A New Cybersecurity Attack Fraud Detection Model using Machine Learning

Muhammet Çakmak[1*], Zafer Albayrak[2], Hakan Can Altunay[3]

[1]Department of Computer Engineering, Giresun University, 28100 Giresun, Türkiye; E-mail: muhammet.cakmak@giresun.edu.tr
* Corresponding author

[2]Department of Computer Engineering, Sakarya University of Applied Sciences, 54100 Sakarya, Turkey; E-mail: zaferalbayrak@subu.edu.tr

[3]Department of Computer Technologies, Carsamba Chamber of Commerce Vocational School, Ondokuz Mayis University, 55200 Samsun, Turkey
E-mail: hakancan.altunay@omu.edu.tr

Abstract: The rapid growth in e-commerce and online payment systems has led to an increase in cyberfraud activities and serious financial losses for commercial enterprises. Solving this critical problem requires the development of effective fraud prevention mechanisms and the investment of online shopping platforms in cybersecurity systems. Fraud detection is usually based on binary classification methods, but reducing the size of high-dimensional datasets is of great importance to detect fraudulent activities with high accuracy. This study proposes a two-stage hybrid framework to prevent credit card fraud. In the first stage, the dimensionality in the dataset is reduced using Autoencoder (AE), and the low-dimensional data obtained from the fully connected layer of the Autoencoder is presented as input data to the Random Forest (RF) classifier in the second stage. The proposed AE+RF model is compared with popular models such as Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN) and XgBoost on the basic performance metrics of precision, recall, F1-Score, AUC-ROC and AUC-PR. Experimental results show that the AE+RF model outperforms its closest competitors by 2.11% in recall, 1.03% in F1-Score, 1.02% in AUC-ROC and 2.08% in AUC-PR. These findings reveal that the proposed framework makes a significant contribution to enhancing the security of e-commerce platforms by providing high accuracy and efficiency in cybersecurity fraud detection.

Keywords: Cybersecurity; Deep Learning; AutoEncoder; Random Forest

1 Introduction

In recent years, with the rapidly developing e-commerce applications, credit card usage and fraud is increasing. The methods used by fraudsters are becoming more complex every day. Unauthorized purchases with credit and debit cards account for

10-15% of total fraud [1]. According to a European Central Bank report, the total value of fraudulent card transactions in the Eurozone amounted to €1.03 billion [2]. Due to the increasing credit card fraud, all small- and large-scale public and private businesses are researching and investing to prevent this possible fraud. Fraud detection systems are an effective way to prevent attacks. Fraud detection systems are vital for card users, business owners and credit card manufacturers.

Like traditional fraud detection systems, big data and artificial intelligence present significant opportunities for preventing fraud. These AI-based systems classify transactions in a dataset as either normal or fraudulent using binary classification methods. The model created to detect fraud classifies transactions as normal or fraudulent based on a data set. However, detecting fraud in the classification process poses various difficulties. The first problem is that the data sets are severely unbalanced due to the small number of fraudulent transactions [3]. Secondly, misclassifying normal and fraudulent transactions significantly impacts the cost of solving the problem [4]. Third, the temporal dependence between normal and fraudulent transactions is high [5]. Fourth, Class-dependent situations change over time. That is, the existence of concept drift occurs [6]. Finally, the problem is that the search preprocessor in the detection process should be high-dimensional [7].

To date, classification and regression problems based on machine learning methods such as Random Forest (RF), Support Vector Machine (SVM), Gradient Boosting (GB), Artificial Neural Network (ANN), Decision Tree (DT) have been solved in a wide variety of fields [8]. In addition, deep learning models such as Recurrent neural networks (RNN), Convolution Neural Networks (CNN), Generative Adversarial Networks (GAN) and Long short-term memory (LSTM) have been used to solve problems involving large amounts of data [9].

Since artificial intelligence models can make very good output predictions from input data, they achieve highly successful results in many areas. Machine learning and deep learning methods have been used in many areas, such as image processing, translation, speech recognition, time series, recommendation system, anomaly detection, healthcare application, and finance. Although machine and deep learning models provide successful results in many areas, they are not widely used in credit card fraud detection.

In order to detect cyber fraud accurately, it is critical to use input data with a lower representation in the imbalanced data set [10]. Autoencoders are a representation learning method that effectively reduces input sizes in data sets where the data input size is large [11]. The structure of an autoencoder consists of an encoder and a decoder. Encoder module at the input stage of an input is used to compress the data, and the decoder module at the output stage is used to reconstruct the compressed examples [12], [13].

This study presents a hybrid framework that combines an autoencoder for dimension reduction and machine learning for classification. The proposed model is presented by combining Autoencoder and random forest approaches. In the

combined model, we first represent the features with the Autoencoder and use these features as input for Random Forest (RF). We also used well-known Recurrent Neural Network (RNN), Artificial Neural Network (ANN), Convolutional Neural Network (CNN), and Xgboost models to compare the performance of the proposed model.

The contributions of the proposed study are presented as follows:

1) Using deep Autoencoder to obtain a better representation of values in a critical dataset.
2) We are combining deep learning and machine learning to detect credit card fraud.
3) Comparison of the results obtained from the Deep Auto Encoder and the PCA model in terms of performance.
4) Conducting comprehensive experimental studies to determine the best model.

Section 2 gives information and literature studies about credit card fraud and detection methods. In Section 3, the subject of Autoencoders is explained. Section 4 shows the experimental study results. Section 5 contains the conclusion of the article.

2 Related Work

The main use of a credit card detection system is the detection of credit card fraud by analyzing the previous fraudulent transactions placed in the dataset. It develops a credit card detection model identifying transactions that might be fraudulent. Machine learning and deep learning models have been used in the last few years [14]. Recent research demonstrates that fraud detection has increasingly relied on advanced artificial intelligence methods, including Artificial Neural Networks (ANN), LightGBM, k-nearest neighbors (kNN), AdaBoost, Support Vector Machines (SVM), Genetic Algorithms (GA), Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), Convolutional Neural Networks (CNN), Deep Neural Networks (DNN), and Generative Adversarial Networks (GAN), to address the complex and highly imbalanced nature of credit card transaction data [15].

