Contents lists available at Science-Gate

International Journal of Advanced and Applied Sciences

Journal homepage: http://www.science-gate.com/IJAAS.html

An enhanced AI-based model for financial fraud detection

Ahmed H. Ali [1, *], Ahmed Ali Hagag [2]

[1]Department of Electrical Quantities Metrology, National Institute of Standards (NIS), Giza, Egypt
[2]Ministry of Communication and Information Technology, Giza, Egypt

ARTICLE INFO

Article history:
Received 29 April 2024
Received in revised form
23 August 2024
Accepted 3 October 2024

Keywords:
Fraud detection
Machine learning
Supervised learning
Credit card transactions
Accuracy

ABSTRACT

The research seeks to identify complex fraudulent activities. Artificial intelligence (AI) techniques, such as machine learning and deep learning, have shown significant potential in enhancing the accuracy and efficiency of fraud detection models. This study introduces a novel AI-based fraud detection model that combines both supervised and unsupervised learning methods. The proposed machine learning system uses these techniques to detect fraudulent transactions. The supervised learning component is trained using a labeled dataset that includes both fraudulent and non-fraudulent transactions. The dataset used in the research contains 284,807 credit card transactions. After preparing the data, four Python-based models were developed. The K-Nearest Neighbors (KNN) model successfully predicted 99.94% of credit card transactions as valid or fraudulent. A random forest (RF) model was also used to assess the legitimacy of transactions, achieving an accuracy score of 99.96% correctly classifying nearly all data points. The Support Vector Machine (SVM) model achieved 99.94% accuracy, misclassifying only 51 cases. The logistic regression (LR) model attained an accuracy of 99.92% with 70 misclassifications and 99.91% with 77 misclassifications. These models demonstrate high accuracy and efficiency.

1. Introduction

Detecting fraudulent activities in the financial sector is critical, but traditional rule-based systems have limitations in detecting complex fraud. The utilization of artificial intelligence (AI) methods, specifically machine learning and deep learning, has demonstrated promise in enhancing the precision and effectiveness of fraud detection processes. Our research introduces a novel fraud detection model that utilizes AI techniques to combine supervised and unsupervised learning methods (Kolachalama and Garg, 2018).

In recent years, machine learning-based systems have become increasingly popular in identifying various types of fraud, as AI has shown significant success in other fields (Ngai et al., 2011; Phua et al., 2010). However, each detection model has certain limitations. One common challenge is dealing with imbalanced datasets, where fraud cases are often much fewer than non-fraud cases (He and Garcia, 2009). To address this, techniques such as under-sampling the non-fraud cases or over-sampling the fraud cases can be applied. For instance, SMOTE (Synthetic Minority Over-sampling Technique) is frequently used to create synthetic fraud cases to balance the dataset (Chawla et al., 2002).

After balancing the dataset using these methods, machine learning models, including neural networks, can be implemented to detect fraud effectively (Fiore et al., 2019; West et al., 2005). By addressing the imbalance in data, these models improve in both accuracy and reliability, ultimately enhancing fraud detection outcomes.

Studies have shown that when using an oversampled dataset, neural networks may predict fewer fraud transactions accurately compared to models trained on under-sampled datasets (He and Garcia, 2009). However, the under-sampling approach can fail to correctly identify a significant number of non-fraudulent transactions, often mislabeling them as fraudulent. This misclassification can result in blocking legitimate customers' accounts, potentially leading to customer complaints, diminished trust, and financial losses for banks or corporations (Ngai et al., 2011).

To address these issues, this study applies outlier detection techniques and removes outliers from the oversampled dataset. This approach aims to enhance

* Corresponding Author.
Email Address: ahmed.hussien@nis.sci.eg (A. H. Ali)
https://doi.org/10.21833/ijaas.2024.10.013
Corresponding author's ORCID profile:
https://orcid.org/0000-0001-8141-750X
2313-626X/© 2024 The Authors. Published by IASE.

detection accuracy by reducing misclassifications, thus improving model reliability and maintaining customer trust in financial services.

Financial institutions face a major challenge in detecting and preventing fraud, as it can lead to serious financial losses and harm to their reputation. Traditional fraud detection systems, based on fixed rules, often struggle to keep up with quickly changing fraud methods and tend to produce many false positives, wasting resources. Thus, a new and improved fraud detection model that uses AI is needed to identify fraud accurately while reducing false positives. This study seeks to answer the following questions: How can an advanced fraud detection model be developed using AI techniques?

The aim of this research is to develop and evaluate an advanced fraud detection system using AI techniques to help financial institutions reduce fraud risks and protect customer assets. The system is designed to learn from past data, detect fraud in real-time, and lower both false positives and negatives, thereby minimizing financial losses. Traditional rule-based systems struggle to keep up with increasingly complex fraud tactics, resulting in many false positives and negatives. Alongside these objectives, the research seeks to enhance the speed and accuracy of fraud detection in financial institutions by utilizing AI algorithms and machine learning models. Another goal is to improve customer experience by reducing false alerts in the fraud detection process.

