (10), it can be observed that a larger AUC indicates better performance of the model.

4.3. Threshold α Setting

In this paper, the credit card fraud detection problem is reformulated as an anomaly detection problem. In the testing phase, we discriminate abnormal samples (fraudulent transactions) by calculating the characteristic distance between x and $\hat{x}$ and z and $\hat{z}$. The following principles are followed:

$$Score = \beta_1 d_{con} + \beta_2 d_{lat} \tag{11}$$

Among them, d_{con} represents the L_1 distance between x and $\hat{x}$, d_{lat} represents the L_2 distance between z and $\hat{z}$, β_1 and β_2 are two hyperparameters, and $Score$ represents the fraud score. In order to restrict this formula, we let $\beta_1 + \beta_2 = 1$. So Formula (5) can be expressed as follows:

$$Score = \beta d_{con} + (1 - \beta) d_{lat} \tag{12}$$

We only need to adjust the parameter β to set a fraud score that conforms to the two dataset in this paper. In the experiment, β is set to 0.5. When $Score \geq \alpha$, the transaction sample is judged as a fraudulent transaction; when $Score < \alpha$, it is regarded as a normal transaction. To determine the threshold α, we visualize the kernel density curve of the fraud score of Kaggle Credit Card Fraud Detection Dataset, as shown in Figure 5. The red curve in the figure represents normal transactions, and the blue curve represents fraudulent

transactions. The choice of threshold α should be at the intersection of the two curves, so in this experiment, α is set to 0.4. The same applies to the IEEE-CIS Fraud Detection Dataset.

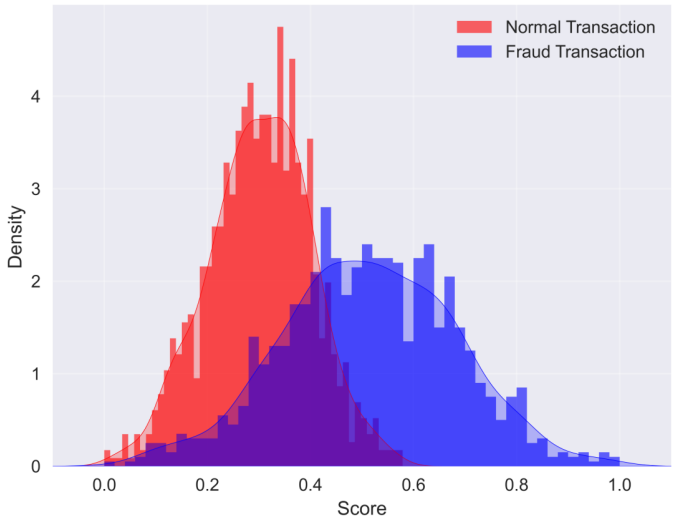


Figure 5. Kernel density curve of fraud score on Kaggle Credit Card Fraud Detection Dataset.

4.4. Model Comparison Experiment

4.4.1. Comparative Experiment on Kaggle Credit Card Fraud Detection Dataset

In this experiment, we compare our proposed UAAD-FDNet with existing traditional machine learning methods and deep learning methods on Kaggle Credit Card Fraud Detection Dataset to demonstrate its effectiveness. The experimental results are shown in Table 3.