In Table 1 present the performance of the other models for credit card fraud detection model such as Convolutional Autoencoders (CAE), Stacked Autoencoders (SAE), and Variational Autoencoders (VAE), Neighborhood Components Analysis (NCA) [16], Principal Component Analysis (PCA) [17], and Convolutional Autoencoder (CAE) [18]. In addition, autoencoders are used in areas such as various health data [19], anomaly detection [20], image classification [21] and voice recognition [22] to reduce data sizes.

Table 1
Comparative analysis of recent credit card fraud detection methods

Contribution	Methods	Dimension Reduction	Year	References
The study proposes a GAN-based data augmentation framework to handle extreme class imbalance in credit card fraud detection, improving recall and AUC-PR.	GAN + ANN	No	2023	[23]
This work applies a convolutional autoencoder to reduce feature dimensionality before classification using ensemble learning methods.	CAE + RF, XGBoost	CAE	2023	[24]
The proposed approach integrates PCA with gradient boosting models to improve fraud detection performance on highly imbalanced datasets.	PCA + LightGBM	PCA	2023	[25]
A hybrid deep learning model combining GAN-generated samples with BiLSTM is introduced to capture temporal fraud patterns.	GAN + BiLSTM	No	2024	[26]
This study employs a stacked autoencoder for feature learning followed by CatBoost for fraud classification, achieving improved recall and F1-score.	SAE + CatBoost	SAE	2024	[27]
The authors propose a transformer-based model for sequential transaction modeling, demonstrating superior performance compared to RNN-based methods.	Transformer	No	2024	[28]
A hybrid framework combining variational autoencoders and ensemble classifiers is introduced to address data imbalance and feature redundancy.	VAE + RF, XGBoost	VAE	2024	[29]
The study presents a GAN–RNN-based fraud detection system that adapts to evolving fraud behaviors and significantly improves minority-class detection.	GAN + GRU	No	2024	[30]
The study presents a hybrid deep learning framework combining GANs with RNNs to enhance credit card fraud detection by addressing data imbalance and evolving fraud patterns. The GAN-GRU model achieves high sensitivity (0.992) and specificity (1.000) on the European credit card dataset, outperforming traditional methods.	Hybrid GANs+RNNs	No	2024	[31]
This work introduces a lightweight deep learning model optimized for real-time credit card fraud detection with reduced computational cost.	CNN + Attention	No	2025	[32]

Motivated by these observations, the present study builds upon the demonstrated effectiveness of autoencoder-based feature learning and ensemble classification. Unlike previous works, this study employs an autoencoder to obtain representative low-dimensional embeddings of transaction data and subsequently develops a hybrid machine learning and deep learning framework that operates on this reduced feature space. By doing so, the proposed approach aims to improve detection performance while maintaining computational efficiency, thereby addressing both feature redundancy and class imbalance challenges in credit card fraud detection.

3 Proposed Model

In this section, the proposed hybrid framework is presented. First, the architecture of the autoencoder is described, highlighting its role in learning compact latent representations for dimensionality reduction. Next, the encoder's learned features are used as input to the Random Forest classifier to perform fraud detection. Finally, the experimental setup and results are reported to demonstrate the effectiveness of the proposed approach.

Autoencoder is a technique used for dimensionality reduction in a dataset [32]. The system is composed of two main components: an encoder and a decoder. The encoder transforms input data into a lower-dimensional representation capturing essential properties with fewer parameters [33]. On the other hand, the decoder takes the lower-dimensional data and converts it back to the original dimension, attempting to recreate the original input data.

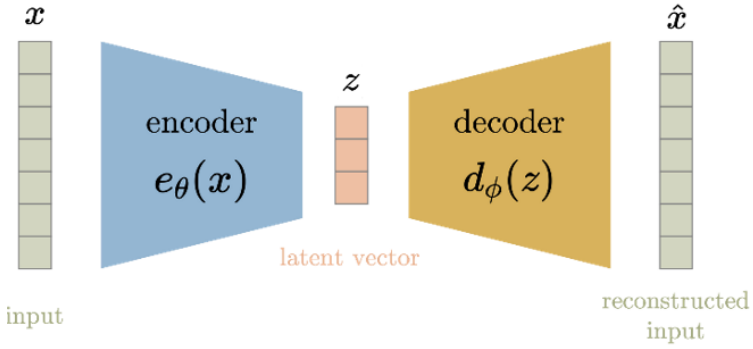

Figure 1
Schematic representation of Autoencoder structure

Autoencoders have various applications, such as feature extraction, noise reduction, and unsupervised learning. An input x is part of a space that has m dimensions. The input data x is transformed by the encoder into a k-dimensional projection z. The first step is to generate a latent representation from the input x. The decoder module uses the latent representation to generate an output that matches the dimensions of input x. The Autoencoder typically consists of three layers: the input layer, the hidden layer, and the output layer. The cost function calculates the difference between the input x and reconstruction $x'$and is expressed as Eq. 1. Various cost functions can be used in autoencoders. Since the dataset was normalized between 0 and 1, a sigmoid activation function was employed in the output layer.

$$L(x, x') = \frac{1}{2}\sum_i (\mathrm{x}_i - \mathrm{x'}_i)^2 \tag{1}$$

3.1 Proposed Framework

To achieve better classification performance, creating a more meaningful representation of input data is crucial before classification. Due to the growing volume of data, representation learning is being widely employed in various fields [34]. The proposed model involves three stages: data reduction using an autoencoder, connected layer generation, and classification via random forest. The proposed model was compared to well-known machine and deep learning models in terms of performance. Here's a step-by-step explanation of the proposed model's operation:

- The Autoencoder is applied to the original credit card fraud dataset.
- Using an auto-encoder, representative, small-dimensioned, and preprocessed new dataset is obtained from the original dataset.
- Fully connected layers in the autoencoder model are obtained.
- Fully connected layers are used as input for the RF classifier.
- The results of the proposed AE+RF and other well-known classifiers are compared.

The experimental study is summarized in Figure 2, which presents a flow chart. To begin with, we split the original dataset into two sections- training and testing.

