ISSN : 2455-9679
SJIF Impact Factor : 6.008

ijrtsm

IJRTSM

INTERNATIONAL JOURNAL OF RECENT TECHNOLOGY SCIENCE & MANAGEMENT

"MACHINE LEARNING APPROACHES FOR DETECTING CREDIT CARD FRAUD: A REVIEW OF ALGORITHMS AND APPLICATIONS"

Neha Bharti [1]***, Dr. S.K. Pandey*** [2]

[1] *M.Tech Scholar, Department of Computer Science and Engineering, VNS group of Institutions, Bhopal, Madhya Pradesh, India*

[2] *Department of Computer Science and Engineering, VNS group of Institutions, Bhopal, Madhya Pradesh, India*

nehukec@gmail.com

ABSTRACT

Credit card fraud continues to pose a major threat to global financial systems, resulting in significant economic losses and consumer mistrust. With the rapid rise in digital transactions, detecting fraudulent activities has become increasingly complex and essential. In recent years, machine learning (ML) has emerged as a powerful tool for identifying and preventing credit card fraud by learning patterns from vast transaction datasets. However, one of the primary challenges in this domain is data imbalance, where genuine transactions vastly outnumber fraudulent ones. This imbalance often leads to biased models and reduced detection rates for fraudulent cases. This paper presents a comprehensive review of existing machine learning techniques used for credit card fraud detection, emphasizing methods that effectively address data imbalance and improve overall model performance. It evaluates a wide range of supervised, unsupervised, and hybrid models, as well as ensemble methods and deep learning frameworks. Additionally, it explores data preprocessing strategies such as oversampling, undersampling, cost-sensitive learning, and anomaly detection. The study highlights each method's strengths and limitations, offering insights into best practices and future research directions. This review aims to support researchers and practitioners in developing robust, accurate, and adaptive fraud detection systems that can effectively respond to evolving fraud patterns in real-world scenarios.

Key Words: *Credit Card Fraud, Fraud Detection, Financial Security, Anomaly Detection, Machine Learning, Data Imbalance.*

I. INTRODUCTION

In today's digital age, the widespread use of credit cards has significantly transformed the financial ecosystem, offering unparalleled convenience in online and offline transactions. However, this increased reliance on electronic payment systems has also opened the door to a growing number of cybercrimes, most notably credit card fraud. Credit card fraud poses a severe threat to consumers, financial institutions, and the global economy, leading to billions of dollars in losses each year. As fraudsters continuously evolve their techniques, detecting and preventing fraudulent transactions has become an increasingly complex and critical challenge. Detection systems, capable of analyzing vast amounts of transactional data to identify patterns, anomalies, and suspicious activities in real time. ML algorithms can learn from historical data to distinguish between legitimate and fraudulent behaviors, thereby enhancing detection accuracy and reducing false positives. Despite their promise, applying machine learning to credit card fraud detection is not without challenges.

One of the most significant issues is data imbalance—a situation where the number of fraudulent transactions is minuscule compared to legitimate ones. This imbalance often leads to biased models that favor the majority class, resulting in poor performance in identifying the minority (fraudulent) class. Additionally, real-world fraud detection systems must contend with issues like concept drift (where fraud patterns change over time), feature selection, scalability, interpretability, and latency in prediction.

This review aims to provide a comprehensive analysis of machine learning techniques used in credit card fraud detection, with a particular focus on how they address data imbalance and improve model performance. It covers a wide range of supervised, unsupervised, and hybrid models, including decision trees, random forests, support vector machines, neural networks, ensemble learning, and deep learning approaches. Furthermore, it explores data preprocessing strategies, such as resampling (over-sampling and under-sampling), cost-sensitive learning, and anomaly detection methods designed to mitigate the impact of data imbalance.

Types of Fraud

This article covers a wide range of fraud categories, including: bankruptcy fraud; application fraud; behavioral fraud; theft/counterfeit; telecommunications fraud; computer intrusions; and credit card fraud.