2. Related works

Numerous studies have focused on detecting credit card fraud (CCF). This section reviews several studies related to CCF detection, with a particular emphasis on research addressing fraud detection in cases of category imbalance. Various techniques are used for credit card fraud detection, and the main approaches can be grouped into deep learning (DL), machine learning (ML), CCF detection, clustering and feature ordering, and user authentication methods (Abakarim et al., 2018). The payment card authorization process typically involves two types of authentication: password-based and biometrics-based. Biometrics-based authentication can be further divided into three types: physiological authentication, behavioral authentication, and a combination of both (Balogun et al., 2019).

Chen and Lai (2021) examined how the continuous use of the Internet in organizations has enabled online banking services, contributing to financial losses due to global increases in financial fraud. With advancements in technology, fraud detection systems can now identify risks such as unauthorized transactions and irregular attacks more effectively. In recent years, data mining and machine learning techniques have been widely applied to address these issues. However, improvements are still needed in areas such as identifying unknown attack patterns, enhancing big data analytics, and increasing computational speed. The study focuses on detecting financial fraud using deep learning algorithms, demonstrating that detection accuracy can be improved with large datasets. The proposed model was compared with existing machine learning and autoencoder models using a real-time credit card fraud dataset, achieving 99% accuracy with detection times of around 45 seconds, as reported by Alfaiz and Fati (2022).

The COVID-19 pandemic has limited physical purchases, making online transactions—and thus credit card fraud—a critical issue in online banking. This study highlights the need for robust fraud detection methods. The researcher tested 66 machine-learning models on a real dataset of European cardholders using stratified K-fold cross-validation. In the first phase, nine machine learning algorithms were assessed, and the top three were further evaluated with 19 resampling techniques. Among the 330 evaluations, the All K-Nearest Neighbors (AllKNN) under-sampling technique combined with CatBoost (AllKNN-CatBoost) emerged as the best model, achieving an AUC of 97.94%, recall of 95.91%, and F1-Score of 87.40%, outperforming previous models.

Taha and Malebary (2020) emphasized the importance of combating credit card fraud and discussed common techniques used to counter this issue. Financial institutions and banks not only provide convenient financial services but also actively work to protect credit cards from fraud. They invest in and develop various technologies, including advanced machine learning systems, which are central to many fraud detection processes. One method employed is Decision Tree (DT), which is easy to implement but requires verification for each transaction. The authors analyzed different models using an imbalanced European Credit Card Fraud Detection (ECCFD) dataset without applying any resampling techniques. Their results showed that DT performed best overall, achieving a recall of 79.21%, a precision of 85.11%, and a quick processing time of 5 seconds. In comparison, K-Nearest Neighbors (KNN) showed higher recall (81.19%) and precision (91.11%) but required much more time (463 seconds).

Another research direction focuses on Logistic Regression (LR) and KNN. Vengatesan et al. (2020) assessed the performance of LR and KNN on an imbalanced European Credit Card Fraud Detection (ECCFD) dataset, finding that KNN achieved the highest accuracy at 95%, with a recall of 72% and an F1 score of 82%. Additionally, Puh and Brkić (2019) analyzed the performance of algorithms such as Random Forest (RF), Support Vector Machine (SVM), and LR on a dataset of European cardholders. They addressed data imbalance using the Synthetic Minority Oversampling Technique (SMOTE) and used LR with modified algorithm parameters, setting the LR parameter C to 100 and applying L2 regularization. Two models were created with LR: one using continuous learning and the other using incremental learning. The results showed an AUC score of 91.14% for continuous learning and 91.07%

for incremental learning. The average accuracy was 73.37% with continuous learning and 84.13% with incremental learning.

Trivedi et al. (2020) examined how advancements in technology and faster communication have contributed to a significant increase in credit card fraud. Credit card fraud detection has become a crucial topic in financial analysis due to the millions of dollars in annual losses it causes for both consumers and financial institutions. Fraudsters continuously create new methods to commit illegal activities, making fraud prevention techniques essential to reducing losses for banks and financial institutions. In this study, the authors propose an effective automated credit card fraud detection method that includes a feedback system based on machine learning. This feedback approach improves the classifier's detection rate and cost-effectiveness.

The study evaluated various methodologies, including RFs, tree classifiers, artificial neural networks, SVM, naive Bayes, LR, and gradient boosting classifiers, using a slightly imbalanced dataset of 284,807 credit card transactions from European account holders. The methods were applied to both raw and pre-processed data. The effectiveness of each approach was measured based on performance metrics such as precision, recall, F1 score, and false positive rate (FPR).

ML encompasses various branches, each designed to handle specific learning tasks. ML frameworks include distinct types of approaches, each suited to different detection challenges. CCF detection, for instance, the RF method is a widely used solution. RF, an ensemble technique that combines multiple DTs, is popular among researchers due to its robustness and accuracy in classification tasks (Breiman, 2001).

To enhance model performance further, combined approaches like the APATE method integrate RF with network analysis, enabling more comprehensive fraud detection (Arora et al., 2020). This multi-model approach leverages the strengths of both ML classification and network-based insights for improved detection outcomes.