After that, we set up all the parameters and inputted the input layer X into the Autoencoder. The Autoencoder then cleaned the features and reduced the dimensions, resulting in the output layer X'. The Autoencoder's hidden layer data is transmitted to the tree nodes via fully connected layers. The nodes in the tree layer of the RF are forwarded to the nodes *(k)* in the decision trees. Decision nodes make a decision by taking an average of the decision results. All parameters are then updated to minimize the cost value between actual and prediction. The entire process repeats iteratively until the optimum result is achieved. The costs that may arise from misclassifying normal and fraudulent transactions must be reduced. To address this issue, we have employed a two-stage framework that involves using an Autoencoder to reduce dimensionality in the first stage. We then feed this dimensionally reduced data into a Random Forest model for further analysis. Our hybrid AE+RF model includes parameter tuning and hyperparameter optimization in the Random Forest component to tackle the problem of misclassification costs. By adjusting the decision threshold and fine-tuning the model parameters, we aim to strike a balance between minimizing false positives and false negatives. Our methodology is designed to improve the overall performance of fraud detection by considering the subtle costs associated with misclassifications.

In the $X = \epsilon\, R^n$ formula, X is the input feature vector for the Autoencoder and n is the feature number of the Autoencoder. H in Eq. 2 indicates the hidden layer of autoencoders. W_E is weight and b_E is defined as bias in the encoder part.

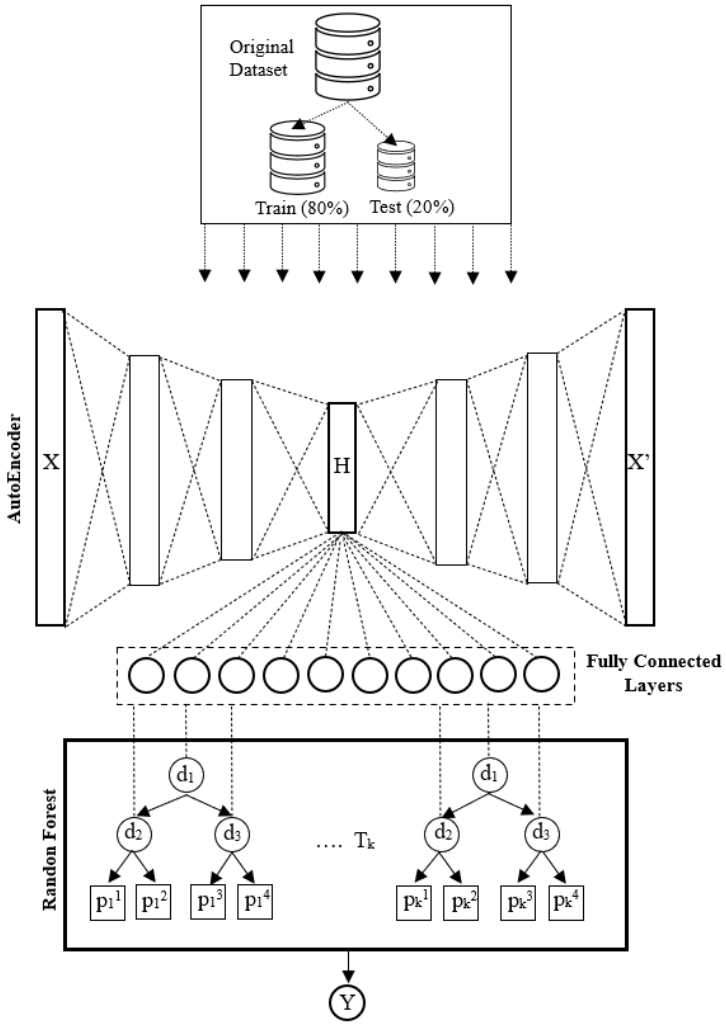

Figure 2
Flow chart of the proposed model

$$H = f_E(W_E X + b_E) \tag{2}$$

By restructuring the hidden layer, the output value X' is obtained. The formula of the output structure is given in Eq. 3. W_D is weight and b_D is the bias value used on the decoder side. The X data from H is optimized to reach X'. Purpose of optimization is to minimize the average reconstruction error.

$$X' = f_D(W_D H + b_D) \tag{3}$$

The measurement formula with conventional squared error is given in Eq. 4.

$$L_{AE}(X, X') = \|X - X'\|^2 \tag{4}$$

H passing through the fully connected layer forms the input nodes for the random forest. The formula for X_T obtained in Eq. 5 is shown. W_F is weight and b_F is the bias value used in the fully connected layers. $g(*)$ is activation function for the model.

$$X_T = g(W_F H + b_F) \tag{5}$$

Decision function $D_d\,(X_T; \Theta) \epsilon\, [0,1]$ means that the decision node reaches d and is sent to the left subtree. Main decision function is shown in Eq. 6.

$$D_d\,(X_T; \Theta) = \sigma(f_d\ (X_T\ )) \tag{6}$$

This formula $\sigma(x) =\ (1 + e^{-x})^{-1}$ express the sigmoid function. $f_d\ (X_T\ )$ is the transfer function for output.

$$f_d\,(X') = (W_T\, X_T\ ) \tag{7}$$

Two subtrees are defined when the tree has 2 child nodes. Tree nodes are calculated by n_depth. If the depth is 2, the number of leaf nodes is 4. The number of leaves is calculated by $2^{n_{depth}+1}$.

4 Experimental Study and Performance Evaluation

The European Cardholders dataset is used for experimental study. Detailed information about the results obtained is given in the following sections. The T4 GPU with 12GB RAM is used in the work. Keras and Tensorflow libraries are used for programming.

4.1 Performance Metrics

Precission, Recall (Sensitivity) and F1-Score parameters [35] are used to evaluate the proposed models. The formula of the evaluation parameters is given in Equations 8, 9 and 10.

$$Pre(k) = \frac{TP(k)}{TP(k) + FP(k)} \tag{8}$$

$$Recall(k) = \frac{TP(k)}{TP(k) + FN(k)} \tag{9}$$

$$F1 - Score(k) = \frac{2 * Pre(k) * Sen(k)}{Pre(k) + Sen(k)} \quad (10)$$

The proposed work uses a two-class evaluation approach to classify fraudulent transactions. True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) metrics are used for evaluation. Table 2 shows the representation of these metrics.