- **Credit Card Fraud:** Offline and online credit card fraud are the two main categories.
- **Offline fraud** is perpetrated by making use of a counterfeit or stolen physical card at a variety of establishments.
- **On-line fraud** commits the crime while the cardholder is not present, over the phone, online, or while shopping.
- **Telecommunication Fraud:** in the absence of the cardholder, when transacted over the phone, online, or during a shopping trip.
- [4] Used a powerful generalized response model to predict management fraud. The "probit and logit" methods are a part of the model. Credit cards and their many varieties are defined at the outset of this paper, which then moves on to discuss relevant research and potential methods and models for identifying legitimate and fraudulent purchases.

Table 1 : Types of Credit Card Fraud

Type of Fraud	Description
Card Not Present (CNP)	Fraud where the physical card is not required, typically during online or phone transactions.
Lost/Stolen Card	Transactions made using a card that has been physically lost or stolen from the owner.
Counterfeit Card	Duplicate cards created using stolen card information via skimming or hacking.
Application Fraud	Fraudulent credit card accounts opened using fake or stolen identity information.
Account Takeover	Criminals gain access to a legitimate user’s account and change credentials to make unauthorized transactions.
Skimming	Card information is stolen using a device placed on ATMs or POS terminals.
Phishing	Fraudsters trick users into revealing card details via fake emails, websites, or messages.
Mail Theft	Theft of new credit cards or billing statements from postal mail.
Fake Merchant Fraud	Fraudsters set up fake businesses to charge stolen cards and withdraw funds.
Refund Fraud	A fraudster manipulates a return/refund process to get unauthorized credits.

Computer Intrusion: Intrusion Is Defined As

The Act Of Entering Without Warrant Or Invitation; That Means “Potential Possibility Of Unauthorized Attempt To Access Information, Manipulate Information Purposefully. Intruders May Be From Any Environment, An Outsider (Or Hacker) And An Insider Who Knows The Layout Of The System.

Computer intrusion can be classified into three categories: misuse intrusions, network intrusions and host intrusions. Misuse intrusions analyze the information gather and compare it to large databases of attack signatures. Network intrusions, individual packets flowing through a network are analyzed. Passive intrusions, detects a potential security breach, logs the information and signals an alert.

Bankruptcy Fraud: This column focuses on bankruptcy fraud. Bankruptcy fraud means using a credit card while being absent. Bankruptcy fraud is one of the most complicated types of fraud to predict. Some methods or techniques may help in fraud prevention. The bank will send its users/customers an order to pay. However, the users will be recognized as being in a state of personal bankruptcy and not able to recover their unwanted loans. The bank will have to cover the losses itself. One of the possible ways to prevent bankruptcy fraud is by doing a pre-check with credit bureau in order to be informed about the past banking history of its customers. [5] presented a model to forecast personal bankruptcy among users of credit card.

Theft Fraud/ Counterfeit Fraud: In this section, we focus on theft and counterfeit fraud, which are related to one other. Theft fraud refers using a card that is not yours. As soon as the owner give some feedback and contact the bank, the bank will take measures to check the thief as early as possible. Likewise, counterfeit fraud occurs when the credit card is used remotely; where only the credit card details are needed. Firstly, use of your copied card number and codes via various web-sites, where no signature or physical cards are required. [6] although in European E-commerce seems to be quite low, at only 0.83 percent along with the average charge-back ratio, significant concerns are notified in detailed analysis. For the listed credit card, the customers are contacted and if they do not react within certain time limit than the card is blocked.

Application Fraud: When someone applies for a credit card with false information that is termed as application fraud. For detecting application fraud, two different situations have to be classified. When applications come from a same user with the same details, that is called duplicates, and when applications come from different individuals with similar details, that is termed as identity fraudsters. [7]describes application fraud as “demonstration of identity crime, occurs when application form(s) contain possible, and synthetic (identity fraud), or real but also stolen identity information (identity theft).” In most of the banks, eligibility for a credit card, applicants need to complete an application form. Application form is mandatory except for social fields. The bank would also ask for certain details as contact details, such as e-mail address, mobile phone number and land-line number. Confidential information will be the password.

Behavioral Fraud: Behavioral fraud occurs when sales are made on a „cardholder present‟ basis and details of legitimate cards have been obtained fraudulent basis.

Credit Card Fraud Detection

Issues with credit cards, both theoretical and practical, are discussed in this section.