Kim et al. (2019) discussed DL algorithms, including convolutional neural networks (CNNs), deep belief networks (DBNs), and deep autoencoders, as powerful tools for data processing, feature learning, and pattern classification. DL aims to explore artificial neural networks, typically relying on the backpropagation model, which is affected by the network's depth (Kousika et al., 2021). However, as network depth increases, issues like insufficient local minima and error dilution can arise, reducing the backpropagation algorithm's efficiency. Deep architectures, such as deep belief networks, are effective for addressing optimization challenges in training parameters.

Traditional ML algorithms, including SVM, DT, and LR, have been widely applied to CCF detection (Ngai et al., 2011). However, these methods often face scalability issues with large datasets. In contrast, CNNs, a DL approach, are suited for handling three-dimensional data, commonly used in image processing tasks. Unlike Artificial Neural Networks (ANNs), CNNs contain convolutional layers with multiple channels, enabling them to capture spatial hierarchies in data (Krizhevsky et al., 2012).

CNNs are especially popular for image processing applications because they require minimal pre-processing and are effective at retaining key features through techniques like feature maps, channels, pooling, stride, and padding. Typically, 2D CNN models are applied to two-dimensional data, including text, images, and video, leveraging feature mapping to learn internal representations independent of feature location (LeCun et al., 2015). For one-dimensional data, 1D CNNs can be employed, which are commonly used in natural language processing (NLP) tasks, particularly for sequence classification challenges.

Lucas and Jurgovsky (2020) explained that in one-dimensional convolutional neural networks (1D-CNN), the kernel filter moves sequentially through data samples from top to bottom, while in two-dimensional CNNs (2D-CNN), it moves both from left to right and top to bottom. Deep belief networks are often viewed as one of the most effective techniques for training deep architectures, whereas traditional machine learning algorithms such as SVM, DT, and LR are less suited for large datasets. CNNs, as deep learning methods, are widely used for processing three-dimensional data, especially in image processing tasks. Structurally similar to artificial neural networks (ANNs), CNNs include hidden layers and varying numbers of channels with specialized convolutional layers.

CNNs are advantageous in image processing, as they require minimal pre-processing while preserving essential features through image reduction for prediction purposes. Key terms in CNNs include feature maps, channels, pooling, stride, and padding. The 2D-CNN is typically used for text, image, and video processing, as it processes two-dimensional data inputs.

The feature mapping process is used to learn internal representations from input data, which can also be applied to one-dimensional data, such as in NLP, where sequence classification poses a challenge (Matloob et al., 2020). Autoencoders are neural networks trained to encode and decode data, identifying anomalous points and classifying transactions as fraud or non-fraud based on reconstruction error. Generative adversarial networks (GANs) consist of two neural networks that work together to enhance prediction accuracy, often through unsupervised learning in a zero-sum game framework. GANs are a key category in deep learning models with promising potential for advancements in DL (Molina et al., 2018).

The DL model structure includes two main modules: a generator (G) and a discriminator (D). During training, these modules form a neural network where the generator creates simulated data, and the discriminator assesses this data against the target data, distinguishing between simulated and real data. Variational Autoencoders (VAEs) are a type of autoencoder with regularized training distribution, ensuring adequate hidden space resources and enabling the generation of new data. Long Short-Term Memory (LSTM) networks, a type of recurrent neural network (RNN), are commonly used in DL models for processing and predicting time-sequence data, addressing the vanishing gradient problem that often limits RNNs to short-term memory.

DL methods such as CNN and LSTM are recommended for handling tasks like image classification, NLP, and Restricted Boltzmann Machines (RBM) to manage large datasets. Additionally, the effect of data pre-processing on classification performance in detecting credit card fraud remains an area requiring further investigation.

Researchers can use various ML techniques, including supervised and unsupervised techniques. Common ML algorithms, such as LR, ANN, DT, SVM, and NB, are used for CCF detection. These techniques can be combined with ensemble techniques to construct robust detection classifiers (Arora et al., 2020). An artificial neural network involves linking multiple neurons and nodes. A feed-forward perceptron multilayer comprises multiple layers, including an input layer, an output layer, and one or more hidden layers. The input nodes represent the exploratory variables, and their precise weights are multiplied. Each hidden layer node is transferred with a certain bias and added together. An activation function is applied to create the output of each neuron for this summation, and the result is then transferred to the next layer. Finally, the output layer provides the algorithm's response. The weights are initially set randomly and then adjusted using algorithms such as backpropagation (Błaszczyński et al., 2021).

The Bayesian belief network is a graphical model that represents the contingency relationships between a set of variables. The independence assumption in naive Bayes is also used. The Bayesian belief network is a graphical model that represents the contingency relationships between a set of variables, while the independence assumption in naive Bayes allows for dependencies among variables. Quantity variables are represented as nodes, and the dependencies of conditions between variables are shown as arcs between nodes. The conditional probability table of each node is linked, which makes the possibilities of the node's variable conditional on the parent's node values (Branco et al., 2020).