Support Vector Machine (SVM) [38], Decision Tree (DT) [39], Extreme Gradient Boosting (XG Boost) [40], K-Nearest Neighbor (KNN) [41], and Random Forest (RF) [42] are several typical machine learning algorithms. Long Short-Term Memory (LSTM) [33], Convolutional Neural Network (CNN) [43], MultiLayer Perceptron (MLP) [44], and AutoEncoder (AE) [45] are four commonly used deep learning methods. From the table, we can find that the performance of machine learning methods on fraud detection tasks is far worse than that of deep learning methods. This may be due to the fact that artificial feature engineering often has difficult in fully modeling the internal relationship between different feature attributes of financial transaction data, meaning that it is difficult for machine learning methods to accurately determine the decision boundary. The deep learning method avoids the process of manually constructing feature engineering, and it can effectively use parameter learning to automatically capture the interdependence between high-dimensional features. Compared with machine learning methods, it has stronger feature learning capabilities and general ability. Among the four deep learning methods, the classification index of LSTM is relatively low. Compared with LSTM, AE has improved by 0.0455/0.0104/0.0244/0.0434 in four indicators. It uses the data x itself as a supervisory signal to guide the training of the neural network, and can learn more compact data representation. From the data, we can see that the fraud detection performance of AE is significantly better than that of LSTM, CNN, and MLP. The UAAD-FDNet proposed in this paper takes advantage of the powerful feature learning ability of AE and the data generation advantages of GAN to fully learn the normal transaction data distribution in an unsupervised learning manner. In the test phase, the detection of fraudulent transactions (abnormal samples) is completed by calculating the hidden vectors z and $\hat{z}$ and the feature distance between the input x and the reconstructed sample $\hat{x}$. Compared with other methods, the method proposed in this paper represents a new training-inference paradigm. In this experiment, we compare UAAD-FDNet with and without feature attention module. The experimental results in the table show that the proposed method has more robust fraud detection performance. Without the help of FA,

our method outperforms AE by 0.0228/0.0019/0.0099/0.0158 on four evaluation metrics, respectively. After introducing FA, the overall performance of the model is significantly better than other existing fraud detection methods. This fully demonstrates the advancement and effectiveness of the UAAD-FDNet proposed in this paper.

Table 3. Comparative experimental results on Kaggle Credit Card Fraud Detection Dataset. (Red bold indicates optimal results. Blue bold indicates suboptimal results).

Method	Model	PR	RC	F1	AUC
Machine Learning	SVM	0.8854	0.7215	0.7951	0.8586
	DT	0.8837	0.7269	0.7977	0.8598
	XG Boost	0.8955	0.7280	0.8031	0.8649
	KNN	0.9032	0.7268	0.8055	0.8709
	RF	0.9112	0.7343	0.8132	0.8827
Deep Learning	LSTM	0.9073	0.7391	0.8146	0.8845
	CNN	0.9217	0.7453	0.8242	0.9075
	MLP	0.9262	0.7461	0.8265	0.9094
	AE	0.9528	0.7495	0.8390	0.9279
	UAAD-FDNet w/o FA (Ours)	**0.9756**	**0.7514**	**0.8489**	**0.9437**
	UAAD-FDNet w/ FA (Ours)	**0.9795**	**0.7553**	**0.8529**	**0.9515**

4.4.2. Comparative Experiment on IEEE-CIS Fraud Detection Dataset

Given that it is difficult to convincingly demonstrate the advantages of the proposed method on a single dataset, we conduct another set of experiments on IEEE-CIS Fraud Detection Dataset.

Table 4 shows the specific results of our experiment. As we analyze above, traditional machine learning methods are generally inferior to deep learning methods. However, the XG Boost achieves the best experimental results and even outperforms some deep learning methods in some indicators. This indicates that in certain specific applications, machine learning methods are still a powerful tool. Compared to these methods, our method can effectively improve performance without using feature attention. This implies that the fraud detection method based on the anomaly detection framework proposed in this paper can fully utilize the normal transaction data for highly biased learning, enabling abnormal samples (fraud samples) to be effectively distinguished. The feature attention module can optimize the feature expression of the model, thereby suppressing adverse effects caused by missing or incorrect data.

Table 4. Comparative experimental results on IEEE-CIS Fraud Detection Dataset. (Red bold indicates optimal results. Blue bold indicates suboptimal results).