Terms

- **Credit Card:** They can buy things online without actually having the cash on hand through the use a credit card. The use of a credit card streamlines the process of automatically extending credit to consumers. Almost all credit cards now include a unique identifier that speeds up online purchases.
- **Fraud:** Any dishonesty perpetrated with the aim to deceive another person or entity for one's own benefit or harm is considered fraudulent. The concept of fraud is defined differently in different legal systems. Deceit is both a criminal offense and a breach of civil law. One typical goal of fraud is to defraud individuals or organizations of their money.

Credit Card Fraud

Even though there are a lot of credit card transactions in the US, the fraud rate is very low. Ukraine has an alarming 19% fraud rate, second only to Indonesia's 18.3%; other high-risk countries facing the threat of credit card theft include Yugoslavia (17.8%), Malaysia (5.9%), and Turkey (9%). The factors that authorize users to make credit card transactions include the credit card number, signatures, the address of the card holder, the expiration date, and so on. Credit card fraud is the unlawful use of a card or card information without the owner's knowledge, which constitutes a criminal deceit. Credit card fraud detection is an area that receives very little public attention because of its sensitive nature. Methods including ANNs, rule-induction approaches, decision trees, SVMs, LRs, and meta-heuristics like k-means clustering, evolutionary algorithms, and closest neighbor algorithms are commonly used to detect fraud. Humans are capable of committing various forms of fraud, including but not limited to stealing, miscommunication, deceit, dishonesty, and the making of clever but deceptive recommendations. Manually verifying the activities and identities of most external parties can be too costly for companies dealing with millions of them. Without a doubt, there is a direct overhead cost associated with researching each questionable transaction. No matter how suspicious a transaction seems, it is not worth investigating if the sum is less than the overhead cost.[8-10]

ISSN : 2455-9679
SJIF Impact Factor : 6.008

II. Literature Review

Abdul Rehman Khalid et al. (2024) this research explores the use of ensemble methods to improve credit card fraud detection. The authors combine techniques like Bagging, Boosting, Random Forest, SVM, and KNN, along with data pre-processing and SMOTE for handling class imbalance. Their model outperforms traditional machine learning methods in terms of accuracy, precision, recall, and F1-score. This study also demonstrates the practical application of the model using a European credit card dataset and emphasizes the importance of ensemble methods in overcoming challenges such as data imbalance and real-time processing in fraud detection.[11]

Xiaomei Feng et al. (2024) this paper introduces a method that utilizes compact data learning (CDL) to address data imbalance in credit card fraud detection. The authors train their model using a European cardholder dataset and compare their CDL-based feature reduction approach with other feature reduction techniques. The results show that their method outperforms several other machine learning algorithms, contributing to more efficient fraud detection systems in the financial industry by handling data imbalance more effectively.[12]

Diana T. Mosa et al. (2024) this study addresses data imbalance in credit card fraud detection by utilizing meta-heuristic optimization (MHO) approaches. The authors analyze Kaggle's CCF benchmark datasets and compare 15 MHO methods with different transfer functions (TFs) to select important features. They employ SVM and Random Forest classifiers for evaluation and achieve an impressive classification accuracy of 97% by reducing the feature set by 90%. The research emphasizes the significance of feature selection and how machine learning can improve fraud detection systems through innovative methods.[13]

Esraa Faisal Malik et al. (2022) This study investigates seven hybrid machine learning models for credit card fraud detection using a real-world dataset. The authors focus on combining the best-performing algorithms from earlier research to create more effective hybrid models. The Adaboost + LGBM hybrid model demonstrated the best overall performance, highlighting the potential of hybrid models in improving fraud detection systems and calling for further research in algorithm combinations tailored for credit card fraud detection.[14]

Igor Mekterović et al. (2021) This paper discusses the challenges of improving credit card fraud detection in the context of data mining models, particularly focusing on imbalanced datasets and feature engineering. The authors analyze real-world data from card-not-present (CNP) fraud transactions and identify areas where businesses can invest in improving fraud detection systems. The paper emphasizes bridging the gap between academic research and practical concerns in implementing fraud detection models.[15]

