Computer Science • 25(2) 2024 https://doi.org/10.7494/csci.2024.25.2.5967

Nirupam Shome
Devran Dey Sarkar
Richik Kashyap
Rabul Hussain Laskar

DETECTION OF CREDIT CARD FRAUD WITH OPTIMIZED DEEP NEURAL NETWORK IN BALANCED DATA CONDITION

Abstract *Due to the huge number of financial transactions, it is almost impossible for humans to manually detect fraudulent transactions. In previous work, the datasets are not balanced and the models suffer from overfitting problems. In this paper, we tried to overcome the problems by tuning hyperparameters and balancing the dataset with a hybrid approach using under-sampling and over-sampling techniques. In this study, we have observed that these modifications are effective in getting better performance in comparison to the existing models. The MCC score is considered an important parameter in binary classification since it ensures the correct prediction of the majority of positive data instances and negative data instances. So, we emphasize on MCC score and our method achieved an MCC score of 97.09%, which is far more (16 % approx.) than other state-of-the-art methods. In terms of other performance metrics, the result of our proposed model has also improved significantly.*

Keywords credit card, fraud detection, deep learning, fraud transactions

Citation Computer Science 25(2) 2024: 253–276

1. Introduction

In the banking system, a credit card is a plastic card issued by banks or any other financial institution. It enables the users to borrow funds from the concerned institutions for availing of any kind of goods and services that the user wants. Credit cards are issued with a condition that the cardholder must pay back the original borrowed amount along with some additional interest amount which is decided by the banks or financial institutions. In the modern world of cashless transactions, credit cards play a very important role.

The total number of credit card users in India in 2019 touched around 52 million and it goes on increasing rapidly day by day. A credit card is a good option for the users for cashless pay, along with a good source for respective banks revenue generation. But with all the benefits and perks credit cards is very susceptible to fraud. Credit card frauds are easier to ensue in a short period and can cause excessive loss to both the cardholder and the bank. Fraudsters always focus on camouflaging fraudulent transactions as legitimate, so it becomes challenging to detect and stop them. The report published by the National Crime Records Bureau (NCRB), mentioned cases of online financial fraud using credit or debit cards have abruptly increased by over 225% in India during the pandemic period of 2020.

A credit card fraud is a fraudulent money transaction that is unauthorized and is carried out without the knowledge of the legal cardholder. These sorts of frauds can be conducted in various ways, it can be the thief stealing the card physically or stealing the card details or information via any means such as cyber-attack, etc. Although there is no foolproof technique to stop credit card fraud, but with the help of machine learning this issue can be addressed and has shown its relevance in the past. The main objective of any credit card fraud detection scheme is to identify or detect any sort of suspicious acts or events in the transactions and report them to the financial institution.

In literature, many credit card fraud detection techniques have been proposed and they claim very good performance. Among the most successful methods, many of them use machine learning models. The paper [22] shows the utilization of artificial neural networks and self-organizing maps for this purpose. The described model utilizes the advanced combination of neural networks for data analysis by the set of facts obtained from the previous transactions. The self-organizing maps are used for pictorial representation of the status of various organizations. According to the authors, the concept of self-organizing maps will be extremely efficient for anomaly detection.

The paper [25] used the optimized light Gradient Boosting Machine (GBM) where Bayesian-based hyper-parameter optimization is utilized in combination with tuned light GBM. This model achieves good accuracy and an F1 score. A comparative analysis among different machine learning algorithms to select which performs the best in detecting credit card fraud is proposed in [20]. A random forest technique-

based fraud detection method is introduced in [14]. The proposed system achieved an accuracy of 90%. Cost-sensitive modelling using a neural network strategy for credit card fraud detection is projected in [10]. The cost-sensitive method initially deals with the imbalanced data issue. The neural network architecture uses two hidden layers in between the input and output layers. To study various neural networks, [5] provides a general idea about ANN, deep neural networks (DNN), convolutional neural networks (CNN), and recurrent neural networks (RNN) for credit card fraud detection.

The credit card fraud detection problem contains developing and training a model using historical credit card transaction records with the knowledge of fraud transactions. Then the trained model is used to detect a new fraudulent transaction. Our aim here is to spot maximum fraudulent transactions while minimizing the incorrect fraud classification. This paper aims to develop a credit card fraud detection model by the knowledge gathered from reviewing various other models, which can identify suspicious events in credit card transactions.

It is well an established fact that better proportional neurons with the appropriate number of hidden layers give better results. With a greater number of hidden layers, more features can be extracted but it will be helpful up to a certain limit, beyond that instead of meaningful extracting features overfitting of data arises with increased system complexity. Overfitting in neural networks can the addressed with data augmentation, simplifying neural networks, weight regularization, dropouts, and early stopping. Overfitting is one of the main reasons for errors like false positives under limited imbalanced data conditions. In imbalanced data conditions results in terms of accuracy may be good due to overfitting but Matthew's correlation coefficient (MCC) score will be poor.

To overcome this issue in our study we have an emphasis on MCC score rather than other metrics and we balance the dataset to get proper results. In our study, we have also tried to optimize the neural network model to regularization weights of the neurons. To achieve this we also introduced a dropout layer in our proposed model. To speed up model training and hyperparameter optimization, we have also used the predictive early-stopping approach.

This paper is organized as follows; Section 2 presents the background research and related work, Section 3 presents details of existing models and their limitations, Section 4 presents the dataset and its pre-processing, and Section 5 demonstrates the description and development of the proposed model. Section 6 shows the comparison of existing models with our proposed model and Section 7 ends the work along with the conclusion and future directions.

2. Background and related work

Credit card fraud is a growing issue in today's modern world where many transactions are going online. Thus, it is very important to develop a full-proof and robust system that can detect and even predict patterns of any sort of credit card fraud.

Credit card fraud can be categorized in the following ways:

a) *PoS Fraud*: It is called Point-of-sale (PoS) fraud. A small skimming device is implanted into normal PoS devices, and these skimming devices hack the user's data. These devices scan and store the card data while customers are doing the transaction.
b) *Phishing and Vishing*: This fraud is accomplished through a cyber network. These involve mimicking the official communication from the bank which acts as an inducement for the target user. In the subsequent stage, the user is asked to click on a link, and these links redirect to a fake website (with an original webpage appearance) and request the user for card details.
c) *Keystroke Logging*: This usually happens when the user clicks on a suspicious link, and he/she unknowingly installs malware on the system. The malware secretly monitors every activity done by the user and records every key pressed on the system by the user, ultimately stealing the card details.
d) *Application Fraud*: In this type of fraud, a fraudster impersonates as a genuine client by using stolen or faked documents to get a credit card. In this fraud, the fraudster is using a legal credit card using false papers.
e) *Theft or Loss of a Card*: It is the physical misplacing of a credit card or getting stolen by someone. In this situation, there is always a chance of fraudsters taking advantage of it and using it to do fraudulent transactions.