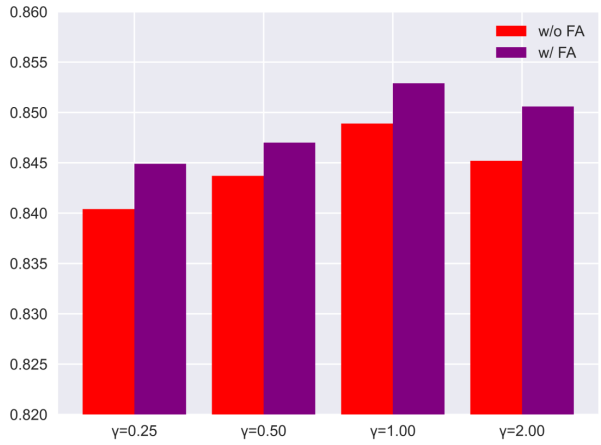
Method	Model	PR	RC	F1	AUC
Machine Learning	SVM	0.9091	0.1906	0.3151	0.5783
	DT	0.5206	0.5470	0.5335	0.7622
	XG Boost	**0.9447**	0.5915	0.7275	0.7892
	KNN	0.8358	0.3711	0.5140	0.6730
	RF	**0.9713**	0.5024	0.6623	0.7405
Deep Learning	LSTM	0.8525	0.5854	0.6941	0.7802
	CNN	0.8779	0.5952	0.7094	0.7837
	MLP	0.9159	0.5796	0.7099	0.8241
	AE	0.9055	0.5873	0.7125	0.8181
	UAAD-FDNet w/o FA (Ours)	0.9415	**0.6027**	**0.7349**	**0.8390**
	UAAD-FDNet w/ FA (Ours)	0.9337	**0.6281**	**0.7510**	**0.8556**

4.5. Model Ablation Experiment

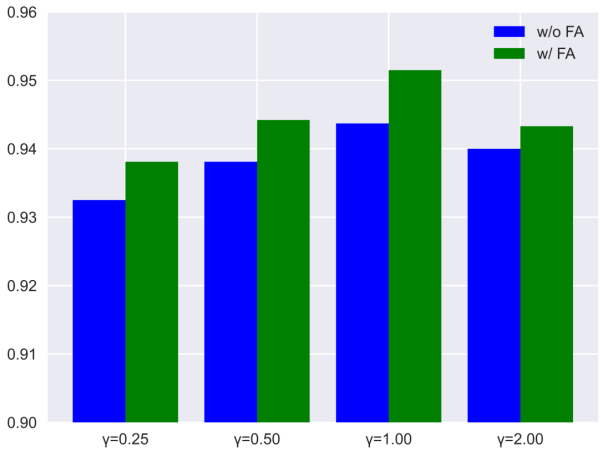
In this section, we explore the impact of channel width and loss function on several classification indicators on Kaggle Credit Card Fraud Detection Dataset. The specific content of the experiment is as follows.

4.5.1. Ablation for Channel

In this section, we explore the effect of the model's channel width on the final fraud detection metrics. We assign a weight factor γ to each layer parameter of the network to adjust the channel ratio of the entire model. In this experiment, γ is set to 0.25, 0.5, 1, and 2, respectively, and four UAAD-FDNets with different widths are constructed accordingly. Figure 6 shows the test results of UAAD-FDNet with FA and without FA on the Kaggle credit card fraud detection dataset, where the F1 indicator is reported in (a), and the AUC indicator is reported in (b). It can be seen from the figure that as γ continues to increase, the fraud detection performance of the model is also continuously improving, because the increase in the number of model parameters enables it to fit more complex data distributions. When γ increases to a certain extent, the performance of the model decreases, which may be because the increase in the number of parameters brings the risk of overfitting to the model, which degrades its performance on the test set. Therefore, when γ increases, strategies such as dropout and regularization to suppress overfitting should be considered. In this experiment, $\gamma = 1$ is the best choice.



(**a**) F1 is reported



(**b**) AUC is reported

Figure 6. Ablation experiment results of UAAD-FDNet with different γ factors on Kaggle Credit Card Fraud Detection Dataset. (**a**) shows the comparison of F1 indicators, and (**b**) shows the comparison of AUC indicators.