Emilija Strelcenia et al. (2023) the authors propose a novel data augmentation model, K-CGAN, to address the issue of imbalanced datasets in credit card fraud detection. They compare various data augmentation methods, including B-SMOTE, K-CGAN, and SMOTE, and find that K-CGAN outperforms the others in terms of F1 score and accuracy. Their results demonstrate the effectiveness of using advanced data augmentation techniques to improve the performance of fraud detection models, particularly in terms of precision and recall.[16]

Ibomoiye Domor Mienye et al. (2024) This research introduces a deep learning framework combining Recurrent Neural Networks (RNNs) with Generative Adversarial Networks (GANs) to improve fraud detection in the face of imbalanced datasets. The GAN component generates synthetic fraudulent transactions to address data imbalance, and the study demonstrates significant improvements in detection accuracy and sensitivity. The authors' GAN-GRU model outperforms traditional methods, showcasing the potential of deep learning techniques in fraud detection.[17]

Mengqiu Li et al. (2024) this paper presents the FEDGAT-DCNN model, which combines dilated convolutions and a Graph Attention Network (GAT) in a federated learning framework for credit card fraud detection. The authors address the issues of sparse data and novel fraud tactics while preserving data privacy. They show that FEDGAT-DCNN outperforms traditional models and other federated learning methods in terms of precision, reliability, and practical applicability for fraud detection in real-world scenarios.[18]

Abdullah Alharbi et al. (2024) In this study, the authors develop a deep learning approach to address the issue of text data in credit card fraud detection. They introduce a text2IMG conversion method that generates small images from text data to handle class imbalance and input them into a convolution neural network (CNN) with class weights. The model's effectiveness and robustness are validated using deep learning and machine learning methods, proving its capacity to improve detection scores in the context of fraud detection.[19]

Table 2Literature Study

Author(s)	**Technique(s)**	**Outcomes**
Victor Chang et al. 2024	Random Under-Sampling, SMOTE	SMOTE: Accuracy 86.75%, F1 Score 73.47%; Under-sampling: Recall 92.86% but lower accuracy; highlights trade-off and statistical fairness.
Abdul Rehman Khalid et al. 2024	Ensemble (Bagging, Boosting, RF, SVM, KNN), SMOTE + Under-sampling	Ensemble outperformed traditional methods in accuracy, precision, recall, F1-score; robust model for credit card fraud detection.
Xiaomei Feng et al. 2024	Compact Data Learning (CDL), Feature Reduction	CDL-enhanced method outperformed existing models; reduced dataset size without compromising performance; beneficial for financial fraud detection.
Diana T. Mosa et al. 2024	Meta-Heuristic Optimization (15 MHO techniques), Feature Selection, SMOTE	Achieved 97% accuracy, reduced feature size by 90%; showed importance of feature selection in fraud detection systems.
Esraa Faisal Malik et al. 2022	Hybrid ML Models (Adaboost + LGBM)	Hybrid Adaboost + LGBM was the best model; emphasized the need for hybrid approaches over standalone ML methods.
Igor Mekterović et al. 2021	Feature Engineering, Cost-Benefit Analysis on Real Data	Identified key areas for fraud detection investment; aligned academic methods with business needs in CNP fraud detection.
Emilija Strelcenia et al. 2023	K-CGAN, B-SMOTE, SMOTE, Data Augmentation	K-CGAN had best F1 Score and Accuracy; B-SMOTE and SMOTE also effective; augmentation methods improved classifier performance on imbalanced data.
Ibomoiye Domor Mienye et al. 2024	GANs + RNN (LSTM, GRU, Simple RNN)	GAN-GRU achieved sensitivity of 0.992 and specificity of 1.000; improved detection in imbalanced datasets.