The bilateral-branch network (BBN) computational system involves establishing a network structure, which may depend on specific algorithms using the given data. After determining the network topology, historical data is used to fit the network, with continuous variables discretized and assumed to follow a normal distribution. In BBN, each node is considered independent of nodes outside its descendants, conditional on its parent nodes within the graph—a principle known as the Markov condition (Lad and Adamuthe, 2020). SVM is a linear classification model commonly used for regression problems. According to the SVM algorithm, the points closest to the line from each class are identified as support vectors (Cartella et al., 2021). This paper emphasizes the integration of unsupervised and supervised techniques for classifying CCF detection.

3. Materials and methods

Several essential steps are required to develop an improved AI-based model for credit card fraud detection. The process begins with acquiring relevant data, followed by data preparation and pre-processing. Once the data is verified as suitable for modeling, the modeling phase can begin. Four models—KNN, RF, SVM, and LR—are created during this phase. Python is used to implement all four models, including the KNN model.

3.1. Dataset

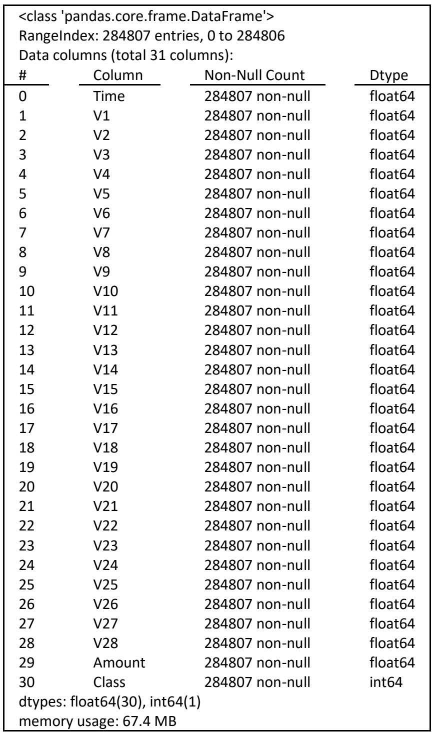
The credit card transaction dataset consists of 284,807 transactions by European cardholders from September 2013, with 492 labeled as fraudulent (Fig. 1). The dataset is highly imbalanced, as fraud accounts for only 0.172% of all transactions. Due to this imbalance, accuracy measurements based on the confusion matrix are not meaningful; instead, accuracy should be assessed using the area under the precision-recall curve. The dataset includes only numerical features derived through Principal Component Analysis (PCA), with features labeled as V1, V2, ... V28. The 'Time' and 'Amount' features were not transformed through PCA, where 'Time' represents the elapsed seconds between each transaction and the dataset's first transaction, and 'Amount' indicates the transaction value. The 'Class' feature is the target variable, labeled as 1 for fraud and 0 for non-fraud.

To effectively utilize this dataset, machine learning models need to be developed and tested to detect fraud accurately while minimizing false positives. Additionally, feature engineering and selection techniques can be applied to improve model performance.

3.2. Data preprocessing

Fig. 2 shows the distribution of the target variable. Non-fraudulent transactions are the most common, comprising 284,315 instances, while fraudulent transactions are much less frequent, with only 492 instances.

The Pearson Correlation method is used to visualize the correlations between features and the target variable Class. The predictor columns do not show high correlation values with each other or with the Class column. However, there is a negative correlation between V2 and Amount and a positive correlation between V7 and Amount. Fig. 3 displays the correlation between features and the target variable Class.



```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 284807 entries, 0 to 284806
Data columns (total 31 columns):
#     Column      Non-Null Count      Dtype
0     Time        284807 non-null     float64
1     V1          284807 non-null     float64
2     V2          284807 non-null     float64
3     V3          284807 non-null     float64
4     V4          284807 non-null     float64
5     V5          284807 non-null     float64
6     V6          284807 non-null     float64
7     V7          284807 non-null     float64
8     V8          284807 non-null     float64
9     V9          284807 non-null     float64
10    V10         284807 non-null     float64
11    V11         284807 non-null     float64
12    V12         284807 non-null     float64
13    V13         284807 non-null     float64
14    V14         284807 non-null     float64
15    V15         284807 non-null     float64
16    V16         284807 non-null     float64
17    V17         284807 non-null     float64
18    V18         284807 non-null     float64
19    V19         284807 non-null     float64
20    V20         284807 non-null     float64
21    V21         284807 non-null     float64
22    V22         284807 non-null     float64
23    V23         284807 non-null     float64
24    V24         284807 non-null     float64
25    V25         284807 non-null     float64
26    V26         284807 non-null     float64
27    V27         284807 non-null     float64
28    V28         284807 non-null     float64
29    Amount      284807 non-null     float64
30    Class       284807 non-null     int64
dtypes: float64(30), int64(1)
memory usage: 67.4 MB
```

