International Journal of Industrial Engineering
2023;7(1):28-34.
ISSN: 2456-8449

Credit Card Fraud Detection System to Enhance Legitimate Transactions Using Artificial Neural Networks

S. John Justin Thangaraj[1], M. Manikandan[2*], D. Mansoor Hussian[3], M. Sivakumar[3]

[1]*Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (SIMATS), Chennai, Tamil Nadu, India.*
[2]*Department of Computer Science and Engineering, VIT University, Bhopal, Madhya Pradesh, India.*
[3]*Department of Computer Science and Engineering, Sri Krishna College of Engineering and Technology, Coimbatore, Tamil Nadu, India.*

[*]*Corresponding author: manimmk9792@gmail.com*

Abstract. The explosive expansion of E-Commerce has led to a corresponding surge in credit card fraud as more people use their cards to make transactions online. Credit card fraud is on the rise as credit cards become the preferred method of payment for both online and in-store purchases. In practice, fraudulent and legitimate transactions coexist, and basic pattern-matching approaches are typically insufficient to identify fraud effectively. While Artificial Neural Networks (ANNs) offer several advantages, it's important to note that they require careful tuning, validation, and ongoing monitoring to ensure optimal performance. Additionally, the interpretability of ANNs can be a challenge, and efforts are being made to enhance model interpretability in the context of fraud detection. Identifying fraudulent transactions while avoiding false positives is the main objective of a credit card fraud detection system that uses ANNs effectively and correctly. Therefore, for credit card issues to reduce their loss, it is crucial that they implement reliable fraud detection systems. Various forms of credit card fraud may now be detected using cutting-edge methods based on Artificial Intelligence (AI), Data mining, Fuzzy logic, Machine learning (ML), sequence alignments, Genetics Programming, etc. An effective system for detecting credit card fraud requires a thorough grasp of all these methods. Using a set of predetermined criteria, this study examines the various neural network techniques currently in use for credit card fraud detection and provides an overall rating for each. The result shows 98% accuracy and 96% recall for the credit card fraud detection system.

Keywords: Electronic Commerce, Artificial Intelligence, Artificial Neural Networks, Credit card fraud and legitimate transactions

INTRODUCTION

Financial institutions now rely heavily on innovations in technology to facilitate the electronic delivery of a wide range of services to their clientele. Credit cards, such as Visa and MasterCard, are among the most widely used forms of payment processing in the financial industry. However, bad actors have found new methods to compromise users' cards, usually for fraudulent reasons, and this is a major problem for businesses and their clientele. The purpose of this study is to propose a new technique for detecting credit card fraud using data on transactions made by customers of a well-known company called Vesta. By employing deep learning approaches, the authors were able to come up with a method that successfully identified 99.1% of transactions as fraudulent [1]. Building trustworthy credit card fraud detection systems is complicated by the issue of unbalanced datasets. This article investigates and assesses the advanced credit card fraud detection systems that make use of ML algorithms and deep reinforcement learning (DRLs), using both fraud and non-fraud labels. Two re-sampling methods are utilized to resample the unbalanced credit card fraud data. This well-rounded dataset is subsequently used in conjunction with ML algorithms to create credit card fraud detection frameworks. Next, the unbalanced credit card fraud dataset is used in conjunction with DRL to develop detection methods [2].

For a comprehensive assessment of the efficacy of these ML and DRL models, it is recommended to use the many classification metrics available. Determine the trustworthiness of ML models for credit card fraud identification based on two re-sampling strategies and DRL models via practical tests. The ML models achieve above 99% accuracy when the original credit card fraud datasets are re-sampled prior to the training/test split. However, when the training credit card fraud datasets are re-sampled using these methods, the ML models provide

Received: 18.02.2023 Revised: 23.03.2023 Accepted: 07.04.2023

subpar results, especially in logistic regressions (1.81% accuracy and 3.55% F1 scores). Results show that the DRL model performs poorly and has low success rates, achieving just 34.8% accuracy [3]. The widespread acceptance and convenience of credit cards have accelerated the journey toward a cashless society. However, this comes with the downside of a rise in fraudulent activity, making the adoption of a systematic fraud detection system crucial for both cardholders and the institutions that issue the cards. In this paper, train machine learning models on the supplied datasets using a variety of machine learning algorithms, such as random forests, logistics regressions, Support Vectors Machines (SVM), and Neural Networks, and then conduct a comparative study of the accuracies and other measures of the models produced by each algorithm. Can foretell which algorithm will be more effective for needs by comparing their F_1 scores. Based on the findings, ANNs are the most effective learning algorithms. Their F_1 score was 0.91 [4].