4.5.2. Ablation for Loss Function

In order to verify the effectiveness of the joint loss composed of three loss functions proposed in this paper, we conduct sufficient ablation experiments. In order to ensure the training paradigm of the GAN, the adversarial loss L_{adv} is always preserved during the experimental process, as this is crucial for the normal training of the entire network. Context loss L_{con} and latent loss L_{lat} are removed from the training process, respectively, to verify their important contributions to the UAAD-FDNet proposed in this paper. Table 5 shows the experimental results on the Kaggle credit card fraud detection dataset. From the table, we can intuitively see that by removing L_{con} and L_{lat}, respectively, the overall network experienced significant performance degradation on all four evaluation indicators. This indicates that during the training stage, both are crucial for the network to learn the distribution of normal transaction samples. If the two loss functions are removed at the same time, the network's performance to fraud detection is even worse than that of traditional machine learning methods. This is because there is a lack of constraints on the reconstructed sample $\hat{x}$ and hidden vector z,$\hat{z}$ in the generator G during the training stage, which results in untrustworthy fraud scores when calculating according to Equation (12), seriously hindering the network from identifying fraudulent transactions. If three different loss functions are used to supervise the network at the same time, the method proposed in this paper can achieve significant improvement with regard to four indicators, which fully proves the effectiveness and criticality of the proposed joint loss function for this task.

Table 5. The ablation experiment results of the loss function on the Kaggle Credit Card Fraud Detection Dataset. (Red bold indicates the optimal result.)

L_{adv}	L_{con}	L_{lat}	PR	RC	F1	AUC
✓			0.7532	0.6451	0.6950	0.7443
✓	✓		0.9088	0.7306	0.8100	0.8769
✓		✓	0.9152	0.7375	0.8168	0.8964
✓	✓	✓	**0.9795**	**0.7553**	**0.8529**	**0.9515**

5. Conclusions

In this paper, we reformulate the credit card fraud detection problem as an anomaly detection problem, and propose a new unsupervised attentional anomaly detection-based credit card fraud detection network (UAAD-FDNet). The network mainly consists of a generator and a discriminator. Among them, the generator uses the autoencoder with Feature Attention to reconstruct the input transaction samples to generate as real transaction data as possible, in this way, it can learn the high-level representation (hidden vector) of normal transaction data. The discriminator is used to form an adversarial training mode with the generator during the training phase to better guide the generator to fit the normal transaction data distribution. Compared with traditional machine learning methods, such as SVM, DT, XG Boost, KNN, and RF, as well as existing deep learning-based methods, such as LSTM, CNN, MLP, and AE, our method has stronger generalization. The experimental results on Kaggle Credit Card Fraud Detection Dataset and IEEE-CIS Fraud Detection Dataset show that the proposed method can effectively avoid the problem of data imbalance, and its fraud detection performance is better. This indicates that, in real scenarios, our method can safeguard the interests of financial users well.

Author Contributions: Conceptualization, S.J. and J.W.; methodology, S.J.; software, S.J.; validation, J.W., M.X., and R.D.; formal analysis, R.D.; investigation, J.W. and R.D.; resources, S.J.; data curation, J.W.; writing—original draft preparation, S.J.; writing—review and editing, J.W.; visualization, J.W.; supervision, M.X.; project administration, S.J.; funding acquisition, S.J. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported in part by the National Natural Science Foundation of PR China (72101121) and Ministry of Education, Humanities and social science projects (21YJC790054).