Some of the previous research works done regarding credit card fraud detection are described and summarization of each of these is also discussed in this section. Paper [26] gives us a concrete idea about credit card fraud, how data is exploited, and the methodologies available for the detection of fraud. In the paper [19], several machine-learning techniques were used for the purpose. The model proposed a hybrid of K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and decision trees for credit card fraud detection. In the paper decision tree, Extreme Learning Machines (ELM), KNN, Multilayer Perceptron (MLP), and SVM were compared with their proposed hybrid model. It is observed in [16] that an imbalanced classification of the data leads to misleading and inaccurate results. They proposed that logistic regression, decision tree algorithm, and Artificial Neural Networks (ANN) are the best algorithms based on different performance matrices. The authors used data balancing for the training purpose of the model. The model proposed by [23] shows random forest algorithm is the best to provide higher accuracy rates in detecting fraud instances. They proposed an ensembling learning approach for the purpose. Ensembling learning consists of Random Forest and neural networks.

In the paper published by Suresh Kumar and Asha RB [4], ANN, SVM, and KNN algorithms were implemented on credit card data. The ANN model performs the best among the three algorithms. The model uses 30 features and the ANN architecture uses 15 hidden layers. The activation function in the model is Rectified Linear Units (ReLU). An autoencoder-based unsupervised fraud detection system based on clustering has been proposed in [30]. The autoencoders used three hidden layers and

K-means for the purpose of clustering. The models perform well compared to the other proposed models developed using the European dataset. In [15], Kernel-based supervised hashing (KSH) for this purpose of detection has been used. KSH models are useful in solving problems with huge data quantity and dimension. Compared to previous models proposed regarding credit card fraud detection, this is relatively a new approach. In the paper [21], the authors proposed thirteen models based on ANN and logistic regression. Results show the performance of the ANN models is superior when compared with logistic regression models.

A multilayer perceptron neural network utilization for fraud detection is used by [18]. A comparative study among multi-layer perceptron (MLP), decision tree, and naïve Bayes algorithms has been done. The accuracy of the MLP on the test data was more compared to other algorithms. In [9], ANN with backpropagation was used for credit card fraud detection. The neural network has three hidden layers. The first hidden layer has 15 neurons. In the model, the hidden layers use ReLU activation functions and in the output layer, the sigmoid activation function was used. The model achieves good accuracy and F1 Score.

From the above studies made on various proposed deep learning models, it is very forthcoming that deep learning models often suffer from overfitting or underfitting. Overfitting occurs when the model performs fine in training data but gives poor results in the testing conditions. In underfitting, the model performance is poor in training data and testing data, in underfitting the model is not trained enough to generalize to other data. In [13, 29], information related to the overfitting and underfitting of a model is provided. The authors mentioned how and why the overfitting and underfitting of models occur during the training phase. The effects of overfitting and underfitting are explained and discussed the techniques available to stop the overfitting and underfitting of the model. A study on the dropout and how it helps in reducing the chance of overfitting the neural networks is described in [24].

3. Baseline models for credit card fraud detection and their limitations

There has been extensive research work on this topic done previously and numerous articles have been published. Among the state-of-the-art literature methods, the three latest works [3, 4, 9] are considered for extensive study and the development of a new model for credit card fraud detection.

3.1. Existing model 1

In the paper [4], the authors proposed the ANN architecture with 15 hidden layers for credit card fraud detection. They used an imbalanced dataset and achieved an accuracy of 97.2% whereas the other parameters like precision and recall are not satisfactory and are very low. On calculating the F1 score, it is found to be 78.59%. The architecture of the model proposed by the above-mentioned work is depicted in

the following Figure 1. The number of neurons in each hidden layer is not mentioned in detail, only the number of hidden layers and their activation function is revealed.

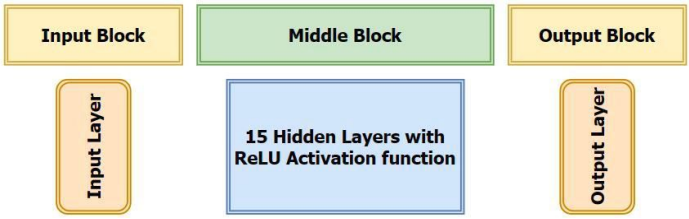

Figure 1. The architecture of baseline model-1

3.2. Existing model 2

The pictorial representation of the architecture mentioned in the paper [9] is shown below in Figure 2. The first hidden layer consists of 15 neurons and the activation function of all three hidden layers is ReLU. The diagram is constructed with the information provided by the authors in their paper. The model uses three hidden layers and achieves an accuracy of 97.2%. In this model, all the result parameters are close to ideal except the MCC score. The MCC value of the model is 81.46, which is very poor compared to the other metrics. Secondly, from the accuracy curve and the loss curve, it is clear that the curves are not smooth and tend to show some sort of overfitting which seems to be increasing with the number of epochs. Overfitting might tend to give inaccurate results when tested in a practical situation.

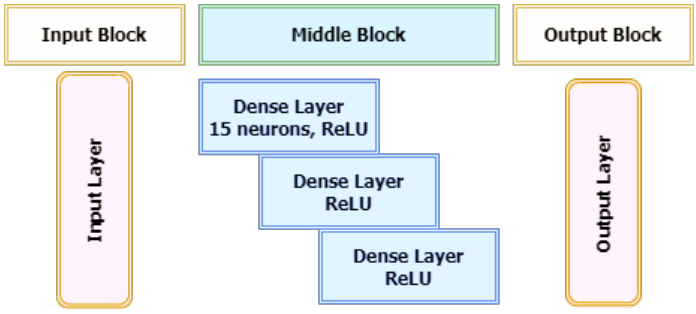

Figure 2. The architecture of baseline model-2

3.3. Existing model 3

In the paper [3], the authors proposed five hidden layers of ANN. The dataset is pre-processed and cleaned before being fed to train and test the model. The classifier has achieved a 95.3% accuracy. Furthermore, 95.2% of the fraud transactions are identified as shown by precision, and 95.55% times model is able to detect fraud transactions correctly as shown by recall rate. From the testing data, the model is also able

to predict 95.5% of fraud transactions correctly. Although the model performance is good with all results reaching above the mark of 95%, it is comparatively low as compared to existing model-2. On the other hand, the MCC score is improved for this model in comparison to others. Secondly, the dataset used for validating the model is again used for the final performance evaluation of the model. In the testing phase, the element of uncertainty in data is missing, as the model has already seen the dataset during validation. The model contains seven dense layers which are linked consecutively and consist of 1024, 512, 256, 128, 64, and 32 neurons, out of which, five are hidden dense layers. The architecture of the hidden dense layers is shown in Figure 3.