The introduction of the COVID-19 epidemic accelerated the already-rapid growth of online purchasing. Credit card payments are the standard in the online retail industry. As fraudulent transactions cost more money, the issue of detecting fraudulent activity on credit cards is more pressing than ever. Tuning the data mining models is where most of the existing literature on this issue has spent its attention. There is a big discrepancy between the results of academic studies and the caution with which corporations embrace new models (particularly black-box ones) [5]. This article took a wider view and considered this issue from an academic and a business perspective, identifying obstacles in fraud detection problems like feature engineering and unbalanced datasets and separating promising from unprofitable areas to invest in when enhancing fraud detection systems. Research is grounded on actual data from the most common kind of fraudulent purchases, card-not-present (CNP) fraud. The information was supplied by a commercial partner, a global card processing firm. To find the most cost-effective way to enhance their fraud detection system put, many data mining models and methods were tested against the described obstacles [6].

This research offers a technique for identifying credit card fraud using an auto-encoder combined with a probabilistic random forest (AE-PRF). First, the auto-encoder is used in the proposed AE-PRF approach to reduce the large dimensionality of features extracted from credit card transaction data. Then, it uses the random forest, a kind of ensemble learning that combines probabilistic classification with the bootstrap aggregating (bagging) principle, to determine whether a set of data is fake. To measure and compare AE-PRF's efficacy, use the credit card fraud datasets. The credit card fraud dataset is very unbalanced due to the enormous number of legitimate transactions considerably outnumbering the small number of fraudulent ones [7]. The problem statement is discussed below. The challenge and problems related to efficiently detecting fraudulent transactions in credit card data are the focus of the problem statement for an ANN-based credit card fraud detection system. Since fraudsters are always changing their strategies, fraud patterns are complex and hard to spot. Conventional models can find it difficult to depict the dynamic and non-linear character of fraudulent activity. Because ANNs—especially deep learning models—are often seen as "black-box" systems, it may be difficult to understand the logic behind their predictions. Concerns about regulatory compliance and stakeholders may arise from this lack of interpretability. False positives, or genuine transactions that are mistakenly reported as fraudulent may cause aggravation for cardholders and cost companies money. It is important to maintain an equilibrium between false positives and false negatives. To guarantee that an artificial neural network-based credit card fraud detection system is accurate, effective, and complies with industry norms and laws, it is imperative that these issues be resolved.

The creation, installation, and refinement of a credit card fraud detection system using ANNs comprises a variety of labor contributions to provide an efficient and trustworthy fraud detection solution. Putting strong data preprocessing methods into practice to organize and clean credit card transaction data. Improving the neural network's performance requires taking critical measures to handle missing values, deal with outliers, and guarantee data quality. Decide on the right layers, neurons, and activation functions for a neural network design. This entails testing out several designs to see which one best addresses the issue of credit card fraud detection. Put systems in place to handle credit card transactions in real-time. This entails making the neural network and its infrastructure as efficient and quick to process transactions as possible. Smooth integration of the credit card fraud detection technology into the financial institutions' current fraud protection processes. Achieving compatibility and seamless integration is essential for realistic implementation. Together, these contributions aid in the creation, improvement, and implementation of an artificial neural network-based credit card fraud detection system. To develop a system that successfully detects and reduces fraudulent activity in credit card transactions, they address several issues and concerns.

The following section will be a literature survey discussed in section 2. After that, the proposed system is

discussed using a Linear Regression algorithm for credit card fraud detection in section 3. Then, the Result and discussion are discussed to improve the legitimate transaction in section 4. Finally, the conclusion provides the overall performance of the healthcare system and recommendations for future work.