Table 2
Confusion Matrix for Performance Evaluation

	Predicted Values	
Actual Values	True-Positive (TP)	False-Negative (FN)
	False-Positive (FP)	True-Negative (TN)

4.2 Dataset Description

In this study, we use the publicly available credit card fraud detection dataset [36], which contains 284,807 transactions, including 492 fraud cases. This indicates a severely imbalanced class distribution, where the fraudulent class represents a very small portion of the overall data. Due to confidentiality constraints imposed by the dataset provider, the original raw features are not accessible. Instead, the dataset includes 28 anonymized numerical features (V1-V28) that were obtained through a Principal Component Analysis (PCA) transformation performed by the data provider. In addition to these PCA-derived components, the dataset contains two further numerical variables: Time and Amount. The Time variable represents the elapsed time between each transaction and the first transaction in the dataset, while Amount denotes the transaction amount. Consequently, each transaction is represented by a 30-dimensional dense numerical feature vector. For model development and evaluation, the dataset is partitioned into training and test subsets using an 80/20 split, where 80% of the samples are used for training and the remaining 20% are reserved for testing. Table 3 summarizes the dataset characteristics, including the number of transactions, class distribution, and feature composition.

Table 3
Original and augmented data in the dataset

	Total Data	**Normal**	**Fraud**	**Imbalance Rate (%)**
Original Dataset	284.807	284.315	492	0.1730
Training (80%)	227.845	227.455	390	0.1711
Test (20%)	56.962	56.860	102	0.1729

4.3 Experimental Study

In the experimental study, the proposed Autoencoder+RF method is compared with well-known deep learning and machine learning models such as CNN, RNN, ANN, and Xgboost to demonstrate its success. The original dataset was split into two parts, with 80% of the data used for training and the remaining 20% used for testing.

An autoencoder was employed to decrease the data size of the original dataset. The data is encoded in the autoencoder layer and then transmitted to the decoder layer through the hidden layer. The fully connected layers produced in the hidden layer are used as inputs for the Random Forest (RF) algorithm. The hyperparameter values of the Autoencoder used in the Autoencoder are shown in Table 4.

Table 4
Original and augmented data in the dataset

Hyper Parameters	**Values**
Optimization Algorithm for Autoencoder	Adadelta
Number of layers (Encoder and Decoder)	6
Loss Function for Autoencoder	Mean Squared Error
Bottleneck Layer's	10
Bottleneck Activation Function	Relu
Training Batch Size for Autoencoder	256

In order to improve the accuracy of classification, it is important to have a meaningful representation of the input data before classification. The proposed model consists of three stages: data reduction using an autoencoder, generating connected layers, and classification using a random forest algorithm. Here's a step-by-step explanation of how the proposed model works: First, an autoencoder is applied to the original credit card fraud dataset. Using the Autoencoder, a new dataset with representative, low-dimensional, and preprocessed data is obtained from the original dataset. Next, fully connected layers in the autoencoder model are generated, which are then used as input for the random forest classifier.

The AE+RF hybrid model with hyperparameter optimization in the Random Forest part and parameter tuning overcomes the problem of misclassification costs. By adjusting the decision threshold and the model adjustments to reduce the false positives and false negatives, our framework ultimately achieves the balance between the two costs. We proposed a novel fraud detection methodology incorporating performance considerations to achieve the optimal outcome.

4.3 Results and Discussions

The hyperparameters of the proposed Autoencoder+RF model and other ANN, CNN, RNN and Xgboost models used for comparison are given in Table 5.

Table 5
Hyperparameters of all machine learning and deep learning models

Classifier	Hyper Parameters	Values
Random Forest	Number of Trees	100
	Random State	42
Xgboost	Number of Trees	100
	Maximum Depth	3
	Learning Rate	0.2
	Subsampling Rate	0.8
	Column Subsampling Rate	0.8
	Alpha Regularization	0.005
	Lambda Regularization	0.1
ANN	Hidden Layer 1	256 neurons
	Hidden Layer 2	128 neurons
	Hidden Layer 3	64 neurons
	Optimizer	Adam
	Activation Function	Sigmoid
	Loss Function	Binary Cross Entropy
	Batch Size	256
	Learning Rate	1e-3
RNN	Number of RNN Cells (SimpleRNN)	64
	Number of Neurons in Dense Layer	32
	Learning Rate	1e-3
	Optimizer	Adam
	Dropout Rate	0.5
CNN	Number of Filters	32
	Learning Rate	1e-3
	Kernel Size	3
	Number of Neurons in Dense Layer	100
	Optimizer	Adam
	Dropout Rate	0.5

To compare the success of the proposed model, all models were first used for classification directly using the original dataset without using an autoencoder. Comparison performance values for all models are given in Table 6. A comparison of models applied to the original data set for credit card fraud detection reveals meaningful findings. RF and Xgboost exhibit the highest precision scores of 0.99, indicating their proficiency in accurately identifying fraudulent transactions. However, regarding recall, which measures the ability to detect actual instances of fraud, Xgboost slightly outperforms RF with a score of 0.89 compared to RF's 0.88. This suggests that Xgboost is better at capturing a larger proportion of true positives. Moreover, Xgboost achieves the highest F1 score of 0.93, signifying a balanced trade-off between precision and recall. When considering the area under the receiver operating characteristic curve (AUC-ROC), Xgboost once again leads with a score of 0.97, highlighting its effectiveness in distinguishing between positive and

negative cases. Additionally, Xgboost surpasses its counterparts in terms of the area under the precision-recall curve (AUC-PR) with a score of 0.88, indicating its superior ability to rank and retrieve fraudulent cases effectively.