Fig. 1: Dataset structure

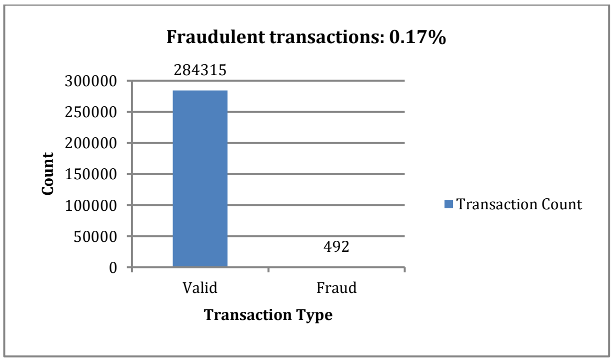


Fig. 2: Class distribution

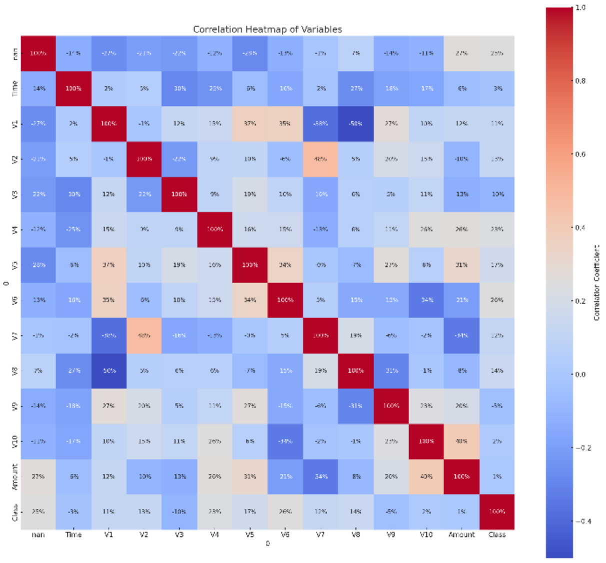


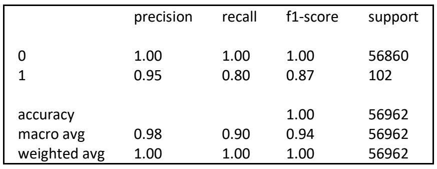
Fig. 3: Correlations between features and the target variable class

3.3. Model evolution

After preparing the data for modeling, four models were created using Python.

3.3.1. KNN

KNN model achieved an accuracy of 99.94% in predicting whether credit card transactions were valid or fraudulent. It accurately classified 100% of valid transactions and 91% of fraudulent ones. KNN, a supervised machine learning algorithm, can perform both classification and regression tasks; here, it was used for classification to identify fraudulent transactions. To determine the optimal K value, experiments were conducted with K=3 and K=7, ultimately selecting K=9 for the final predictions, which produced an impressive accuracy of 99.94%. However, besides accuracy, it is crucial to evaluate additional performance metrics and conduct further data analysis to ensure the model is neither overfitting nor underfitting. Fig. 4 illustrates the KNN model's 99.94% accuracy in classifying transactions as valid or fraudulent.

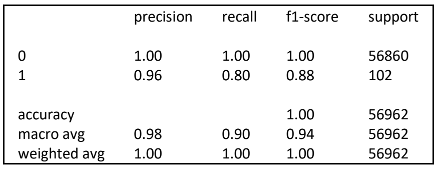


```
              precision    recall    f1-score    support

0             1.00         1.00      1.00        56860
1             0.95         0.80      0.87        102

accuracy                             1.00        56962
macro avg     0.98         0.90      0.94        56962
weighted avg  1.00         1.00      1.00        56962
```

Fig. 4: The KNN model accurately predicted 99.94% of the transactions as being valid or fraudulent

3.3.2. RF model

RF model was used to assess the legitimacy of credit card transactions, achieving an accuracy of 99.96%. This indicates that the model successfully classified all transactions in the dataset, correctly identifying 100% of valid transactions and 94% of fraudulent ones. The RF model was chosen over LR for this task, although LR was also applied to predict the validity of credit card transactions. The LR model attained an accuracy of 99.92%, accurately classifying 100% of valid transactions but only 91% of fraudulent transactions. Fig. 5 illustrates the RF model's 99.96% accuracy in predicting transactions as valid or fraudulent.

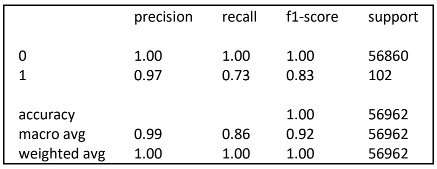


```
              precision    recall    f1-score    support

0             1.00         1.00      1.00        56860
1             0.96         0.80      0.88        102

accuracy                             1.00        56962
macro avg     0.98         0.90      0.94        56962
weighted avg  1.00         1.00      1.00        56962
```

Fig. 5: The RF model accurately predicted 99.96% of the transactions as being valid or fraudulent

3.3.3. SVM

SVM is a supervised machine learning method that uses continuous learning algorithms to perform classification and regression tasks. SVM can handle both linear and non-linear classification by maximizing the margins between classes to minimize classification errors (Mahesh, 2020). In this context, the SVM model achieved an accuracy of 99.94%, with only 51 misclassified cases. When predicting transaction authenticity, the model achieved 99.93% accuracy, correctly identifying 100% of legitimate transactions and 93% of fraudulent ones. Although the SVM model demonstrates high accuracy with a misclassification rate of only 0.06%, it may still miss some fraudulent cases. Therefore, combining the SVM model with additional fraud detection techniques may further enhance its performance. Fig. 6 illustrates the SVM model's 99.93% accuracy in classifying transactions as valid or fraudulent.