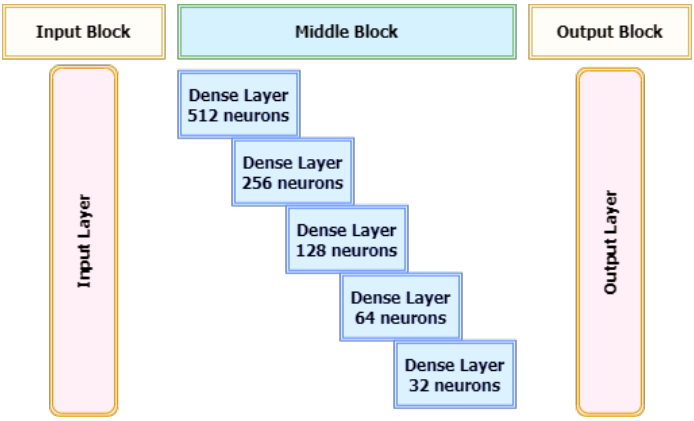

Figure 3. The architecture of baseline model-3

4. The dataset used for our analysis

The literature has employed a variety of datasets to evaluate and test credit card fraud detection methods. Typically, researchers use their own data. However, there are datasets that are freely accessible that might potentially be used. It is vital to note that having access to a reliable dataset is essential for enabling researchers to clearly compare their findings and demonstrate the superiority of one method over another. Finding datasets for financial research is challenging since the availability of such a dataset is still a problem in the field of credit card fraud detection. This is because the majority of banks do not share their data due to privacy issues and the sensitivity of client information. The data used for credit card fraud detection are of two types, real transaction data and Synthetic data. real transaction data are collected from real-world transactions whereas synthetic database is developed artificially rather than being gathered from actual situations. To develop a real-life system for credit card fraud detection, it is appropriate to develop a model on real-life data rather synthetic one. So, in this study, we have developed the model on real-life transaction data.

The dataset utilized for developing and validating the model is obtained from Kaggle [15], which contains information about various transactions made by European

users in September 2013. The dataset has transactions that occurred for two days. The dataset contains 284807 transaction details, of which 492 fraud transactions and the remaining are legitimate. The dataset is highly imbalanced because the fraud class is much less compared to the legitimate class. The fraud is 0.172% of all the transaction records.

The dataset is Principal Component Analysis (PCA) transformed and contains only numerical data. For the sake of the confidentiality of the users, the original features and some of the background information of the data are not provided by Kaggle. The principal components obtained after PCA are features V1, V2, V3, ..., V27, and V28. Features such as 'Time' and 'Amount' are not PCA transformed. The dataset contains 31 feature columns and 284807 rows of credit card transaction details. In the column 'Class', fraud transactions are denoted as '1', and legitimate transactions are denoted as '0'.

4.1. Dataset preparation

As the dataset is highly imbalanced, directly using the data for training the model can lead to erroneous results. To avoid these possibilities, the dataset needs to be balanced before usage. When the distribution of classes in the dataset is uneven, data imbalance results. According to [1], the class imbalance may be a natural phenomenon or result from the challenging nature of data collection due to high costs, privacy issues, and labor demands. The dataset of credit card transactions is frequently unbalanced since there are relatively few fraudulent transactions compared to legitimate ones. A variety of techniques, including under-sampling [27] and over-sampling [7], are proposed to address this problem. The extremely unbalanced credit card datasets and the fact that each instance of the dataset contains important data (such as transactions held by the same cardholder) make these solutions difficult to implement [28]. The third data balancing approach is the hybrid of under-sampling and over-sampling techniques. In this analysis, we use the hybrid approach for data balancing. The method proposed in [17] is implemented to balance our dataset. The author used Random Under Sampling (RUS) to test numerous SMOTE variants, including the original variant, the borderline1 and borderline2 variants, SVM-SMOTE, SMOTEENN, and SMOTETomek. A hybrid balancing model was evaluated using the Balanced Bagging Ensemble, which is internally balanced using RUS and SMOTE. The outcomes demonstrated the scalability and superior performance of hybrid approaches.

From the dataset, 10% of the data containing both fraud and legitimate transactions were sorted and stored in a separate CSV file to test the model. The transaction details available in the testing dataset are then balanced to get 28126 fraud and 28481 legitimate transactions and a total of 56607 transactions are available for testing purposes.

The transaction in the remaining 90% is again balanced by above mentioned technique, resulting in 243650 fraud transactions and 255834 legitimate transactions, a total of 499484 transactions. This data is further divided into training and validation datasets. The dataset is divided into 80% for training and 20% for validation purposes.

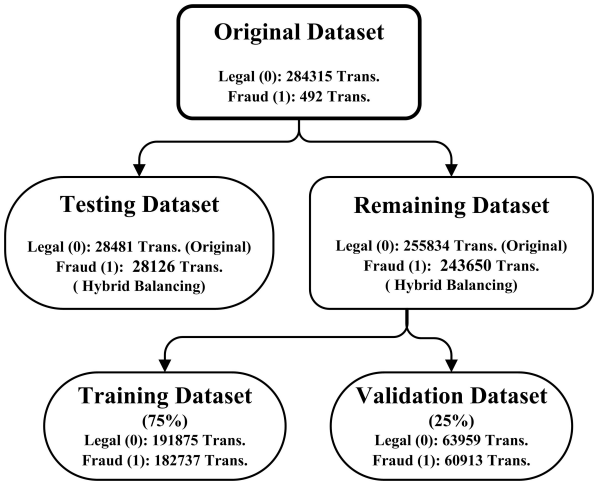

Figure 4. Representation of the dataset and its distribution into training, testing, and validation

5. Development of proposed model for credit card fraud detection

For the correct functioning of the neural network model, the number of hidden layers and associated neurons must be appropriate: neither more nor less. The development processes by which the number of hidden layers and associated neurons conform to the functional requirements of specific problems are complex and still, it is not addressed completely. To attain good results in neural networks a large number of hidden layers and neurons are required [11, 12]. We have studied and taken as a reference from a few research papers during the development of our proposed model. For finalizing the proposed models of credit card fraud detection, numerous deep learning models have been implemented. After extensive study associated with neural networks and simultaneously working on the drawbacks of the existing models, we came up with a final proposed model. After analyzing the three existing (baseline) models [3, 4, 9], it is observed that the existing (baseline) model-3 [20] is well described and properly explained. Therefore, necessary modification work is done on the existing (baseline) model-3 to build our proposed model. The other existing (baseline) models, namely 1 and 2 [4, 9] are used for comparison of results to show the improvement because these also used the same dataset.

The baseline model-3 consists of five hidden dense layers with one input and one output dense layer. Based on this, the investigative model-1 is developed which also consists of five hidden dense layers but the number of neurons in each layer is reduced. The results obtained from the investigative model-1 are analyzed and compared with existing model-1 and existing model-2, thereby encouraging further analysis because the results are not as expected.

To further improve the performance, the investigative model-2 is developed. This model uses six dense hidden layers with ReLU as the activation function. The investigative model-2 gives better results compared to our investigative model-1. However due to the increase in the number of layers and neurons, the computational time of the model is more, and the model takes more epochs to learn.