Table 3 how data imbalance is addressed and the advancements in machine learning techniques:

Method/Technique	**Description**	**Advantages**	**Disadvantages**	**Applications**
Random Oversampling	Duplicating samples from the minority class to balance class distribution.	Simple to implement, can improve model performance on imbalanced datasets.	Can lead to overfitting due to duplicated samples, may not generalize well.	Fraud detection, medical diagnosis (imbalanced classes).
Random Undersampling	Randomly removing samples from the majority class to balance	Reduces the size of the dataset, reducing training time.	Loss of potentially useful data from the majority class, may	Customer churn prediction,

	the dataset.		lead to underfitting.	spam detection.
SMOTE (Synthetic Minority Over-sampling Technique)	Generates synthetic samples for the minority class based on nearest neighbors.	Creates diverse synthetic samples, reduces overfitting compared to random oversampling.	May create unrealistic samples or noise if not tuned properly.	Credit card fraud detection, rare event prediction.
ADASYN (Adaptive Synthetic Sampling)	A variation of SMOTE that focuses on generating synthetic samples for difficult-to-classify instances.	Addresses class imbalance more effectively by focusing on harder examples.	Computationally expensive, can generate noisy or redundant samples.	Intrusion detection, medical diagnosis.
Class Weight Adjustment (in Algorithms like SVC, Logistic Regression)	Assigns different weights to each class during model training to penalize misclassifications of the minority class.	Prevents the model from biasing toward the majority class, easy to implement.	Requires careful tuning of the weights to avoid overemphasizing the minority class.	Imbalanced binary classification tasks.
Ensemble Methods (Bagging, Boosting, Stacking)	Combining multiple weak models to form a strong predictive model, with techniques like **Bagging**, **Boosting**, and **Stacking**.	Boosts performance by leveraging multiple weak models, helps in reducing bias and variance.	Can be computationally expensive and harder to interpret.	Text classification, fraud detection.
Anomaly Detection Techniques	Treats the minority class as anomalies and focuses on detecting them as outliers.	Effective for highly imbalanced datasets where minority class is rare.	May not be suitable for cases where the minority class isn't well-defined as an anomaly.	Fraud detection, network intrusion detection.
Balanced Random Forest	A variant of Random Forest that adjusts for imbalances by modifying how trees are constructed.	Good at handling class imbalance, reduces overfitting by adjusting class distribution in each tree.	Can still be prone to overfitting if not tuned correctly.	Classification tasks with imbalanced classes.
Cost-sensitive Learning	Modifies learning algorithms to take the cost of misclassifying different classes into account.	More control over class misclassification cost, improves focus on minority class.	Requires domain knowledge to assign appropriate costs, may not be suitable for all tasks.	Risk prediction, financial fraud detection.
Transfer Learning	Leverages pre-trained models on a large dataset to help with imbalanced data tasks.	Can reduce the need for large imbalanced datasets, can generalize better from transfer learning.	May require large pre-trained models and significant computational resources.	Image classification, NLP tasks with imbalanced labels.

III. Conclusion

The research papers reviewed collectively emphasize the growing concern of credit card fraud in an era of increasing online transactions and technological advancements. As fraud detection systems face the challenge of imbalanced datasets, with fraudulent transactions being significantly fewer than legitimate ones, various innovative techniques have been explored to enhance detection accuracy and reduce the impact of data imbalance.Most studies reported significant improvements over traditional methods. For example, ensemble models outperformed individual classifiers in various metrics, including precision, recall, accuracy, and F1-score. Similarly, data augmentation strategies like K-CGAN and SMOTE demonstrated superior results in precision and recall, with K-CGAN specifically showing the best performance in terms of F1 score and accuracy.Many of the studies highlighted the importance of effective feature selection and feature reduction techniques. Meta-heuristic optimization methods were frequently used to identify the most critical features, leading to models that not only performed better but also required fewer resources for training.The studies also emphasize the need for continuous innovation and adaptability in fraud detection systems. As fraud tactics evolve, so too must the techniques used to detect them. The ongoing development of hybrid machine learning models and the exploration of novel data augmentation and feature selection methods will likely be key to future advancements in this field.

REFERENCES

[1] Amlan Kundu, Suvasini Panigrahi, Shamik Sural and Arun K. Majumdar, “BLAST-SSAHA Hybridization for Credit Card Fraud Detection,” IEEE Transactions On Dependable And Secure Computing, vol. 6, Issue no. 4, pp.309-315, October-December 2009S.

[2] A. Shen, R. Tong, and Y. Deng, "Application of classification models on credit card fraud detection," June 2007.

[3] Dahl, J.: Card Fraud. In: Credit Union Magazine (2006).