Table 6
Performance comparison of Models

Models	Precision	Recall	F1	AUC-ROC	AUC-PR
RF	0.99	0.88	0.93	0.94	0.86
Xgboost	0.99	0.89	0.93	0.97	0.88
ANN	0.92	0.89	0.91	0.98	0.85
RNN	0.92	0.89	0.90	0.96	0.78
CNN	0.93	0.88	0.90	0.98	0.82

The table 7 below reports the results of the proposed Autoencoder+RF AE+RF models when the original data set is examined and with 10 and 20 reduced dimensions. Here, the Autoencoders model is reduced by 10 layers and, subsequently, the reduced dimensions approach is used to get 10 dimensions and used as an input for our proposed model and other models based on a two-stage model. Therefore, the Table below notes our proposed AE+RF models and other models when 10 reduced dimension established. The reveals that the AE+RF model proposed is extremely strong and accurate, given its high quality in performance metrics. All demonstrated performance measures have excellent precision, recall, and F1 scores of 0.99, 0.97, and 0.98, respectively, implying that this model can make even more correct predictions rather than alerting the false alarms and can be applied to identify the fraudulent transactions. Its high AUC-ROC and AUC-PR scores, which were 0.99 and 0.98, respectively, confirm that it has a high capacity to discriminate between the favorable versus the unfavorable outcomes. Indeed, the AE+RF model proposed is a very high-quality application which has the high-quality performance metrics. The AE+XGBoost method also proves to be accurate, given the fact that its precision score is very high, and the model provides a 0.95 recall score. Regarding the results achieved concerning the precision measure, the AE+ANN and AE+RNN configurations exhibit an identical score of 0.99. At the same time, their recall metrics are equal to 0.93 and 0.81.

Table 7
Performance results were obtained using Autoencoders (10) dimensions

Models	Precision	Recall	F1	AUC-ROC	AUC-PR
Proposed AE-RF	**0.99**	**0.97**	**0.98**	**0.99**	**0.98**
AE+Xgboost	0.99	0.95	0.97	0.98	0.96
AE+ANN	0.99	0.93	0.96	0.98	0.93
AE+RNN	0.98	0.81	0.87	0.97	0.92
AE+CNN	0.94	0.93	0.92	0.99	0.95

We applied another 20-layer reduction process to the Autoencoders model. (Table 8) Next is the amount of 20 dimensions that we use as input of our proposed model and other models. We compared the performance of AE-RF with the others using the table of 20 reduced sizes and obtain the comparison results. The model proposed, the AE+RF, has obtained the best performance in measurement methods with a perfect measure to the precision measure the F1 and recall with a good point nine hundreds. So, the model is excellent in predicting fraud transactions and has a promising number of false alarms. On the other hand, all our points have a high predictive relationship performance on the AUC-ROC and the AUC-PR measured in 0.99. The AE+Xgboost worked extremely well producing high precision, recall, and F1 rates in 0.99. This indicates that the AE+Xgboost can identify true positives rigorously and place it as the second-highest dependence after the suggested AE+RF. The AE+ANN produced outstanding results with precision and recall and an F1 rate of 0.99. From this, it is evident that the model minimizes the rate of false positives and catches true correct positives. On the other hand, the AE+RNN and AE+CNN obtained F1 scores of 0.98 and 0.66 and obtain recall rates of 0.98 and 0.61.

Table 8
Performance results were obtained using Autoencoders (20) dimensions

Models	Precision	Recall	F1	AUC-ROC	AUC-PR
Proposed AE-RF	**1.00**	**0.99**	**0.99**	**0.99**	**0.99**
AE+Xgboost	0.99	0.99	0.99	0.99	0.99
AE+ANN	0.97	0.97	0.99	0.99	0.99
AE+RNN	0.98	0.98	0.98	0.99	0.98
AE+CNN	0.97	0.61	0.66	0.99	0.99

In this study, we described how our proposed RF+AE model performs compared to 10 and 20-dimension reduction PCA models. Table 9 presents the results of the RF+AE model and other models based on 10-dimensional reduction of PCA. Our designed AE+RF model confirmed excellent precision 99%, recall 97%, leading to an F1 score of 98%. Moreover, we demonstrated excellent discrimination with an AUC-ROC of 99% and high AUC at the PR curve of 98%. Those applying PCA together with out-of-the-box machine learning models, including XGBoost, ANN, RNN, and CNN, had a somewhat worse performance. Of these, the PCA+XGBoost model demonstrated the best outcomes in terms of precision and recall, totaling 95% and 86%, respectively. It yielded an F1 score of 90% and AUC-ROC and AUC-PR scores of 95% and 74%, respectively. It is worth mentioning that the PCA+ANN, PCA+RNN, and PCA+CNN models show relatively similar performance between each other while trading-off precision-recall. Nevertheless, the best performance is demonstrated by the AE+RF model with PCA of only 10 dimensions, compared to overall models.

The results for the AE+RF and other models using 20 dimensions for PCA reduction can be seen in Table 10. AE+RF model with a 20-D autoencoder had an outstanding

recall and F1 score of 0.99 each. The remarkable figures show the finding of an exceptional model in identifying fraudsters and non-fraud transactions.

Table 9
Performance results were obtained using PCA (10) dimensions

Models	Precision	Recall	F1	AUC-ROC	AUC-PR
PCA+Xgboost	0.95	0.86	0.90	0.95	0.74
PCA+ANN	0.93	0.87	0.89	0.95	0.81
PCA+RNN	0.92	0.87	0.89	0.95	0.79
PCA+CNN	0.92	0.85	0.88	0.97	0.71
Proposed AE+RF	**0.99**	**0.97**	**0.98**	**0.99**	**0.98**

Further, using the AUC-ROC, the current model had a figure of 0.99, which is an outstanding discrimination capacity. The higher AUC-PR of the model at 0.99 indicates the excellent model in handling the trade-off between precision and recall. The other models including PCA+XGBoost, PCA+ANN, PCA+RNN, and PCA+CNN had lower measures of performance. For instance, PCA +Xgboost had a fabulous precision at 98% and a reasonable recall at 89%, giving the F1-score of 93%. Although this model achieved a well-balanced performance, it also demonstrated a high discriminatory power, with the AUC-ROC being 0.97. Therefore, the proposed AE+RF model with a 20-dimensional autoencoder framework is the most effective, as it demonstrated the highest precision, recall, and discrimination.