Further, we develop a model (Proposed Model) which consists of five dense hidden layers and one dropout layer. Now considering the results, it is seen that the proposed model is the best-performing model among all the designed and implemented models in terms of performance and computational time. The model also requires a lesser number of epochs to learn and performs better. To compare the existing and proposed models, Matthew's correlation coefficient (MCC) is one of the best parameters as it considers all the matrices (true positives, false negatives, true negatives, and false positives) available in the confusion matrix for evaluation. In the paper [8], a study regarding the reliability of MCC in binary classification has been done. The study suggests MCC is the most suitable parameter for binary classification since it ensures the correct prediction of the majority of positive data instances and the majority of negative data instances. Whereas, the F1 score merges precision and recall in a more interpretable way than MCC. For unknown or unquantifiable data conditions MCC is preferred over the F1 score as it is a more balanced calculation of classifiers, no matter which class is positive. So, we have given more emphasis on the MCC score over other performance evaluation matrices and decided on model performance on it. In Figure 5, we have represented the steps of the progressive development of our proposed models from the baseline model. It gives a clear picture of the design sequence of our analysis.

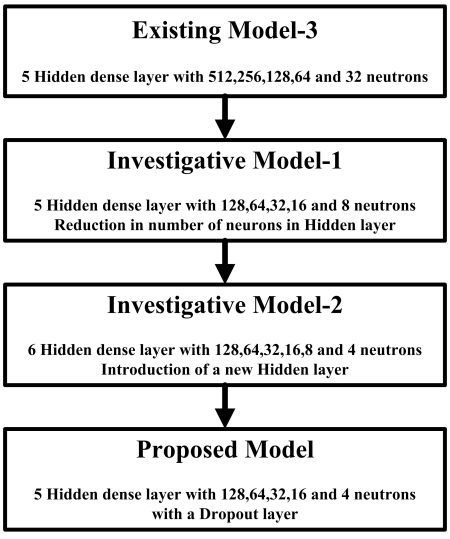

Figure 5. Progressive development of the proposed model

5.1. Description of investigative model-1 for credit card fraud detection

If a neural network model has too many neurons in the hidden layers results in overfitting, high variance, and increases the time it takes to train the network. This happens when the network has large information-handling capabilities but with limited training data, it is not possible to train the neurons in the hidden layers. In existing model-3, the number of neurons is very large as compared to the available data which results in inappropriate learning. So, in our investigative model-1, we have reduced the volume of neurons of each layer to address the issue. By selecting the proper activation function and other parameters, we can generate improved results as compared to the existing model-3.

The investigative model-1 consists of five hidden dense layers that are linked consecutively with 128, 64, 32, 16, and 8 neurons. The hidden layers contain the ReLU activation function, and the last layer is activated with the sigmoid activation function. The batch size of 700, 32 epochs, and the early stopping [13, 29] are used to circumvent the difficulties of overfitting and underfitting the model. Figure 6 shows the architecture of the investigative model-1.

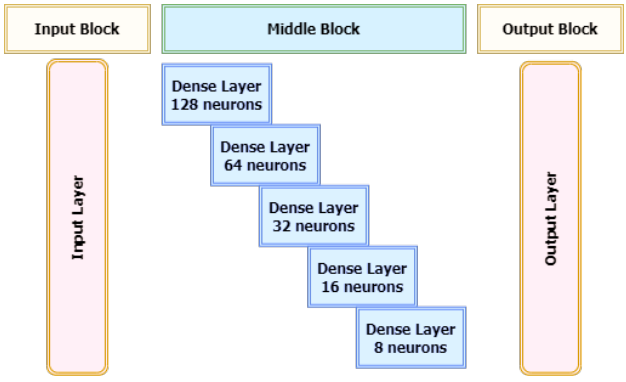

Figure 6. The architecture of Investigative model-1

The model when tested on the testing dataset gives satisfactory results (shown in Tables 1 and 2). Table 1 is the confusion matrix obtained for the investigative model-1, the model correctly identifies 28385 legitimate transactions which were genuinely legal and identifies 25256 transactions as fraudulent which were genuinely fraudulent. The model predicts only 96 fraudulent transactions which were actually legitimate and predicts 2870 transactions as legitimate but actually fraudulent.

Table 1

Confusion matrix of investigative model-1

		Predicted Label	
		Legitimate	Fraudulent
Actual Label	Legitimate	28385	96
	Fraudulent	2870	25256

Table 2 shows the result parameters of the investigative model-1, the model shows a precision of more than 99% and an F1 score of around 94% with an MCC value of 89.9%. but the accuracy of the model is slightly lower than the existing models. From the results, it is clear that our model is working fine and can differentiate between fraudulent and legal transactions.

Table 2
Result of investigative model-1

Accuracy	Precision	Recall	F1 Score	MCC
94.7603	99.6213	89.795	94.4537	89.947

The learning curves of the model shown in Figure 7 and Figure 8 also show that the model does not have any overfitting conditions. However the fluctuation of the validation curve shows that the model can be improved further, and the addition of an extra hidden layer may solve the issue raised.

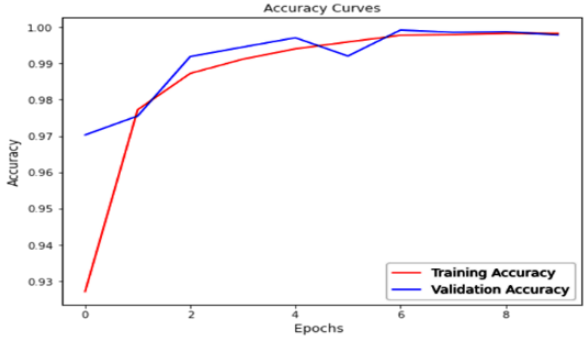

Figure 7. Accuracy curve of investigative model-1

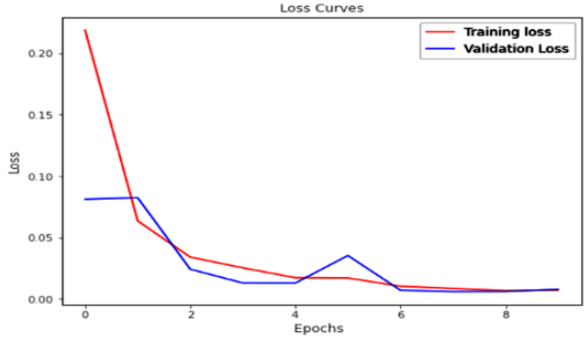

Figure 8. Loss curve of investigative model-1

5.2. Description of investigative model-2 for credit card fraud detection

In deep learning, better performance can be achieved if the entire problem is understood by the network without having overfitting and underfitting conditions [2, 12].

An increased number of hidden layers leads to an increase in the learning components that extract evidence from the previous activation to forward it to the next layer, which results in an improvement in performance and a reduction of bias. To handle big datasets and complex problems, a network with a large number of hidden layers normally shows high accuracy because the added layers provide more parameters to the model. And it allows the model to fit more complex functions. Inspired by this fact, we have increased the number of layers in investigative model-2 and checked the performance of the model with the same dataset (see Fig. 9).