LITERATURE SURVEY

Synthetic minority oversampling methods (SMOTE), adaptive synthetics (ADASYNs), and Tomek links (T-Links) are only a few of the data re-sampling strategies used on the credit card fraud datasets to better balance the quantity of legitimate and fraudulent transactions to boost the performances of the AE-PRF. Whether or not re-sampling strategies are employed to the dataset, experimental findings reveal that AE-PRF performs similarly. Because of this, AE-PRF is well-suited for handling asymmetric data sets. The areas under the receivers' operating characteristics curves, the Matthews correlation coefficients, and the true positive and false negative rates all show that AE-PRF is superior to similar approaches [8]. Credit card fraud is a widespread and rapidly expanding problem. Data Science, along with its close cousin Machine Learning, offers a potential solution to this kind of challenge. The results of four popular machine learning methods—Decision tree, Random Forests, K-nearest neighbor, and Logistic regressions—on much-skewed data sets are compared in this work. To do this, combine relevant aspects of cardholder transactions, including the time, location of the user, kind of goods purchased, cost, vendor, and buying patterns of the client base. To determine whether a transaction is fraudulent, the data is fed into a variety of algorithms that use a combination of accuracy and sensitivity to decide.

Credit card usage has skyrocketed as the globe increasingly digitizes and monetary transactions become paperless. As a result, financial institutions have been suffering from a growing number of fraud-related losses. As a result, there is a need to examine and distinguish fraudulent from legitimate transactions. Here, it provides a thorough analysis of the many techniques now in use to spot credit card fraud. The Hidden Markov Models, Decision Tree, Logistics Regressions, SVM, Genetics algorithm, Neural Network, Random Forest, and Bayesian Belief Networks are all examples of such methods. An in-depth look at the different methods is provided. At the end of each article, summarize the arguments for and against the topic at hand [9]. Credit card processing for online purchases is a convenient and time-saving option for customers. The potential for credit card fraud has grown in tandem with the popularity of using plastic. Both victims and financial institutions might lose a lot of money due to credit card theft. Due to factors such as easily accessible public information, high-class imbalance information, shifting fraud natures, and high false alarm rates, detecting such frauds has been the primary focus of this work. ML-based strategies for credit card identification abound in the relevant literature, including Extreme Learning Techniques, Decision Trees, Random Forests, SVM, Logistics Regressions, and XG Boosts, among others [10].

However, a cutting-edge deep learning algorithm is still required to enhance accuracy and cut down on fraud losses. The most significant effort has been put into using the cutting-edge deep learning algorithms developed recently for this exact goal. To get productive results, a comparison study was conducted between machine learning and deep learning algorithms. The European card benchmark datasets for fraud detection are used in in-depth empirical investigations. An ML algorithm was initially applied to the datasets to increase the reliability of fraud detection. The performance of the fraud detection system is then enhanced by using three convolutional neural network-based designs. The detection accuracy was significantly improved when many layers were added. By experimenting with different configurations of hidden layers, time periods, and advanced models, he has performed a thorough empirical investigation. Accuracy, f1-scores, precisions, and areas under the curves (AUC) were all optimized to 99.9%, 85.71%, 93%, and 98%, respectively, as shown by the assessment of the study work. When applied to credit card recognition challenges, the suggested models achieve better results than the current advanced machine learning and deep learning methods. Also, trials using data-balancing techniques and deep learning algorithms to reduce the number of false negatives is conducted. It is feasible to put the suggested methods into practice for actual credit card fraud detection [11].

People's movement has been curtailed to some level because of the COVID-19 epidemic, making it more difficult to buy products and services offline and giving rise to a culture of growing reliance on internet services. Credit card fraud is a major problem in any industry that deals with money transfers via the Internet. Therefore, there is a pressing need to perfect the use of machine learning to avoid almost all instances of fraudulent credit card transactions. This study looks at 66 different machine learning models over two distinct phases of testing. Every model incorporates stratified K-fold cross-validations and real-world credit card fraud detection datasets of European cardholders [12]. First, nine different machine learning algorithms for spotting fraudulent financial

transactions are put to the test. The top three algorithm techniques are chosen for usage in the next phase when they are put through a battery of 19 different resampling procedures. The All K-Nearest Neighbor (AllKNN) under-sampling approach combined with CatBoosts (AllKNN-CatBoosts) is regarded as the best-recommended model out of 330 evaluation metrics values that took about one month to achieve. To that end, the AllKNN-CatBoosts model is evaluated by contemporaries. The findings show that the suggested model is superior to the baseline models in terms of its AUC (97.94%), Recalls (95.91%), and F1-Scores (87.40%) [13].