Table 10
Performance results were obtained using PCA (20) dimensions

Models	Precision	Recall	F1	AUC-ROC	AUC-PR
PCA+Xgboost	0.98	0.89	0.93	0.97	0.84
PCA +ANN	0.94	0.90	0.92	0.95	0.82
PCA +RNN	0.91	0.89	0.90	0.98	0.81
PCA +CNN	0.94	0.90	0.92	0.99	0.85
Proposed AE+RF (20 dimensions)	**1.00**	**0.98**	**0.99**	**0.99**	**0.99**

Table 11 below represents the achieved performance of the proposed AE+RF model with 10 and 20 dimensions compared with PCA models of 10 to 20 dimensions. To summarize the results, the AE+RF model with 10 dimensions shows the best precision, recall, and F1 score, with 0.99, 0.97, and 0.98 outcomes. These scores are better compared to all PCA models with 10, 12, 14, 16, 18, and 20 dimensions. Furthermore, the PCA+RF models do not show as good performance in fine recall balance as the AE+RF model, and this work provides AUC-ROC and AUC-PR scores of 0.99 and 0.98. Additionally, as seen, the AE-RF model with 20 dimensions presents the best outcomes, which are perfect and extremely high precision of 1.00 and recall of 0.99. This model has excellent discriminatory power with an AUC-ROC and AUC-PR score of 0.99. The PCA model all had low performance values compared to all proposed AE+RF models. Ultimately, the proposed AE-RF model

performed the best in this study, and at 20 dimensions, it exhibited the highest precision, recall, and discriminative power.

Table 11
Performance results were obtained using PCA (20) dimensions

Models	Precision	Recall	F1	AUC-ROC	AUC-PR
PCA+RF (10 dimensions)	0.94	0.80	0.86	0.93	0.74
PCA+RF (12 dimensions)	0.99	0.86	0.91	0.94	0.84
PCA+RF (14 dimensions)	0.98	0.86	0.91	0.94	0.80
PCA+RF (16 dimensions)	0.97	0.89	0.93	0.92	0.81
PCA+RF (18 dimensions)	0.97	0.87	0.92	0.95	0.81
PCA+RF (20 dimensions)	0.97	0.91	0.94	0.95	0.86
Proposed AE+RF (10 dimensions)	**0.99**	**0.97**	**0.98**	**0.99**	**0.98**
Proposed AE+RF (20 dimensions)	**1.00**	**0.99**	**0.99**	**0.99**	**0.99**

As shown in Tables 6-11, AE-based representations improve the performance of all downstream models. We selected Random Forest (RF) as the primary classifier because it provides the most favorable and consistent results on fraud-critical metrics, especially recall and AUC-PR. For example, with AE (10 dimensions) (Table 7), AE+RF achieves the highest recall (0.97) and AUC-PR (0.98) compared to AE+XGBoost (0.95, 0.96) and AE+ANN (0.93, 0.93). With AE (20 dimensions) (Table 8), AE+RF reaches near-perfect performance (Precision=1.00, Recall=0.99, F1=0.99, AUC-PR=0.99). In addition, AE-derived features show a clear synergy with RF, substantially outperforming PCA+RF (Tables 10-11). Considering both performance and practical robustness (lower sensitivity to tuning and stable generalization), RF was chosen as the second-stage classifier in the proposed AE+RF pipeline. The comparison of the results obtained is as shown in Table 12 above. The results obtained from the comparison indicate that the proposed AE+RF model produces better results than the models in literature. NR stands for not reported.

Table 12
Comparison of studies using credit card fraud dataset in the literature

Study	Year	Method	Accuracy	Precision (%)	Recall	Specificity	F1-Score
[37]	2023	AED-LGB (Autoencoder + LightGBM)	99.93	NR	80.39	99.97	NR
[38]	2023	UAAD-FDNet (Unsupervised attentional anomaly detection)	NR	97.95	75.53	NR	85.29
[39]	2023	GAN + ANN (Data Augmentation)	NR	94.10	94.10	94.10	94.10
[40]	2023	CAE + Random Forest	99.21	99.21	99.21	99.21	99.21
[41]	2023	PCA + LightGBM	98.87	98.87	98.87	98.87	98.87
[42]	2024	GAN + BiLSTM	NR	98.40	97.10	NR	97.74
[43]	2024	Stacked AE + CatBoost	99.34	99.34	99.34	99.34	99.34

[44]	2024	Transformer-based Sequential Model	99.41	99.41	99.41	99.41	99.41
[45]	2024	VAE + Random Forest	99.08	99.08	99.08	99.08	99.08
[46]	2024	GAN + GRU	NR	99.20	99.00	100.0	99.10
[47]	2025	CNN + Attention	98.95	98.95	98.95	98.95	98.95
[48]	2025	AE + Ensemble + XAI	99.48	99.48	99.48	99.48	99.48
[49]	2025	Lightweight DNN + SMOTE	98.62	98.62	98.62	98.62	98.62
[50]	2025	Hybrid AE + XGBoost	99.26	99.26	99.26	99.26	99.26
	2025	**Proposed Model**	99.98	1.00	0.99	0.99	0.99

Conclusion

The study proposed a two-stage framework to prevent credit card fraud. In the first stage, an autoencoder was used to reduce the dimensions of the data in a dataset. The dimensionally reduced data obtained from the fully connected layer of the Autoencoder was then fed into a Random Forest (RF) model for further analysis in the second stage. The goal was to reduce the costs that could arise from misclassification of normal and fraudulent transactions. The hybrid AE+RF model, which includes parameter tuning and hyperparameter optimization in the Random Forest component, was designed to strike a balance between minimizing false positives and false negatives by adjusting the decision threshold and fine-tuning the model parameters.

The proposed model was compared with other models such as ANN, CNN, RNN, and Xgboost, which were also applied using autoencoders. The comparison was based on various performance metrics such as precision, recall, F1-Score, AUC-ROC, and AUC-PR values. The results showed that the 10-dimensional proposed AE+RF model outperformed its closest competitor by 2.11% in recall value, 1.03% in F1-Score value, 1.02% in AU-ROC value, and 2.08% in AUC-PR value. However, the precision values were the same. On the other hand, when comparing the 20-dimensional results, the AE+RF model had 1% better precision value, while other performance values were the same.

The dataset used in this study is highly imbalanced, with fraudulent transactions representing only 0.17% of the total samples. While the Autoencoder effectively reduces dimensionality, the imbalance may limit the model's ability to capture meaningful representations for the minority fraud class. Although resampling techniques like SMOTE or undersampling were not applied in this study, they could potentially improve the model's performance by balancing the dataset, particularly in the context of training the classifier to better detect fraudulent transactions.