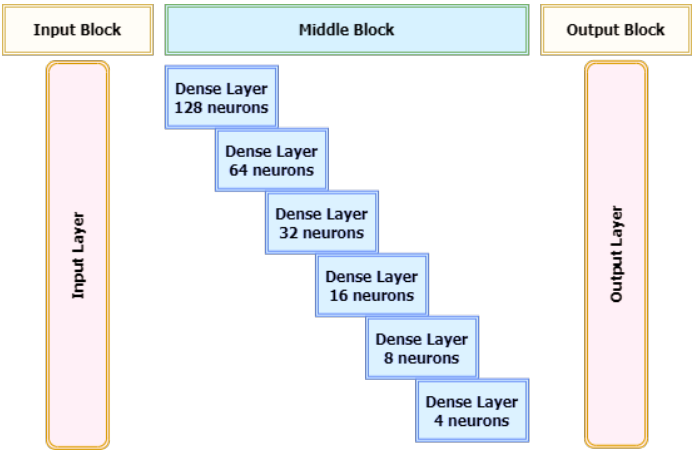

Figure 9. The architecture of investigative model-2

The investigative model-2 comprises six hidden dense layers that are linked consecutively with 128, 64, 32, 16, 8, and 4 neurons. The last layer is activated with a sigmoid activation function. With 32 epochs and 700 batch size, the early stopping technique [13, 29] is used to circumvent the difficulties of overfitting and underfitting the model. figure-9 shows the architecture of the investigative model-2.

The confusion matrix obtained for the investigative model-2 is shown in Table 3, the model correctly identifies 28378 legitimate transactions which were actually legal and identifies 26978 transactions as fraudulent which were actually fraudulent. The model predicts only 103 fraudulent transactions which were actually legitimate and predicts 1148 transactions as legitimate but actually fake.

Table 3
Confusion Matrix of Investigative Model-2

		Predicted Label	
		Legitimate	Fraudulent
Actual Label	Legitimate	28378	103
	Fraudulent	1148	26978

From Table 4 it is seen that the accuracy and F1 Score obtained is 97.7%, whereas the precision score is 99.6%. The results have improved when compared to the previous model. In the learning curves shown in Figure 10 and Figure 11, the number of epochs taken is around 17.5 which is early stopped. The curves are smoother thereby rejecting the possibility of overfitting. But as the number of hidden layers is increased the computational time of the model is increased.

Table 4
Result of Investigative Model-2

Accuracy	Precision	Recall	F1 Score	MCC
97.79	99.65	95.92	97.73	95.64

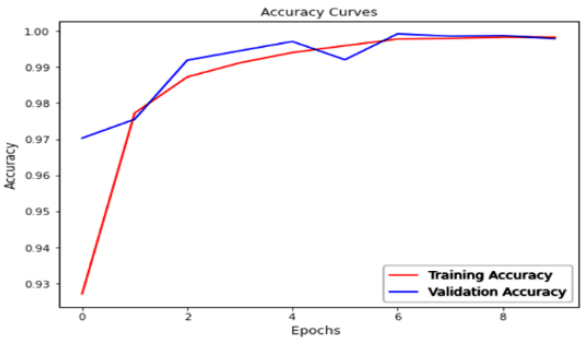

Figure 10. Accuracy curve of investigative model-2

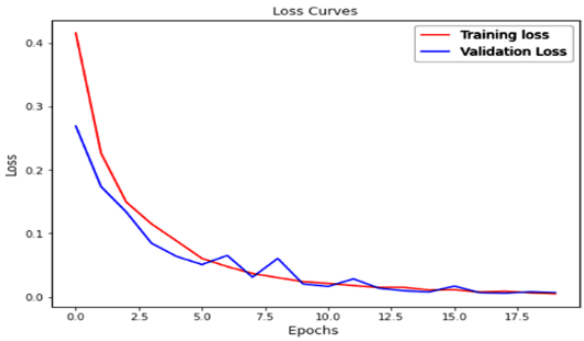

Figure 11. Loss curve of investigative model-2

5.3. Description of proposed model for credit card fraud detection

In many studies, it is observed that an increased number of layers provides a shortcut to improve the capabilities of the model with less amount of data. But it is also true that with the increased number of layers, the model may not extract meaningful features and system instead of learning the patterns, tries to memorize the information

provided to it in training. One of the most important parameters of model designing is that the last dense layer must have the precise number of neurons and the proper activation function to get the best results. Each hidden neuron added will increase the number of weights, thus it is recommended to use the least number of hidden neurons that accomplish the task. Using more hidden neurons than required will add more complexity. Here in our proposed model, we have reduced the number of layers to five in comparison to investigative model-2 for better optimization. To further reduce the model complexity and improve performance, the introduction of the dropout layer is the solution [24]. Dropout works by randomly setting the outgoing edges of hidden units (neurons that make up hidden layers) to 0 at each update of the training phase. The dropout mechanism stops all neurons in a layer from synchronously optimizing their weights and results in gradual improvement in performance and reduction in loss [6]. Dropout forces a neural network to learn more robust features that are useful in conjunction with many different random subsets of the other neurons. If dropout is increased beyond a certain threshold results in improper model fitness. A higher dropout rate results in a higher variance to some of the layers and also degrades training process. To build a more capable network for better generalization and less likely to overfit the training data, we have used one dropout layer in our proposed model.

The architecture of the proposed model consists of five hidden dense layers that are connected sequentially with 128, 64, 32, 16, 4 neurons and one dropout layer. The last layer is activated with a sigmoid activation function. With 32 epochs and 700 batch size, and the early stopping [14], a dropout layer [21] is added to circumvent the difficulties of overfitting and underfitting the model. Figure 12 shows the architecture of the proposed model.

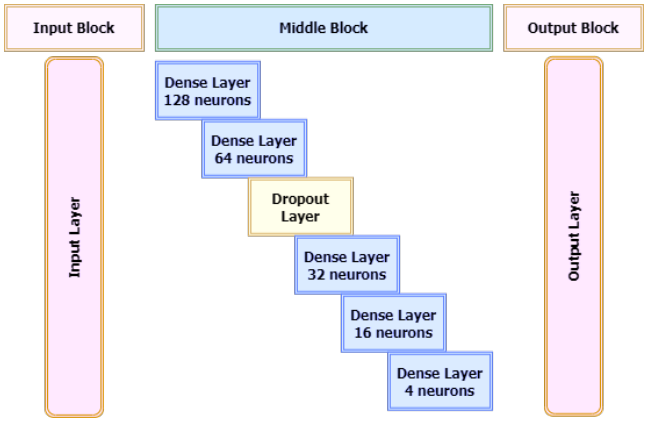

Figure 12. The architecture of the proposed model for credit card fraud detection

The detailed structure of the proposed model is shown in Table 5. The proposed feed forward neural network has the following architecture: the input to the model is

text data with dimension 29. Here we have five dense layers connected sequentially with 3840, 8256, 2080, 528, and 68 learnable parameters in each layer respectively. The total number of learnable parameters in our model is 14777, which is sufficient to learn the patterns from the data. ReLU Activation function is used followed by all the dense layers for better generalization and to capture diverse patterns in the data, leading to better generalization in the presence of unseen examples. The final activation function in the output layer is 'sigmoid', indicating that the model is intended for binary classification of legitimate and fraudulent levels. After the second dense layer, we added one dropout layer with a rate of 0.25 that indicates 25% of neurons are randomly set to zero during training to prevent overfitting. The kernel initializer parameter is set to 'uniform' for all dense layers indicating that the weights of the neurons in those layers are initialized from a uniform distribution.