As a result of the worldwide spreading of the COVID-19 epidemic and the rapid development of online payment systems in recent years, the average daily volume of online transactions and credit card payments has increased dramatically. Unsurprisingly, credit card theft has also increased, wreaking havoc on financial institutions, credit card issuers, and, ultimately, retailers and suppliers. As a result, there is a pressing need to set up reliable measures that protect the honesty of online credit card transactions. To solve the problems of credit card fraud detection, the authors of this work offer a metaheuristic strategy that combines ML with swarm intelligence. To fine-tune SVM, extreme learning machines, and extreme gradients-boosting ML models, a fresh, improved firefly method, called the group search firefly algorithm, was developed. The real-world datasets for credit card fraud detection were collected from the transactions of credit card users in Europe and boosted models were evaluated on them [14]. Since the original dataset is so unbalanced, the SMOTE was used to increase the sample sizes to better examine how well-tailored ML models perform. Recent state-of-the-art techniques were used to evaluate the suggested groups' search firefly metaheuristic. The study has made use of common metrics for measuring machine learning performance, such as classifier accuracy, recall, precision, and area under the curve. Evidence from experiments shows that models tweaked using the proposed approach outperformed those hybridized with competing meta-heuristics [15].

PROPOSED SYSTEM

The use of stolen card numbers is a major issue for credit card companies. Credit card theft of all varieties is expected to cost Americans over $850 million this year, a 10% rise over 1991 levels. Although negligible in comparison to the $8.5 billion in 1992 credit card losses attributable to charge-offs of a substantially delinquent account. As a share of all charges, fraud is on the rise, expanding at a rate that outpaces the expansion of the credit card industry. The fraud issue ballooned from 8 basis points in 1988 to over 20 by 1991. There are a few main types of credit card fraud, while there are numerous subtypes. A certain "base level" of fraud activity may be attributed to lost and stolen card information. This foundational level may grow or shrink depending on macroeconomic factors (high unemployment periods, for example, are associated with higher fraud losses owing to lost and stolen cards). Despite more advanced card manufacturing technology (hologram on the card) and encryptions of information on the magnetic stripes, fraud due to counterfeit cards has been a rising concern in recent years. Figure 1 shows the system architecture of the proposed system.

Counterfeiting is, without a doubt, a more widespread and systemic issue in certain regions, in contrast to the "opportunistic" and, hence, random character of most card theft resulting from lost or stolen cards. Thefts involving cards taken from the mail have increased dramatically in recent years. This so-called NRI (non-receipts of issues) fraud occurs when issuing new cards and when replacing lost or stolen cards. There are higher rates of NRI in certain parts of the nation than in others. Because of the severity of the situation in certain regions, card issuers have resorted to using couriers instead of regular mail to send cards and have created unique activation schemes for their customers. If a consumer does not contact their bank within three business days to verify card receipt, the card will be deactivated (marked in the bank's authorization system as an account for which transaction requests would be rejected). By asking a few questions about the card member's applications (if it's a new problem) or cardholder information files, the bank may verify that the person calling in is a legal cardholder. Even though these solutions are somewhat pricey, they have successfully reduced NRI losses. False applications for cards are submitted as part of other fraud operations.

Criminals in these circumstances get access to victims' real identity and financial data, then use that data to apply for a card in victims' names, indicating a postal drop at which to send the card. If a card is granted under these conditions, not even card member activation can keep it out of the wrong hands since the fraudster who created the fake application information may present it during the activation call. Mail and phone order scams can contribute to economic losses due to fraud. No card imprint can be produced as proof of purchase since the buyer was not present before the merchants at the time of the transactions. Address verification by phone with the cardholders at the time of the transaction is one method used to prevent fraud. All the above are typical manifestations of account

or cardholder fraud. There is also shoplifting that begins with the retailer itself. The "laundering" of fraudulent merchant receipts may result in the acquisition of significant quantities of money for nonexistent transactions. During valid transactions, account information may sometimes be duplicated at the merchant's location, thanks to merchant collaboration. This data is then utilized to create phony copies of actual cards, which are then used in other fraudulent situations.