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Haoxiang, W.; Smys, S. Overview of configuring adaptive activation functions for deep neural networks-a comparative study. *J. Ubiquitous Comput. Commun. Technol.* **2021**, *3*, 10–22.
2. Zhang, R.; Zheng, F.; Min, W. Sequential behavioral data processing using deep learning and the Markov transition field in online fraud detection. *arXiv* **2018**, arXiv:1808.05329.
3. Sun, W.; Yang, C.G.; Qi, J.X. Credit risk assessment in commercial banks based on support vector machines. In Proceedings of the 2006 International Conference on Machine Learning and Cybernetics, Dalian, China, 13–16 August 2006; pp. 2430–2433.
4. Smys, S.; Raj, J.S. Analysis of deep learning techniques for early detection of depression on social media network-a comparative study. *J. Trends Comput. Sci. Smart Technol.* **2021**, *3*, 24–39.
5. Thennakoon, A.; Bhagyani, C.; Premadasa, S.; Mihiranga, S.; Kuruwitaarachchi, N. Real-time credit card fraud detection using machine learning. In Proceedings of the 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 10–11 January 2019; pp. 488–493.
6. Sailusha, R.; Gnaneswar, V.; Ramesh, R.; Rao, G.R. Credit card fraud detection using machine learning. In Proceedings of the 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 13–15 May 2020; pp. 1264–1270.
7. Rtayli, N.; Enneya, N. Enhanced credit card fraud detection based on SVM-recursive feature elimination and hyper-parameters optimization. *J. Inf. Secur. Appl.* **2020**, *55*, 102596.
8. Ileberi, E.; Sun, Y.; Wang, Z. A machine learning based credit card fraud detection using the GA algorithm for feature selection. *J. Big Data* **2022**, *9*, 24.
9. Kim, E.; Lee, J.; Shin, H.; Yang, H.; Cho, S.; Nam, S.k.; Song, Y.; Yoon, J.A.; Kim, J.I. Champion-challenger analysis for credit card fraud detection: Hybrid ensemble and deep learning. *Expert Syst. Appl.* **2019**, *128*, 214–224.
10. Maniraj, S.; Saini, A.; Ahmed, S.; Sarkar, S. Credit card fraud detection using machine learning and data science. *Int. J. Eng. Res.* **2019**, *8*, 110–115.
11. Tiwari, P.; Mehta, S.; Sakhuja, N.; Kumar, J.; Singh, A.K. Credit card fraud detection using machine learning: A study. *arXiv* **2021**, arXiv:2108.10005.
12. Eckerli, F.; Osterrieder, J. Generative adversarial networks in finance: An overview. *arXiv* **2021**, arXiv:2106.06364.
13. Zou, J.; Zhang, J.; Jiang, P. Credit card fraud detection using autoencoder neural network. *arXiv* **2019**, arXiv:1908.11553.
14. Liu, X.; Zhang, F.; Hou, Z.; Mian, L.; Wang, Z.; Zhang, J.; Tang, J. Self-supervised learning: Generative or contrastive. *IEEE Trans. Knowl. Data Eng.* **2021**, *35*, 857–876.
15. Albahli, S.; Nazir, T.; Mehmood, A.; Irtaza, A.; Alkhalifah, A.; Albattah, W. AEI-DNET: A novel densenet model with an autoencoder for the stock market predictions using stock technical indicators. *Electronics* **2022**, *11*, 611.
16. Chen, R.C.; Chen, T.S.; Lin, C.C. A new binary support vector system for increasing detection rate of credit card fraud. *Int. J. Pattern Recognit. Artif. Intell.* **2006**, *20*, 227–239.
17. Khan, A.; Singh, T.; Sinhal, A. Implement credit card fraudulent detection system using observation probabilistic in hidden markov model. In Proceedings of the 2012 Nirma University International Conference on Engineering (NUiCONE), Ahmedabad, India, 6–8 December 2012; pp. 1–6.
18. Zareapoor, M.; Shamsolmoali, P. Application of credit card fraud detection: Based on bagging ensemble classifier. *Procedia Comput. Sci.* **2015**, *48*, 679–685.
19. Yee, O.S.; Sagadevan, S.; Malim, N.H.A.H. Credit card fraud detection using machine learning as data mining technique. *J. Telecommun. Electron. Comput. Eng.* **2018**, *10*, 23–27.
20. Lu, C.; Xia, M.; Lin, H. Multi-scale strip pooling feature aggregation network for cloud and cloud shadow segmentation. *Neural Comput. Appl.* **2022**, *34*, 6149–6162.
21. Qu, Y.; Xia, M.; Zhang, Y. Strip pooling channel spatial attention network for the segmentation of cloud and cloud shadow. *Comput. Geosci.* **2021**, *157*, 104940.
22. Wang, Z.; Xia, M.; Lu, M.; Pan, L.; Liu, J. Parameter Identification in Power Transmission Systems Based on Graph Convolution Network. *IEEE Trans. Power Deliv.* **2022**, *37*, 3155–3163.
23. Chen, J.; Xia, M.; Wang, D.; Lin, H. Double Branch Parallel Network for Segmentation of Buildings and Waters in Remote Sensing Images. *Remote Sens.* **2023**, *15*, 1536.
24. Zhang, C.; Weng, L.; Ding, L.; Xia, M.; Lin, H. CRSNet: Cloud and Cloud Shadow Refinement Segmentation Networks for Remote Sensing Imagery. *Remote Sens.* **2023**, *15*, 1664.
25. Ma, Z.; Xia, M.; Lin, H.; Qian, M.; Zhang, Y. FENet: Feature enhancement network for land cover classification. *Int. J. Remote Sens.* **2023**, *44*, 1702–1725.
26. Wang, D.; Weng, L.; Xia, M.; Lin, H. MBCNet: Multi-Branch Collaborative Change-Detection Network Based on Siamese Structure. *Remote Sens.* **2023**, *15*, 2237.