Table 5

The detailed structure of the proposed model for credit card fraud detection

Sl. No.	Layer (type)	Input Shape	Output Shape	Learnable Parameters
1	Input Layer	(None, 29)	(None, 29)	0
2	Dense Layer-1 ReLU Activation	(None, 29)	(None, 128)	3840
3	Dense Layer-2 ReLU Activation	(None, 128)	(None, 64)	8256
4	Dropout Layer Dropout rate: 0.25	(None, 64)	(None, 64)	0
5	Dense Layer-3 ReLU Activation	(None, 64)	(None, 32)	2080
6	Dense Layer-4 ReLU Activation	(None, 32)	(None, 16)	528
7	Dense Layer-5 ReLU Activation	(None, 16)	(None, 4)	68
8	Output Layer Sigmoid Activation	(None, 4)	(None, 1)	5

The confusion matrix obtained for the proposed model is shown in Table 6, the model correctly identifies 28229 legitimate transactions which were actually legal and identifies 27552 transactions as fraudulent which were actually fraudulent. The model predicts only 252 fraudulent transactions which are actually legitimate and predicts 574 transactions as legitimate that are fraudulent.

Table 6

Confusion matrix of proposed model

		Predicted Label	
		Legitimate	Fraudulent
Actual Label	Legitimate	28229	252
	Fraudulent	574	27552

The model shown in Figure 12 gives an accuracy of 98.5% and precision of 99.1% with a recall of 97.96% and F1 scores tending around 99%. The MCC score obtained is around 97.09% which is the highest among all models. The learning curves shown in Figure 13 and Figure 14, signify no overfitting. Although there is an initial spike in the learning curve other than this the learning curves get smoother with each epoch and finally early stopped at epoch 13. The model takes less time to get trained compared to the other models and also generates better results in all aspects.

Table 7
Result of the proposed model

Accuracy	Precision	Recall	F1 Score	MCC
98.54	99.09	97.96	98.52	97.09

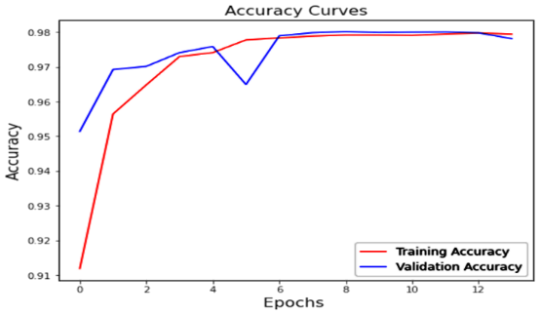

Figure 13. Accuracy curve of the proposed model

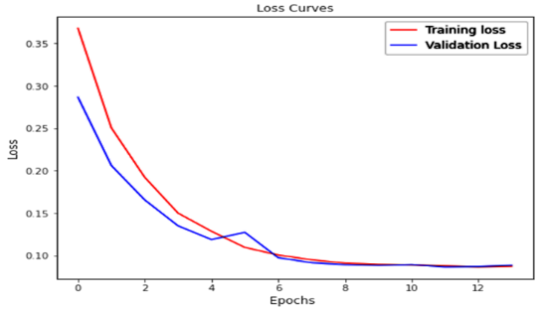

Figure 14. Loss curve of the proposed model

6. Comparison of existing models with our models

In this section, we have discussed the results obtained from our analysis, and Table 8 compares the results obtained from the investigative models and proposed model with existing models highlighting the limitations of existing models and improvements of our proposed model.

Table 8
Result comparison of different credit card fraud detection models

	Accuracy	Precision	Recall	F1 Score	MCC	
Baseline Models						Limitations
Model-1 [4]	97.23	81.15	76.19	78.59	80.26	• The use of a highly imbalanced dataset leads to high accuracy
Model-2 [9]	97.2	97.96	97.96	97.96	81.46	• No data preprocessing is done; directly using the imbalanced data leads to model overfitting (the model considers every transaction as legitimate). • The model is overfitted and can be observed from the learning curve. • Less amount of testing data
Model-3 [3]	95.3	95.2	95.55	95.5	82.47	• The dataset is divided into Training (70%) and Testing (30%), both are used for training and validating the model. The validation data is again used for evaluating the model performance
Factfinding Models						Refinements
Investigative Model-1	94.76	99.62	89.80	94.45	89.95	• In the proposed model, the dataset is pre-processed and balanced before training the model.
Investigative Model-2	97.79	99.61	95.92	97.73	95.64	• Early stopping during the training process and one dropout layer is introduced between hidden layers to stop overfitting or underfitting the model.
Proposed Model	98.54	99.09	97.96	98.52	97.09	• The number of neurons is optimized in the hidden layers for better performance (converges fast)

In this comparison, we have first identified the limitations of the existing models. The limitation of existing model-1 is in its training dataset. As the dataset consists of a very small number of fraud transactions, the model is not able to learn about fraud transaction patterns. Data preprocessing is not performed for existing model-2 and uses imbalanced data that leads to model overfitting. And the results obtained are on very less testing samples which leads to erroneous results. The shortcoming of the existing model-3 is the use of the same data for validation and testing the model. In testing, the model is not exposed to unknown data that gives unreliable results. To overcome these limitations, we have suggested a suitable model with adequate pre-processing and an optimized learning procedure. Before training the model, the

dataset for the proposed model is pre-processed and balanced. To prevent the model from overfitting and underfitting, early stopping is implemented during the training phase, and one dropout layer is included in between hidden layers. For improved performance (rapid convergence), the number of neurons in the hidden layers is adjusted by different sets of investigative models and finally proposed an optimized model with a definite number of neurons in each dense layer. We have achieved significant improvements in model performance with limited trainable parameters.

The Matthews Correlation Coefficient (MCC) is a more dependable statistical rate that conveys a high score only if the estimate is accurate in all of the four confusion matrix classes and both the size of the positive class and negative class [8] are proportional. Figure 15 shows the pictorial representation of the results obtained for existing models, investigative models, and the proposed model.

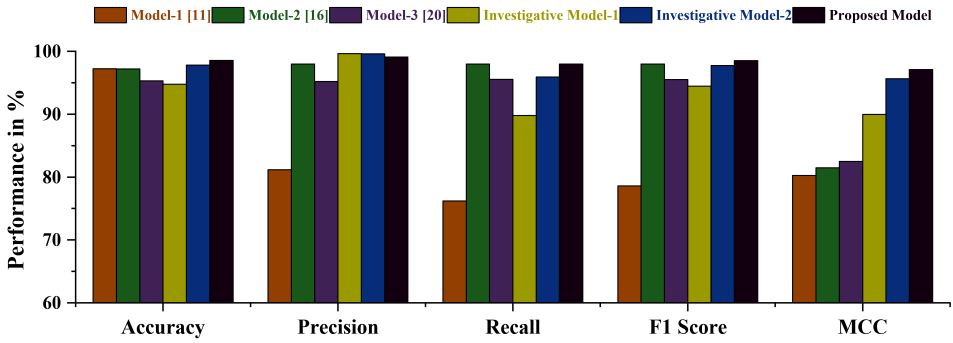

Figure 15. Performance analysis of various models

Comparing the MCC values, it is pragmatic that the proposed model is much better when compared with the existing models. We have achieved significant improvement of 16.83%, 15.63%, and 14.62% in MCC, compared to existing model-1, 2 and 3 respectively.