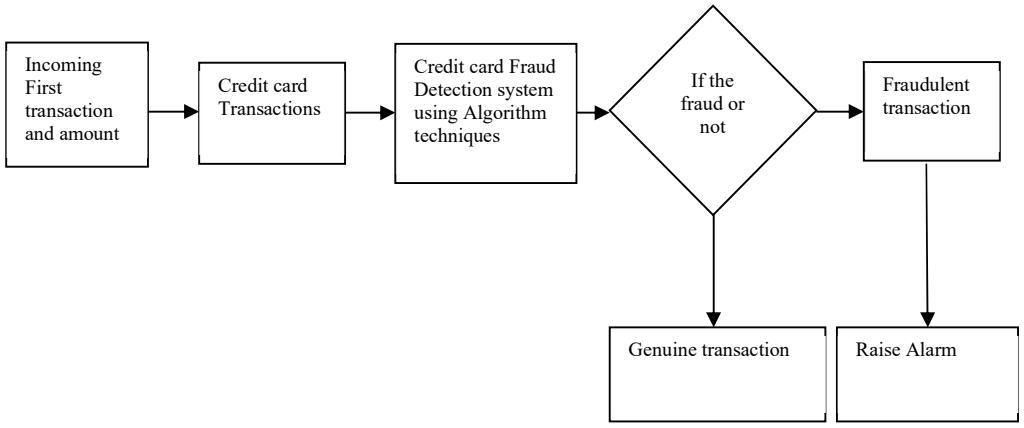

FIGURE 1. System architecture of the proposed system

All these counterfeit cards have transaction histories that can be traced back to usage at the provided merchant's institution, making the colluding merchant a "point of compromise" in this operation. In addition, there is a broad category of dishonesty that is often referred to as abuse. The cardholders use the card to make purchases for which payment is not planned. Sometimes, this is done deliberately just before the cardholder files for bankruptcy. Rather than being counted as credit card fraud, losses attributable to this kind of "bankruptcy fraud" are included in charge-off totals. Bankruptcy fraud was estimated to be at $2.65 billion in 1992, making it the largest single kind of fraud loss at that time.

RESULTS AND DISCUSSIONS

Identifying fraudulent conduct is not a simple process because of the wide variety of fraudulent activities. Most financial institutions use anti-fraud checks as part of their standard procedure when reviewing credit card applications. (Investigators have uncovered false applications filed by members of organized crime by examining application forms for telltale ways of handwriting. This method has been used to uncover some of the fake applications submitted by Nigerian fraud rings. However, after a card has been authorized, most financial institutions depend on frequent reviews of account activity to decide whether fraud is suspected. Every portfolio activity is assessed against a set of rules that banks have devised. These types of verifications might include minimal limitations on the daily volume of transactions. This report on excessive transactions might be restricted to just those that exceed a certain monetary threshold. The fraudulent conduct in the portfolio's history is analyzed to generate these guidelines. However, most financial institutions merely do the bare minimum of statistical analysis when formulating fraud rules; hence, rule sets often consist of a few elementary threshold conditions on account data.

As can be expected, more advanced methods of fraud detection can significantly enhance performance. In particular, the subject of fraud detection is a great application for a well-selected neural network solution when considered as a challenge in pattern recognition. Many issues in the banking and insurance sectors are being reframed as pattern recognition challenges that might benefit from neural network approaches. Mellon Bank commissioned a feasibility study to ascertain whether a neural network could effectively identify fraud in the bank's credit card portfolio. Figure 2 and Figure 3 show the accuracy and recall performance, respectively.

$$Accuracy = \frac{Total\ correct\ predictions}{Total\ number\ of\ predictions} \quad (1)$$