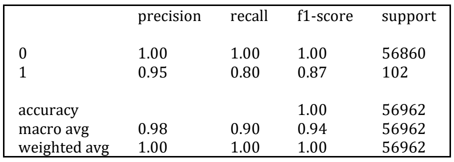
```
              precision    recall  f1-score   support

0             1.00         1.00    1.00       56860
1             0.97         0.73    0.83       102

accuracy                           1.00       56962
macro avg     0.99         0.86    0.92       56962
weighted avg  1.00         1.00    1.00       56962
```

Fig. 6: The SVM model accurately predicted 99.93% of the transactions as being valid or fraudulent

3.3.4. LR

LR is a statistical model used to assess the relationship between a binary or categorical dependent variable and one or more independent variables, which can be either qualitative or quantitative (Domínguez-Almendros et al., 2011). In this case, the LR model achieved an accuracy of 99.92% with 70 misclassifications, and 99.91% accuracy with 77 misclassifications. It correctly classified 100% of legitimate transactions and accurately identified 91% of fraudulent transactions.

Although the LR model shows high accuracy in predicting transaction authenticity with a misclassification rate of only 0.08%, it may still miss some fraudulent cases. To enhance performance, combining the LR model with additional fraud detection techniques may be beneficial. Fig. 7 illustrates the model's 99.91% accuracy in distinguishing between valid and fraudulent transactions.



```
              precision    recall  f1-score   support

0             1.00         1.00    1.00       56860
1             0.95         0.80    0.87       102

accuracy                           1.00       56962
macro avg     0.98         0.90    0.94       56962
weighted avg  1.00         1.00    1.00       56962
```

Fig. 7: The LR model accurately predicted 99.91% of the transactions as being valid or fraudulent

4. Results

To identify the optimal model for detecting fraudulent credit card transactions, all models were compared against each other. Precision was used as the primary metric to evaluate the total number of correctly predicted instances, as represented in the confusion matrix. The confusion matrix includes four components: true positive (TP), true negative (TN), false positive (FP), and false negative (FN).

TP represents fraudulent transactions that were accurately identified as fraudulent. TN represents non-fraudulent transactions that were accurately identified as non-fraudulent. FP occurs when a non-fraudulent transaction is misclassified as fraudulent. FN occurs when a fraudulent transaction is misclassified as non-fraudulent.

Table 1 presents the components needed to compute model accuracy, which can be calculated using the following equation:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

Table 1: Confusion matrix

Predicted	Positive	Negative
Positive	True positive	False negative
Negative	False positive	True negative

Table 2 presents the accuracy scores of the models developed for detecting fraudulent transactions, all of which performed exceptionally well, with accuracy rates exceeding 99%. The RF model achieved the highest accuracy at 99.96%, making it the best-performing model in this study. This model successfully classified the majority of transactions, both valid and fraudulent. KNN model followed with the second-highest accuracy of 99.94%.

Table 2: Accuracy scores

Model		Accuracy
RF	RF	99.96%
LR	LR	99.91%
SVM	SVM	99.93%
KNN	KNN	99.94%

The SVM model ranked third with an accuracy of 99.93%. Both the KNN and SVM models demonstrated strong performance, highlighting their robustness in managing the complexities of fraud detection. Fig. 8 displays the accuracy scores for each model in the proposed approach. The LR model, while performing well, had the lowest accuracy among the models tested, with a score of 99.91%. However, the accuracy difference across models is minimal, indicating that all are highly capable of accurately detecting fraudulent transactions. Despite the high accuracy of each model, it is important to consider other factors, such as handling class imbalance, computational efficiency, and real-world applicability. The RF model's superior accuracy suggests it may be the best choice for this dataset. Yet, the slight accuracy variations among models highlight that factors like speed and resource demands may also influence the final selection.

Additionally, addressing the highly imbalanced data, composed mostly of valid transactions, is crucial. Future research could investigate techniques such as resampling or synthetic data generation to create more balanced datasets, thereby enhancing the models' robustness for real-world applications.