The accuracies of the existing models 1 and 2 are very close to the proposed model, this is due to the balanced datasets which are used to train the proposed model whereas, for the existing models, the dataset is highly imbalanced. From the results shown in Table 8, it is observed that the proposed model has a comparatively better precision value than the existing model-1 (improved 17.94%) and decent enhancement in precision score with respect to existing model-2 (improved 1.13%) and existing model-3 (improved 3.89%). This shows that our model is more capable of differentiating between false positives and true positives. Although achieving both high precision and high recall at the same time is difficult, but still our system is capable of maintaining the recall to a comparatively good value (slightly lesser than precision). The F1 score considers both precision and recall in its calculation. The F1 Score obtained for the proposed model is also upright and better than all existing

models (19.93% for existing model-1, 0.56% for existing model-2, and 3.02% for existing model-3).

Due to a smaller number of fraud transactions compared to genuine transactions in real-world circumstances, most of the available datasets are highly imbalanced. So, data balancing is extremely important before training the model. The use of such imbalanced data can lead to a biased model towards genuine class, which consequent to wrong predictions. A clear picture of a biased model can be seen in existing model-1, although the accuracy of the model is more than 95.23%, but the other parameters are not satisfactory.

The 'Time' column in the dataset consists of seconds elapsed between subsequent transactions. On analyzing the effect of various features on the output, it is seen that the 'Time' information does not play any significant role in the performance of a model. Therefore, removing the 'Time' column from the dataset and using the other 29 features proposed in the model gives better results.

For effective learning maximum volume of data is essential to train the model. But to prevent the model from overfitting, an early stopping technique during the training is necessary. Comparing the learning curves of the existing model-2 [9] with the proposed model shows that the proposed model does not have an overfitting problem.

From the development of investigative model-2 to the proposed model, it is seen that the introduction of a dropout layer improves the results to a great extent. The proposed model uses one less hidden layer thereby reducing the computation complexity.

From the comparison of the proposed model with the three existing models, we can draw the following conclusion:

a) The accuracy of the proposed model is better than the existing models.
b) The precision score of the proposed models is improved to a great extent than the existing models.
c) The recall values and the F1 Scores of the proposed model are slightly better than the existing model-2, 3 and much more than the existing model-1.
d) The MCC score of the proposed model is very high related to all the existing models (16.83%, 15.63%, and 14.62%). The MCC value of the existing model-2 decreases to 80.26% whereas all other performance parameters of it are around 97%. This variation shows the model discrepancy in prediction.
e) We have accomplished our analysis on balanced data, but existing model results are on unbalanced data. So, our proposed model is more generalized to predict fraud efficiently.

7. Conclusion and future work

Credit card fraud is an important problem in the world, and modern machine learning techniques like neural networks are a potent way of detecting fraudulent transactions

among a large number of transactions. From the literature, it is evident that neural networks like Deep Neural Networks (DNN) are more capable of establishing the association between input features and output even when there is no direct relationship between them. The development of our work started with several motivations: to develop a good and efficient model, to study different techniques used previously for this purpose, to rectify the limitations of the considered existing models, and to design a neural network to solve the problem of credit card fraud detection.

Whenever it comes to binary classification, the MCC score plays a vital role in determining the predicting capacity of the model. From the paper published by Davide Chicco and Giuseppe Jurman [8], it is evidenced that the MCC is a more reliable statistical parameter that produces a high score only if the model predicts effectively. Based on the MCC score, it is conclusive that our proposed model is much better compared to other existing models. Our proposed model achieves much more reliable and satisfactory results in terms of all parameters in comparison to other models.

Even after the finest efforts made in developing a robust credit card fraud detection model, there are still some limitations and shortcomings present. To train our model, the dataset is manually balanced, but in real-life circumstances, manual balancing is not feasible, so the use of other data balancing techniques like SMOTE and ADASYN can be implemented. Secondly, training the model on more available data will help in finding the drawbacks of the model. Finally, some traditional machine learning techniques like Decision Tree and Support Vector Machines (SVM) are also capable of giving good results, so developing a hyper-model consisting of neural networks and Decision Tree/SVM will help in developing a robust and efficient credit card fraud detection model. Though this study is concentrated on credit card fraud detection, the method developed in our study can help detect and protect the interest of any other plastic or virtual card holders.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Conflict of interest

The authors declare no competing interests.

Authors' contributions

All authors have contributed to this work

Funding

There is no external fund for this research.

Data availability statement

https://www.kaggle.com/datasets/mlg-lb/creditcardfraud

References

[1] Abd Elrahman S.M., Abraham A.: A review of class imbalance problem, *Journal of Network and Innovative Computing*, vol. 1(2013), pp. 332–340, 2013.

[2] Adil M., Ullah R., Noor S., Gohar N.: Effect of number of neurons and layers in an artificial neural network for generalized concrete mix design, *Neural Computing and Applications*, pp. 1–9, 2022.

[3] Alkhatib K.I., Al-Aiad A.I., Almahmoud M.H., Elayan O.N.: Credit Card Fraud Detection Based on Deep Neural Network Approach. In: *2021 12th International Conference on Information and Communication Systems (ICICS)*, pp. 153–156, IEEE, 2021. doi: 10.1109/ICICS52457.2021.9464555.

[4] Asha R.B., Suresh Kumar K.: Credit Card Fraud Detection Using Artificial Neural Network, *Global Transitions Proceedings*, vol. 2(1), pp. 35–41, 2021. doi: 10.1016/j.gltp.2021.01.006.

[5] Aslam S., Herodotou H., Ayub N., Mohsin S.M.: Deep learning based techniques to enhance the performance of microgrids: a review. In: *2019 International Conference on Frontiers of Information Technology (FIT)*, pp. 1160–1165, IEEE, 2019. doi: 10.1109/fit47737.2019.00031.

[6] Brownlee J.: Dropout Regularization in Deep Learning Models with Keras, *Machine Learning Mastery*, vol. 20, 2016.

[7] Chawla N.V., Bowyer K.W., Hall L.O., Kegelmeyer W.P.: SMOTE: synthetic minority over-sampling technique, *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002. doi: 10.1613/jair.953.

[8] Chicco D., Tötsch N., Jurman G.: The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation, *BioData Mining*, vol. 14(1), 13, 2021. doi: 10.1186/s13040-021-00244-z.