As part of the study, PCA dimension reduction was applied to all other models to compare their success with the proposed model. The performance of ANN, CNN, RNN, and Xgboost models, with 10- and 20-dimensional PCA, were compared with the proposed AE+RF model. The AE+RF model achieved better precision, recall,

F1-Score, AUC-ROC, and AUC-PR values in both dimensions. These results indicate that the combination of representation learning, deep learning, and machine learning methods can be effective in preventing credit card fraud. Therefore, the study recommends using Autoencoder and transfer learning applications together for future studies

References

[1] B. Mohanty, “Role of artificial intelligence in financial fraud detection,” 2023

[2] E. Central Bank, “Seventh report on card fraud,” 2021, doi: 10.2866/603420

[3] R. Van Belle, B. Baesens, and J. De Weerdt, “CATCHM: A novel network-based credit card fraud detection method using node representation learning,” *Decis Support Syst*, Vol. 164, Jan. 2023, doi: 10.1016/j.dss.2022.113866

[4] A. S. A. Issa and Z. Albayrak, "CLSTMNet: A Deep Learning Model for Intrusion Detection," Journal of Physics: Conference Series, Vol. 1973, 2021, Art. no. 012244, doi: 10.1088/1742-6596/1973/1/012244

[5] Özalp, A. N. and Z. Albayrak, Detecting cyber attacks with high-frequency features using machine learning algorithms. Acta Polytechnica Hungarica, 2022. 19(7): pp. 213-233

[6] Avcı, İ. and M. Koca, Cybersecurity Attack Detection Model, Using Machine Learning Techniques. Acta Polytechnica Hungarica, 2023. 20(7): pp. 29-44

[7] Issa, A.S.A. and Z. Albayrak, DDoS attack intrusion detection system based on hybridization of CNN and LSTM. Acta Polytechnica Hungarica, 2023. 20(2): p. 1-19

[8] Alarbi, A. and Z. Albayrak, Core Classifier Algorithm: A Hybrid Classification Algorithm Based on Class Core and Clustering. Applied Sciences, 2022. 12(7): p. 3524

[9] Yousif Hosain Dakhil, Cakmak M. (2025) XAI-XGBoost: an innovative explainable intrusion detection approach for securing internet of medical things systems. *Scientific Reports*, 15, 1-17, Doi: 10.1038/s41598-025-07790-0

[10] M. Ali, R. Borgo, and M. W. Jones, “Concurrent time-series selections using deep learning and dimension reduction,” *Knowl Based Syst*, Vol. 233, Dec. 2021, doi: 10.1016/j.knosys.2021.107507

[11] Y. Bengio, A. Courville, and P. Vincent, “Representation Learning: A Review and New Perspectives,” Jun. 2012 [Online] Available: http://arxiv.org/abs/1206.5538

[12] S. Yang, Y. Zhang, H. Wang, P. Li, and X. Hu, “Representation learning via serial robust autoencoder for domain adaptation,” *Expert Syst Appl*, Vol. 160, Dec. 2020, doi: 10.1016/j.eswa.2020.113635

[13] Cakmak M. A New Lightweight Hybrid Model for Pistachio Classification Using Transformers and EfficientN.et IEEE Access, 13, 85857-85872, Doi:10.1109/ACCESS.2025.3567774 (2025)

[14] F. K. Alarfaj, I. Malik, H. U. Khan, N. Almusallam, M. Ramzan, and M. Ahmed, "Credit Card Fraud Detection Using State-of-the-Art Machine Learning and Deep Learning Algorithms," *IEEE Access*, Vol. 10, 2022, doi: 10.1109/ACCESS.2022.3166891

[15] E. Strelcenia and S. Prakoonwit, "A Survey on GAN Techniques for Data Augmentation to Address the Imbalanced Data Issues in Credit Card Fraud Detection," *Mach Learn Knowl Extr*, Vol. 5, No. 1, 2023, doi: 10.3390/make5010019

[16] B. Bestami Yuksel, S. Bahtiyar, and A. Yilmazer, "Credit Card Fraud Detection with NCA Dimensionality Reduction," in *ACM International Conference Proceeding Series*, 2020, doi: 10.1145/3433174.3433178

[17] A. Saxena, "Credit Card Fraud Detection using Machine Learning and Data Science," *Int J Res Appl Sci Eng Technol*, Vol. 10, No. 12, 2022, doi: 10.22214/ijraset.2022.48293

[18] Z. Salekshahrezaee, J. L. Leevy, and T. M. Khoshgoftaar, "The effect of feature extraction and data sampling on credit card fraud detection," *J Big Data*, Vol. 10, No. 1, 2023, doi: 10.1186/s40537-023-00684-w

[19] A. Arafa, N. El-Fishawy, M. Badawy, and M. Radad, "RN-Autoencoder: Reduced Noise Autoencoder for classifying imbalanced cancer genomic data," *J Biol Eng*, Vol. 17, No. 1, 2023, doi: 10.1186/s13036-022-00319-3

[20] A. Oluwasanmi, M. U. Aftab, E. Baagyere, Z. Qin, M. Ahmad, and M. Mazzara, "Attention autoencoder for generative latent representational learning in anomaly detection," *Sensors*, Vol. 22, No. 1, 2022, doi: 10.3390/s22010123

[21] P. Y. Chen and J. J. Huang, "A hybrid autoencoder network for unsupervised image clustering," *Algorithms*, Vol. 12, No. 6, 2019, doi: 10.3390/a12060122

[22] C. Zhang and L. Xue, "Autoencoder with emotion embedding for speech emotion recognition," *IEEE Access*, Vol. 9, 2021, doi: 10.1109/ACCESS.2021.3069818

[23] Fiore, U., De Santis, A., Perla, F., Zanetti, P., & Palmieri, F. (2023) A GAN-based oversampling approach for imbalanced credit card fraud detection.