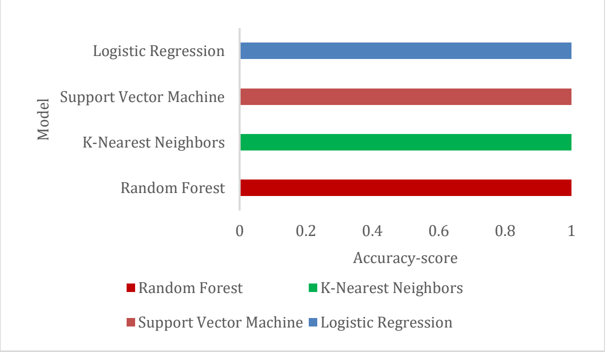


Fig. 8: Accuracy scores by model

5. Discussion

Numerous studies have focused on identifying CCF. This section reviews various research papers on CCF detection, particularly those addressing fraud detection under conditions of class imbalance. Several approaches are commonly used to detect credit card fraud. To examine the most relevant studies, core approaches can be classified into DL, ML, clustering, feature ranking, CCF detection, and user authentication methods (Abakarim et al., 2018), including an authorization process. For credit card authentication, there are two primary methods: password-based and biometric authentication, with the latter divided into physiological, behavioral, and combined authentication types (Balogun et al., 2019). Chen and Lai (2021) noted that with the growing use of the Internet in enterprises, online banking services have expanded, leading to an increase in financial crime. As advancements in financial technology and fraud detection continue, systems are now more effective in identifying threats, including unauthorized and irregular transactions. Data mining and machine learning techniques have been applied to tackle these issues; however, additional improvements are necessary to manage challenges like identifying unknown attack patterns, big data analytics, and computational speed constraints.

The deep neural network (DCNN) approach has been applied to address these challenges using data mining and machine learning. Yet, given the current limitations in identifying previously undiscovered attack patterns and handling big data, further development is required. In this study, all models achieved excellent results in predicting both valid and fraudulent transactions, with accuracy rates exceeding 99%. The RF model achieved the highest accuracy at 99.96%, correctly classifying the majority of transactions. However, since the data was heavily weighted towards valid transactions, a more balanced dataset could enhance model performance.

The manuscript proposes an enhanced fraud detection model using AI, specifically deep learning techniques, to improve the accuracy of detecting fraud in financial transactions. While it provides an in-depth overview of different fraud detection approaches, a comparative analysis of the proposed model against current state-of-the-art systems in real-world settings would further strengthen the study. Presently, many state-of-the-art systems employ machine learning and deep learning models, such as DTs, RFs, SVMs, RNNs, and CNNs, to analyze transaction data and identify fraud patterns (Kumar et al., 2018; Ahmad et al., 2020).

Compared to these systems, the proposed model offers several benefits. It combines feature engineering with deep learning techniques, such as CNNs and LSTMs, to analyze both historical and real-time transaction data, enabling it to identify complex patterns and anomalies that traditional machine learning models may miss.

6. Conclusion

In conclusion, this research introduced an AI-based fraud detection model that integrates both supervised and unsupervised learning techniques. The model was trained on a labeled dataset containing 284,807 credit card transactions, both fraudulent and non-fraudulent. Four Python-based models were developed and evaluated for their effectiveness in identifying fraudulent transactions.

KNN model achieved an impressive accuracy of 99.94%, accurately predicting the legitimacy of credit card transactions. The RF model performed even better, with an accuracy of 99.96%, demonstrating its capability to classify nearly all parameters in the dataset. The SVM model also showed strong performance, reaching an accuracy of 99.94% with only 51 misclassifications. LR attained accuracies of 99.92% and 99.91%, though with slightly higher misclassifications (70 and 77, respectively).

Overall, the proposed AI-based fraud detection models demonstrated high accuracy and effectiveness in identifying complex fraudulent activities in credit card transactions. These models hold substantial potential for enhancing the accuracy of fraud detection systems, which could help financial institutions and customers avoid significant losses.

Compliance with ethical standards

Conflict of interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Abakarim Y, Lahby M, and Attioui A (2018). An efficient real time model for credit card fraud detection based on deep learning. In the Proceedings of the 12th International Conference on Intelligent Systems: Theories and Applications, ACM, Rabat, Morocco: 1-7. https://doi.org/10.1145/3289402.3289530

Alfaiz NS and Fati SM (2022). Enhanced credit card fraud detection model using machine learning. Electronics, 11(4): 662. https://doi.org/10.3390/electronics11040662

Arora V, Leekha RS, Lee K, and Kataria A (2020). Facilitating user authorization from imbalanced data logs of credit cards using artificial intelligence. Mobile Information Systems, 2020: 8885269. https://doi.org/10.1155/2020/8885269

Balogun AO, Basri S, Abdulkadir SJ, and Hashim AS (2019). Performance analysis of feature selection methods in software defect prediction: A search method approach. Applied Sciences, 9(13): 2764. https://doi.org/10.3390/app9132764

Błaszczyński J, de Almeida Filho AT, Matuszyk A, Szeląg M, and Słowiński R (2021). Auto loan fraud detection using dominance-based rough set approach versus machine learning methods. Expert Systems with Applications, 163: 113740. https://doi.org/10.1016/j.eswa.2020.113740

Branco B, Abreu P, Gomes AS, Almeida MS, Ascensão JT, and Bizarro P (2020). Interleaved sequence RNNs for fraud detection. In the Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, CA, USA: 3101-3109. https://doi.org/10.1145/3394486.3403361

Breiman L (2001). Random forests. Machine Learning, 45: 5-32. https://doi.org/10.1023/A:1010933404324