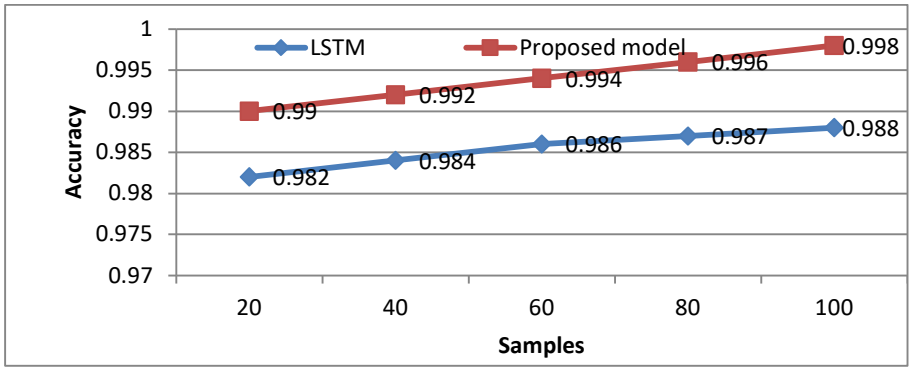

FIGURE 2. Accuracy performance of the proposed system

$$Recall = \frac{True\ positive}{True\ positive + False\ Negative} \quad (2)$$

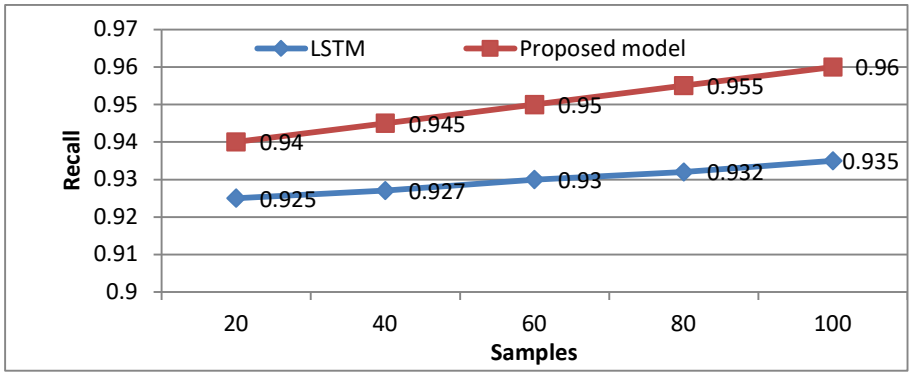

FIGURE 3. Recall the performance of the proposed system.

To conduct the feasibility study, a neural networks-based system was trained on a subset of good and fraud accounts and then subjected to a blind test on an un-sampled subset of transactions. Initially, no fraud labels were supplied in the blind test set, although good/fraud labels were included with the transactions in the training sets. The purpose of the research was to model the performance of a post-authorization processing fraud detection system for a bank. Authorization requests are accompanied by a stream of authorization data elements that detail the nature of the transaction (such as the amount and the merchant code) and the cardholder (identified by account number). The credit card dataset is shown in Table 1.

TABLE 2: Credit Card Clients Dataset

Dataset	Total instances	Defaulter clients	Healthy clients
Imbalanced datasets	30000	6636	23364
Undersampled datasets	13272	6636	6636
Oversampled datasets	46728	23364	23364

An authorization system feed may provide access to extra data fields, such as the time of day. Financial institutions do not keep records of their permission files. The bank's credit card processing systems only store transactions that are sent there by the merchants for settlements. Thus, information about financial dealings was compiled by pulling relevant data from Mellon's settlement files. In this snippet, we can only use the permission data that was saved in the final settlement file to build a prototype. Particularly, the day but not the time of the transaction, the amount of the transaction, and the merchant's codes were all accessible. No information about refused transactions (authorization requests) was accessible either.