[9] Dubey S.C., Mundhe K.S., Kadam A.A.: Credit card fraud detection using artificial neural network and backpropagation. In: *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp. 268–273, IEEE, 2020. doi: 10.1109/iciccs48265.2020.9120957.

[10] Ghobadi F., Rohani M.: Cost sensitive modeling of credit card fraud using neural network strategy. In: *2016 2nd international conference of signal processing and intelligent systems (ICSPIS)*, pp. 1–5, IEEE, 2016. doi: 10.1109/icspis.2016.7869880.

[11] Gupta T.K., Raza K.: Optimizing deep feedforward neural network architecture: A tabu search based approach, *Neural Processing Letters*, vol. 51, pp. 2855–2870, 2020. doi: 10.1007/s11063-020-10234-7.

[12] Huang G.B.: Learning capability and storage capacity of two-hidden-layer feedforward networks, *IEEE Transactions on Neural Networks*, vol. 14(2), pp. 274–281, 2003. doi: 10.1109/tnn.2003.809401.

[13] Jabbar H., Khan R.Z.: Methods to avoid over-fitting and under-fitting in supervised machine learning (comparative study), *Computer Science, Communication and Instrumentation Devices*, vol. 70, 2015. doi: 10.3850/978-981-09-5247-1_017.

[14] Kumar M.S., Soundarya V., Kavitha S., Keerthika E.S., Aswini E.: Credit card fraud detection using random forest algorithm. In: *2019 3rd International Conference on Computing and Communications Technologies (ICCCT)*, pp. 149–153, IEEE, 2019. doi: 10.1109/iccct2.2019.8824930.

[15] Li Z., Liu G., Wang S., Xuan S., Jiang C.: Credit Card Fraud Detection via Kernel-Based Supervised Hashing. In: *2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, pp. 1249–1254, IEEE, 2018. doi: 10.1109/smartworld.2018.00217.

[16] Makki S., Assaghir Z., Taher Y., Haque R., Hacid M.S., Zeineddine H.: An experimental study with imbalanced classification approaches for credit card fraud detection, *IEEE Access*, vol. 7, pp. 93010–93022, 2019. doi: 10.1109/access.2019.2927266.

[17] Mohammed R.A., Wong K.W., Shiratuddin M.F., Wang X.: Scalable Machine Learning Techniques for Highly Imbalanced Credit Card Fraud Detection: A Comparative Study. In: X. Geng, B.H. Kang (eds.), *PRICAI 2018: Trends in Artificial Intelligence. PRICAI 2018,* Lecture Notes in Computer Science, pp. 237–246, Springer, Cham, 2018. doi: 10.1007/978-3-319-97310-4_27.

[18] Mubarek A.M., Adalĭ E.: Multilayer perceptron neural network technique for fraud detection. In: *2017 International Conference on Computer Science and Engineering (UBMK)*, pp. 383–387, IEEE, 2017. doi: 10.1109/ubmk.2017.8093417.

[19] Prusti D., Rath S.K.: Web service based credit card fraud detection by applying machine learning techniques. In: *TENCON 2019-2019 IEEE Region 10 Conference (TENCON)*, pp. 492–497, IEEE, 2019. doi: 10.1109/tencon.2019.8929372.

[20] Sadgali I., Nawal S., Benabbou F.: Fraud detection in credit card transaction using machine learning techniques. In: *2019 1st International Conference on Smart Systems and Data Science (ICSSD)*, pp. 1–4, IEEE, 2019. doi: 10.1109/icssd47982.2019.9002674.

[21] Sahin Y., Duman E.: Detecting credit card fraud by ANN and logistic regression. In: *2011 International Symposium on Innovations in Intelligent Systems and Applications*, pp. 315–319, IEEE, 2011. doi: 10.1109/inista.2011.5946108.

[22] Saraswathi E., Kulkarni P., Khalil M.N., Nigam S.C.: Credit card fraud prediction and detection using artificial neural network and self-organizing maps. In: *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)*, pp. 1124–1128, IEEE, 2019. doi: 10.1109/iccmc.2019.8819758.

[23] Sohony I., Pratap R., Nambiar U.: Ensemble learning for credit card fraud detection. In: *CODS-COMAD '18: Proceedings of the ACM India Joint International Conference on Data Science and Management of Data*, pp. 289–294, 2018. doi: 10.1145/3152494.3156815.

[24] Srivastava N., Hinton G., Krizhevsky A., Sutskever I., Salakhutdinov R.: Dropout: A Simple Way to Prevent Neural Networks from Overfitting, *The Journal of Machine Learning Research*, vol. 15(56), pp. 1929–1958, 2014. http://jmlr.org/papers/v15/srivastava14a.html.

[25] Taha A.A., Malebary S.J.: An intelligent approach to credit card fraud detection using an optimized light gradient boosting machine, *IEEE Access*, vol. 8, pp. 25579–25587, 2020. doi: 10.1109/access.2020.2971354.

[26] Xinwei Z., Yaoci H., Xu W., Qili W.: HOBA: A novel feature engineering methodology for credit card fraud detection with a deep learning architecture, *Information Sciences*, vol. 557, pp. 302–316, 2021. doi: 10.1016/j.ins.2019.05.023.

[27] Yen S.J., Lee Y.S.: Under-sampling approaches for improving prediction of the minority class in an imbalanced dataset. In: *Intelligent Control and Automation: International Conference on Intelligent Computing, ICIC 2006 Kunming, China, August 16–19, 2006*, pp. 731–740, Springer, 2006.

[28] Yeşilkanat A., Bayram B., Köroğlu B., Arslan S.: An adaptive approach on credit card fraud detection using transaction aggregation and word embeddings. In: I. Maglogiannis, L. Iliadis, E. Pimenidis (eds.), *Artificial Intelligence Applications and Innovations. AIAI 2020. IFIP Advances in Information and Communication Technology*, pp. 3–14, Springer, 2020. doi: 10.1007/978-3-030-49161-1_1.

[29] Ying X.: An overview of overfitting and its solutions, *Journal of Physics: Conference Series*, vol. 1168(2), 22022, 2019. doi: 10.1088/1742-6596/1168/2/022022.

[30] Zamini M., Montazer G.: Credit card fraud detection using autoencoder based clustering. In: *2018 9th International Symposium on Telecommunications (IST)*, pp. 486–491, IEEE, 2018. doi: 10.1109/istel.2018.8661129.

Affiliations

Nirupam Shome
Assam University, Department of Electronics and Communication Engineering, Silchar 788011, India, nirupam-shome@yahoo.com

Devran Dey Sarkar
Assam University, Department of Electronics and Communication Engineering, Silchar 788011, India, devransarkar@gmail.com

Richik Kashyap
Assam University, Department of Electronics and Communication Engineering, Silchar 788011, India, rknits2010@gmail.com

Rabul Hussain Laskar
National Institute of Technology, Department of Electronics and Communication Engineering, Silchar, 788010, India, rhlaskar@ece.nits.ac.in

Received: 08.12.2023
Revised: 06.02.2024
Accepted: 20.02.2024