Cartella F, Anunciacao O, Funabiki Y, Yamaguchi D, Akishita T, and Elshocht O (2021). Adversarial attacks for tabular data: Application to fraud detection and imbalanced data. Arxiv Preprint Arxiv:2101. https://doi.org/10.48550/arXiv.2101.08030

Chawla NV, Bowyer KW, Hall LO, and Kegelmeyer WP (2002). SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16: 321-357. https://doi.org/10.1613/jair.953

Chen JIZ and Lai KL (2021). Deep convolution neural network model for credit-card fraud detection and alert. Journal of Artificial Intelligence, 3(2): 101-112. https://doi.org/10.36548/jaicn.2021.2.003

Domínguez-Almendros S, Benítez-Parejo N, and Gonzalez-Ramirez AR (2011). Logistic regression models. Allergologia et Immunopathologia, 39(5): 295-305. https://doi.org/10.1016/j.aller.2011.05.002 **PMid:21820234**

Fiore U, De Santis A, Perla F, Zanetti P, and Palmieri F (2019). Using generative adversarial networks for improving classification effectiveness in credit card fraud detection. Information Sciences, 479: 448-455. https://doi.org/10.1016/j.ins.2017.12.030

He H and Garcia EA (2009). Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering, 21(9): 1263-1284. https://doi.org/10.1109/TKDE.2008.239

Kim J, Kim HJ, and Kim H (2019). Fraud detection for job placement using hierarchical clusters-based deep neural networks. Applied Intelligence, 49(8): 2842-2861. https://doi.org/10.1007/s10489-019-01419-2

Kolachalama VB and Garg PS (2018). Machine learning and medical education. NPJ Digital Medicine, 1: 54. https://doi.org/10.1038/s41746-018-0061-1 **PMid:31304333 PMCid:PMC6550167**

Kousika N, Deepa S, Deephika C, Dhatchaiyine BM, and Amrutha J (2021). A system for fake news detection by using supervised learning model for social media contents. In the 5th International Conference on Intelligent Computing and Control Systems, IEEE, Madurai, India: 1042-1047. https://doi.org/10.1109/ICICCS51141.2021.9432096

Krizhevsky A, Sutskever I, and Hinton GE (2012). ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, 25: 1097–1105.

Lad SS and Adamuthe AC (2020). Malware classification with improved convolutional neural network model. International Journal of Computer Network and Information Security, 9(6): 30-43. https://doi.org/10.5815/ijcnis.2020.06.03

LeCun Y, Bengio Y, and Hinton G (2015). Deep learning. Nature, 521(7553): 436-444. https://doi.org/10.1038/nature14539 **PMid:26017442**

Lucas Y and Jurgovsky J (2020). Credit card fraud detection using machine learning: A survey. Arxiv Preprint Arxiv:2010.06479. https://doi.org/10.48550/arXiv.2010.06479

Mahesh B (2020). Machine learning algorithms-A review. International Journal of Science and Research, 9(1): 381-386. https://doi.org/10.21275/ART20203995

Matloob I, Khan SA, and Rahman HU (2020). Sequence mining and prediction-based healthcare fraud detection methodology. IEEE Access, 8: 143256-143273. https://doi.org/10.1109/ACCESS.2020.3013962

Molina D, LaTorre A, and Herrera F (2018). SHADE with iterative local search for large-scale global optimization. In the IEEE Congress on Evolutionary Computation, IEEE, Rio de Janeiro, Brazil: 1-8. https://doi.org/10.1109/CEC.2018.8477755

Ngai EW, Hu Y, Wong YH, Chen Y, and Sun X (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. Decision support systems, 50(3): 559-569. https://doi.org/10.1016/j.dss.2010.08.006

Phua C, Lee V, Smith K, and Gayler R (2010). A comprehensive survey of data mining-based fraud detection research. ArXiv Preprint ArXiv:1009.6119. https://doi.org/10.48550/arXiv.1009.6119

Puh M and Brkić L (2019). Detecting credit card fraud using selected machine learning algorithms. In the 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics, IEEE, Opatija, Croatia: 1250-1255. https://doi.org/10.23919/MIPRO.2019.8757212

Taha AA and Malebary SJ (2020). An intelligent approach to credit card fraud detection using an optimized light gradient boosting machine. IEEE Access, 8: 25579-25587. https://doi.org/10.1109/ACCESS.2020.2971354

Trivedi NK, Simaiya S, Lilhore UK, and Sharma SK (2020). An efficient credit card fraud detection model based on machine learning methods. International Journal of Advanced Science and Technology, 29(5): 3414-3424.

Vengatesan K, Kumar A, Yuvraj S, Kumar V, and Sabnis S (2020). Credit card fraud detection using data analytic techniques. Advances in Mathematics: Scientific Journal, 9(3): 1185-1196. https://doi.org/10.37418/amsj.9.3.43

West D, Dellana S, and Qian J (2005). Neural network ensemble strategies for financial decision applications. Computers and Operations Research, 32(10): 2543-2559. https://doi.org/10.1016/j.cor.2004.03.017