CONCLUSIONS

ANNs are used in credit card fraud detection systems. The success of the model in detecting fraudulent transactions is usually measured by evaluating its output using a variety of performance criteria. The system's overall accuracy is for detecting credit card fraud. The proportion of accurately categorized transactions (both regular and fraudulent) is indicated by high accuracy. However, accuracy may not be the only important metric particularly in unbalanced datasets. The proportion of fraudulent transactions that were fraudulent as opposed to those that were accurately forecasted. A low false negative rate, which is indicative of a high recall, means that most genuine fraudulent transactions are successfully captured by the system. It's crucial to remember that the outcomes might change depending on the neural network architecture's complexity, the quality of the data preprocessing, the training and testing datasets, and other variables. The assessment metrics provide a thorough picture of the credit card fraud detection system's operation and direct modifications to raise its efficacy in practical situations. Furthermore, since fraud trends change over time, ongoing monitoring and modifications are essential to sustaining high performance. Provide sophisticated continuous monitoring systems that provide real-time anomaly and fraud detection capabilities. This covers the use of proactive alerting and streaming analytics.

REFERENCES

[1]. K.I. Alkhatib, AI. Al-Aiad, M.H. Almahmoud, and O.N. Elayan, 2021, "Credit card fraud detection based on deep neural network approach," *12th International Conference on Information and Communication Systems*, pp. 153-156.

[2]. T.K. Dang, T.C. Tran, L.M. Tuan, M.V. Tiep, 2021, "Machine learning based on resampling approaches and deep reinforcement learning for credit card fraud detection systems," *Applied Sciences*, **11(21)**, pp. 1-5.

[3]. P. Sharma, S. Banerjee, D. Tiwari, and J.C. Patni, 2021, "Machine learning model for credit card fraud detection-a comparative analysis," *Int. Arab J. Inf. Technol*, **18(6)**, pp. 789-796.

[4]. I. Mekterović, M. Karan, D. Pintar, and L. Brkić, 2021, "Credit card fraud detection in card-not-present transactions: Where to invest?" *Applied Sciences*, **11(15)**, pp. 1-6.

[5]. T.H. Lin, and J.R. Jiang, 2021, "Credit card fraud detection with autoencoder and probabilistic random forest," *Mathematics*, **9(21)**, pp. 1-7.

[6]. AS. Rathore, A. Kumar, D. Tomar, V. Goyal, K. Sarda, D. Vij, 2021, "Credit card fraud detection using machine learning," *In 10th International Conference on System Modeling & Advancement in Research Trends (SMART)*, pp. 167-171.

[7]. P. Tiwari, S. Mehta, N. Sakhuja, J. Kumar, and AK. Singh, 2021, "Credit card fraud detection using machine learning: a study," *arXiv preprint arXiv:2108.10005*, pp. 1-8.

[8]. F.K. Alarfaj, I. Malik, H.U. Khan, N. Almusallam, M. Ramzan, and M. Ahmed, 2022, "Credit card fraud detection using state-of-the-art machine learning and deep learning algorithms," *IEEE Access*, **10**, pp. 39700-39715.

[9]. N.S. Alfaiz, and S.M. Fati, 2022, "Enhanced credit card fraud detection model using machine learning," *Electronics*, **11(4)**, pp. 1-9.

[10]. D. Jovanovic, M. Antonijevic, M. Stankovic, M. Zivkovic, M. Tanaskovic, and N. Bacanin, 2022, "Tuning machine learning models using a group search firefly algorithm for credit card fraud detection," Mathematics, **10(13)**, pp. 1-6.

[11]. R. Bin Sulaiman, V. Schetinin, and P. Sant, 2022, "Review of machine learning approach on credit card fraud detection," *Human-Centric Intelligent Systems*, **2**, pp. 55-68.

[12]. O. Voican, 2021, "Credit Card Fraud Detection using Deep Learning Techniques," *Informatica Economica*, **25(1)**, pp. 1-5.

[13]. VS. Karthik, A. Mishra, US. Reddy, 2022, "Credit card fraud detection by modeling behavior pattern using hybrid ensemble model," *Arabian Journal for Science and Engineering*, pp. 1-1.

[14]. I. Benchaji, S. Douzi, B. El Ouahidi, and J. Jaafari, 2021, "Enhanced credit card fraud detection based on attention mechanism and LSTM deep model," *Journal of Big Data*, pp. 1-21.

[15]. M.J. Madhurya, H.L. Gururaj, B.C. Soundarya, K.P. Vidyashree, A.B. Rajendra, 2022, "Exploratory analysis of credit card fraud detection using machine learning techniques," *Global Transitions Proceedings*, **3(1)**, pp. 31-37.