Expert Systems with Applications, 213, 118927. https://doi.org/10.1016/j.eswa.2022.118927-8

[24] Zhang, Y., Wang, S., & Chen, X. (2023) Credit card fraud detection using convolutional autoencoders and ensemble classifiers. *Knowledge-Based Systems*, 268, 110444, https://doi.org/10.1016/j.knosys.2023.110444

[25] Kou, Y., Lu, C. T., Sirwongwattana, S., & Huang, Y. P. (2023) Feature reduction and gradient boosting for large-scale credit card fraud detection.

Information Sciences, 635, 119-134, https://doi.org/10.1016/j.ins.2023.02.087

[26] JRoy, A., Sun, J., Mahoney, R., & Albarelli, A. (2024) GAN-BiLSTM: A hybrid deep learning framework for temporal credit card fraud detection. *Pattern Recognition*, 146, 109950, https://doi.org/10.1016/j.patcog.2023.109950

[27] Liu, H., Li, Y., & Zhang, Z. (2024) Stacked autoencoder and CatBoost-based hybrid model for imbalanced fraud detection. *Applied Soft Computing*, 150, 110987, https://doi.org/10.1016/j.asoc.2023.110987

[28] Chen, Q., Zhou, X., & Wang, J. (2024) Transformer-based sequential modeling for credit card fraud detection.IEEE Transactions on Neural Networks and Learning Systems, 35(4), 5432-5445, https://doi.org/10.1109/TNNLS.2023.3312147

[29] Alazab, M., Awajan, A., Mesleh, A., & Hobbs, M. (2024) A hybrid VAE–ensemble learning approach for highly imbalanced fraud detection datasets. IEEE Access, 12, 41755-41768, https://doi.org/10.1109/ACCESS.2024.3369123

[30] Singh, A., Thakur, N., & Sharma, A. (2024) Adaptive GAN-GRU framework for evolving credit card fraud detection. Neural Computing and Applications, 36, 15789-15805, https://doi.org/10.1007/s00521-024-09713-8

[31] I. D. Mienye and T. G. Swart, "A Hybrid Deep Learning Approach with Generative Adversarial Network for Credit Card Fraud Detection," *Technologies (Basel)*, Vol. 12, No. 10, Oct. 2024, doi: 10.3390/technologies12100186

[32] Wang, T., Li, K., & Xu, Y. (2025) Lightweight attention-based CNN for real-time credit card fraud detection. Future Generation Computer Systems, 152, 204-216, https://doi.org/10.1016/j.future.2024.12.018

[33] D. Bank, N. Koenigstein, and R. Giryes, "Autoencoders," Mar. 2020, [Online] Available: http://arxiv.org/abs/2003.05991

[34] W. Li *et al.*, "A perspective survey on deep transfer learning for fault diagnosis in industrial scenarios: Theories, applications and challenges," *Mech Syst Signal Process*, Vol. 167, 2022, doi: 10.1016/j.ymssp.2021.108487

[35] J. Ha, M. Kambe, and J. Pe, *Data Mining: Concepts and Techniques*. 2011, doi: 10.1016/C2009-0-61819-5

[36] A. Dal Pozzolo, O. Caelen, Y. A. Le Borgne, S. Waterschoot, and G. Bontempi, "Learned lessons in credit card fraud detection from a practitioner perspective," *Expert Syst Appl*, Vol. 41, No. 10, 2014, doi: 10.1016/j.eswa.2014.02.026

[37] Du, H., et al. (2023) AutoEncoder and LightGBM for Credit Card Fraud Detection Problems. *Symmetry*, 15(4), 870,

https://doi.org/10.3390/sym15040870

[38] Jiang, Y., et al. (2023) UAAD-FDNet: Unsupervised Attentional Anomaly Detection for Credit Card Fraud Detection. *Systems*, 11(6), 305, https://doi.org/10.3390/systems11060305

[39] Fanai, S., & Abbasimehr, H. (2023) A novel combined approach based on deep Autoencoder and deep classifiers for credit card fraud detection. *Expert Systems with Applications*, 217, 119562, https://doi.org/10.1016/j.eswa.2023.119562

[40] Afriyie, E., et al. Detecting and predicting fraud in credit card transactions using supervised machine learning algorithms. *Decision Analytics Journal*, 8, 100163, https://doi.org/10.1016/j.dajour.2023.100163 (2023)

[41] Karthika, S., & Senthilselvi, R. "Credit card fraud detection". *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-023-14365-6 (2023)

[42] Alamer, A. A., et al. (2023) RaKShA: A Trusted Explainable LSTM Model to Classify Fraud Patterns in Credit Card Transactions. Mathematics, 11(8), 1901, https://doi.org/10.3390/math11081901

[43] Zhang, J., et al. Fraud Detection with Improved Variational Auto-Encoder Generative Adversarial Network. *IEEE Access*. https://doi.org/10.1109/ACCESS.2023.3302339 (2023)

[44] Credit Card Fraud Detection Using Kernel Extreme Learning Machine with Dandelion Optimizer. Neural Computing and Applications. https://doi.org/10.1007/s00521-023-08567-3

[45] A. Btoush ve R. Zhou, "A Hybrid ML and DL Approach for Credit Card Fraud Detection with Optimized Performance," *Applied Sciences*, Vol. 15, No. 3, p. 1081, 2025. doi: 10.3390/app15031081

[46] El Kafhali, S., et al. (2024) "Credit card fraud detection". Information, 15(4), 227, https://doi.org/10.3390/info15040227

[47] Lian, X., et al. (2024) "Credit card fraud detection Improved LightGBM". IEEE Access. https://doi.org/10.1109/ACCESS.2024.3487212

[48] Mim, M. W., et al. (2024) A soft voting ensemble learning approach for credit card fraud detection. Heliyon, 10(4), e25466. https://doi.org/10.1016/j.heliyon.2024.e25466

[49] Abadlia, I., & Smairi, Z. (2025) Enhanced particle swarm optimization-based hyperparameter optimized stacked autoencoder for credit card fraud detection. *International Journal of Data Science and Analytics*, 20, 1239-1253, https://doi.org/10.1007/s41060-024-00524-x

[50] Shi, L., Luo, R., & Pau, L.-F. (2025) An attention-based balanced variational autoencoder method for credit card fraud detection. Applied Soft Computing, 177, 113190, https://doi.org/10.1016/j.asoc.2025.113190