27. Fu, K.; Cheng, D.; Tu, Y.; Zhang, L. Credit card fraud detection using convolutional neural networks. In *Proceedings of the Neural Information Processing: 23rd International Conference, ICONIP 2016, Kyoto, Japan, 16–21 October 2016*; Proceedings, Part III 23; Springer: Berlin/Heidelberg, Germany, 2016; pp. 483–490.
28. Chouiekh, A.; Haj, E.H.I.E. Convnets for fraud detection analysis. *Procedia Comput. Sci.* **2018**, *127*, 133–138.
29. Saia, R.; Carta, S. Evaluating Credit Card Transactions in the Frequency Domain for a Proactive Fraud Detection Approach. In Proceedings of the SECRYPT, Madrid, Spain, 26–28 July 2017; pp. 335–342.
30. Fiore, U.; De Santis, A.; Perla, F.; Zanetti, P.; Palmieri, F. Using generative adversarial networks for improving classification effectiveness in credit card fraud detection. *Inf. Sci.* **2019**, *479*, 448–455.
31. Saia, R.; Carta, S. Evaluating the benefits of using proactive transformed-domain-based techniques in fraud detection tasks. *Future Gener. Comput. Syst.* **2019**, *93*, 18–32.
32. Esenogho, E.; Mienye, I.D.; Swart, T.G.; Aruleba, K.; Obaido, G. A neural network ensemble with feature engineering for improved credit card fraud detection. *IEEE Access* **2022**, *10*, 16400–16407.
33. Mohmad, Y.A. Credit Card Fraud Detection Using LSTM Algorithm. *Wasit J. Comput. Math. Sci.* **2022**, *1*, 39–53.
34. Schapire, R.E. A brief introduction to boosting. In Proceedings of the Ijcai, Stockholm, Sweden, 31 July–6 August 1999; Volume 99, pp. 1401–1406.
35. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 6000–6010.
36. Zhang, S.; Wang, L. STPGTN—A Multi-Branch Parameters Identification Method Considering Spatial Constraints and Transient Measurement Data. *Comput. Model. Eng. Sci.* **2023**, *136*, 2635–2654.
37. Najadat, H.; Altiti, O.; Aqouleh, A.A.; Younes, M. Credit card fraud detection based on machine and deep learning. In Proceedings of the 2020 11th International Conference on Information and Communication Systems (ICICS), Irbid, Jordan, 7–9 April 2020; pp. 204–208.
38. Hearst, M.A.; Dumais, S.T.; Osuna, E.; Platt, J.; Scholkopf, B. Support vector machines. *IEEE Intell. Syst. Their Appl.* **1998**, *13*, 18–28.
39. Quinlan, J.R. *C4. 5: Programs for Machine Learning*; Elsevier: Amsterdam, The Netherlands, 2014.
40. Meng, C.; Zhou, L.; Liu, B. A case study in credit fraud detection with SMOTE and XGboost. *J. Phys. Conf. Ser.* **2020**, *1601*, 052016.
41. Cover, T.; Hart, P. Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* **1967**, *13*, 21–27.
42. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32.
43. Chen, J.I.Z.; Lai, K.L. Deep convolution neural network model for credit-card fraud detection and alert. *J. Artif. Intell.* **2021**, *3*, 101–112.
44. Kasasbeh, B.; Aldabaybah, B.; Ahmad, H. Multilayer perceptron artificial neural networks-based model for credit card fraud detection. *Indones. J. Electr. Eng. Comput. Sci.* **2022**, *26*, 362–373.
45. Fanai, H.; Abbasimehr, H. A novel combined approach based on deep Autoencoder and deep classifiers for credit card fraud detection. *Expert Syst. Appl.* **2023**, *217*, 119562.