[4] Ghosh and D.L. Reilly, “Credit Card Fraud Detection with a Neural-Network,” Proc. 27th Hawaii International Conference on System Sciences: Information Systems: Decision Support and Knowledge-Based Systems, vol. 3, pp. 621-630, 1994.

[5] Kaiyong Deng, Ru Zhang, Hong Guo,Kaiyong Deng,R Zhang, Dongfang Zhang,WenFeng Jiang,Xinxin Niu “Analysis and Study on Detection of Credit Fraud in Ecommerce 2011

[6] Leila Seyedhossein, Mahmoud Reza Hashemi Mining Information from Credit Card Time Series for Timelier Fraud Detection International Symposium on Telecommunications 2010.

[7] Linda Delamaire (UK), Hussein Abdou (UK), John Pointon (UK), “Credit card fraud and detection techniques: a review”2009.

[8] M.F. Gadi, X. Wang, and A.P. Lago, "Comparison with parametric optimization in credit card fraud detection, 2008.

[9] Md Delwar Hussain Mahdi, Karim Mohammed Rezaul, Muhammad Azizur Rahman “Credit Fraud Detection in the Banking Sector in UK: A Focus on E-Business.” 2010.

[10] Mirjana Pejic-Bach, “Profiling intelligent systems applications in fraud detection and prevention: survey of research articles”, 2010.

[11] Abdul Rehman Khalid,Nsikak Owoh,,Omair Uthmani,Moses Ashawa,Jude Osamor,John Adejoh (2023) “Enhancing Credit Card Fraud Detection: An Ensemble Machine Learning Approach” 2024, 8(1), 6; https://doi.org/10.3390/bdcc8010006, 3 January 2024

[12] Xiaomei Feng,Song-Kyoo Kim (2024) “Novel Machine Learning Based Credit Card Fraud Detection Systems” 2024, 12(12), 1869; https://doi.org/10.3390/math12121869, 15 June 2024

[13] Diana T. Mosa,Shaymaa E. Sorour,Amr A. Abohany,Fahima A. Maghraby (2024) “CCFD: Efficient Credit Card Fraud Detection Using Meta-Heuristic Techniques and Machine Learning Algorithms” 2024, 12(14), 2250; https://doi.org/10.3390/math12142250, 19 July 2024

[14] Esraa Faisal Malik,Khai Wah Khaw,Bahari Belaton,Bahari Belaton,Wai Peng Wong,XinYing Chew (2022) “Credit Card Fraud Detection Using a New Hybrid Machine Learning Architecture” 2022, 10(9), 1480; https://doi.org/10.3390/math10091480, 28 April 2022

[15] Igor Mekterović,Mladen Karan,Damir Pintar,Ljiljana Brkić (2021) “Credit Card Fraud Detection in Card-Not-Present Transactions: Where to Invest” 021, 11(15), 6766; https://doi.org/10.3390/app11156766, 23 July 2021

ISSN : 2455-9679
SJIF Impact Factor : 6.008

[16] Emilija Strelcenia,Simant Prakoonwit (2023) "Improving Classification Performance in Credit Card Fraud Detection by Using New Data Augmentation" 2023, 4(1), 172-198; https://doi.org/10.3390/ai4010008, 31 January 2023

[17] Ibomoiye Domor Mienye,Theo G. Swart (2024) "A Hybrid Deep Learning Approach with Generative Adversarial Network for Credit Card Fraud Detection" 2024, 12(10), 186; https://doi.org/10.3390/technologies12100186, 2 October 2024

[18] Mengqiu Li,John Walsh (2024) "FEDGAT-DCNN: Advanced Credit Card Fraud Detection Using Federated Learning, Graph Attention Networks and Dilated Convolutions" 2024, 13(16), 3169; https://doi.org/10.3390/electronics13163169, 11 August 2024

[19] Abdullah Alharbi,Majid Alshammari,Ofonime Dominic Okon,Amerah Alabrah,Hafiz Tayyab Rauf,Hashem Alyami,Talha Meraj (2022) "A Novel text2IMG Mechanism of Credit Card Fraud Detection: A Deep Learning Approach" 2022, 11(5), 756; https://doi.org/10.3390/electronics11050756, 1 March 2022