International Journal of Engineering Technology and Management Science
Website: ijetms.in Issue: 2 Volume No.9 March - April – 2025
DOI:10.46647/ijetms.2025.v09i02.107 ISSN: 2581-4621

Credit Card Fraud Detection Using State-of-the-Art Machine Learning and Deep Learning Algorithms

Shaik Zaheer Basha[1], Bushra Tahseen[2]
[1]MCA Student, Dr.K.V.Subba Reddy Institute of Technology, Kurnool, Andhra Pradesh, India
[2]Associate Professor, Dr.K.V.Subba Reddy Institute of Technology, Kurnool, Andhra Pradesh, India

ABSTRACT
People can use credit cards for online transactions as it provides an efficient and easy-to-use facility.With the increase in usage of credit cards, the capacity of credit card misuse has also enhanced. Credit card frauds cause significant financial losses for both credit card holders and financial companies. In this research study, the main aim is to detect such frauds, including the accessibility of public data, high-class imbalance data, the changes in fraud nature, and high rates of false alarm. The relevant literature presents many machines learning based approaches for credit card detection, such as Extreme Learning Method, Decision Tree, Random Forest, Support Vector Machine, Logistic Regression and XG Boost. However, due to low accuracy, there is still a need to apply state of the art deep learning algorithms to reduce fraud losses. The main focus has been to apply the recent development of deep learning algorithms for this purpose. Comparative analysis of both machine learning and deep learning algorithms was performed to find efficient outcomes. The detailed empirical analysis is carried out using the European card benchmark dataset for fraud detection. A machine learning algorithm was first applied to the dataset, which improved the accuracy of detection of the frauds to some extent. Later, three architectures based on a convolutional neural network are applied to improve fraud detection performance. Further addition of layers further increased theaccuracy of detection. A comprehensive empirical analysis has been carried out by applying variations in the number of hidden layers, epochs and applying the latest models. The evaluation of research work shows the improved results achieved, such as accuracy, f1-score, precision and AUC Curves having optimized values of 99.9%,85.71%,93%, and 98%, respectively. The proposed model outperforms the state-of-the-art machine learning and deep learning algorithms for credit card detection problems. In addition, we have performed experiments by balancing the data and applying deep learning algorithms to minimize the false negative rate. The proposed approaches can be implemented effectively for the real-world detection of credit card fraud.
Keywords: Credit, fraud, accuracy

Introduction

Credit card fraud (CCF) is a type of identity theft in which someone other than the owner makes an unlawful transaction using a credit card or account details. A credit card that has been stolen, lost, or counterfeited might result in .fraud. Card-not-present fraud, or the use of your credit card number in e-commerce transactions has also become increasingly common as a result of the increase in online shopping. Increased fraud, such as CCF, has resulted from the expansion of e-banking and several online payment environments, resulting in annual losses of billions of dollars. In this era of digital payments, CCF detection has become one of the most important goals. As a business owner, it cannot be disputed that the future is heading towards a cashless culture. As a result, typical payment methods will no longer be used in the future, and therefore they will not be helpful for expanding a business. Customers will not always visit the business with cash in their pockets. They are now placing a premium on debit and credit card payments. As a result, companies will need to

DOI:10.46647/ijetms.2025.v09i02.107 ISSN: 2581-4621

update their environment to ensure that they can take all types of payments. In the next years, this situation is expected to become much more severe [1].

In 2020, there were 393,207 cases of CCF out of approximately 1.4 million total reports of identity theft [4]. CCF is now the second most prevalent sort of identity theft recorded as of this year, only following government documents and benefits fraud [5]. In 2020, there were 365,597 incidences of fraud perpetrated using new credit card accounts [10]. The number of identity theft complaints has climbed by 113% from 2019 to 2020, with credit card identity theft reports increasing by 44.6% [14]. Payment card theft cost the global economy $24.26 billion last year. With 38.6% of reported card fraud losses in 2018, the United States is the most vulnerable country to credit theft.

As a result, financial institutions should prioritize equipping themselves with an automated fraud detection system. The goal of supervised CCF detection is to create a machine learning (ML) model based on existing transactional credit card payment data. The model should distinguish between fraudulent and non fraudulent transactions, and use this information to decide whether an incoming transaction is fraudulent or not. The issue involves a variety of fundamental problems, including the system's quick reaction time, cost sensitivity, and feature pre-processing. ML is a field of artificial intelligence that uses a computer to make predictions based on prior data trends [1]

ML models have been used in many studies to solve numerous challenges. Deep learning (DL) algorithms applied applications in computer network, intrusion detection, banking, insurance, mobile cellular networks, health care fraud detection, medical and malware detection, detection for video surveillance, location tracking, Android malware detection, home automation, and heart disease prediction. We explore the practical application of ML, particularly DL algorithms, to identify credit card thefts in the banking industry in this paper. For data categorisation challenges, the support vector machine (SVM) is a supervised ML technique. It is employed in a variety of domains, including image recognition [25], credit rating [5], and public safety [16]. SVM can tackle linear and nonlinear binary classification problems, and it finds a hyper plane that separates the input data in the support vector, which is superior to other classifiers. Neural networks were the first method used to identify credit card theft in the past [4]. As a result, (DL), a branch of ML, is currently focused on DL approaches.

In recent years, deep learning approaches have received significant attention due to substantial and promising outcomes in various applications, such as computer vision, natural language processing, and voice. However, only a few studies have examined the application of deep neural networks in identifying CCF. [3]. It uses a number of deep learning algorithms for detecting CCF. However, in this study, we choose the CNN model and its layers to determine if the original fraud is the normal transaction of qualified datasets. Some transactions are common in datasets that have been labelled fraudulent and demonstrate questionable transaction behavior . As a result, we focus on supervised and unsupervised learning in this research paper.

The class imbalance is the problem in ML where the total number of a class of data (positive) is far less than the total number of another class of data (negative). The classification challenge of the unbalanced dataset has been the subject of several studies. An extensive collection of studies can provide several answers. Therefore, to the best of our knowledge, the problem of class imbalance has not yet been solved. We propose to alter the DL algorithm of the CNN model by adding the additional layers for features extraction and the classification of credit card transactions as fraudulent or otherwise. The top attributes from the prepared dataset are ranked using feature selection techniques. After that, CCF is classified using several supervised machine-driven and deep learning models.

In this study, the main aim is to detect fraudulent transactions
using credit cards with the help of ML algorithms and deep learning algorithms. This study makes the following contributions:

DOI:10.46647/ijetms.2025.v09i02.107 ISSN: 2581-4621

_ Feature selection algorithms are used to rank the top features from the CCF transaction dataset, which help in class label predictions.
_ The deep learning model is proposed by adding a number of additional layers that are then used to extract the features and classification from the credit card farad detection dataset.
_ To analyse the performance CNN model, apply different architecture of CNN layers.
_ To perform a comparative analysis between ML with DL algorithms and proposed CNN with baseline model, the results prove that the proposed approach outperforms existing approaches.
_ To assess the accuracy of the classifiers, performance evaluation measures, accuracy, precision, and recall are used. Experiments are performed on the latest credit cards dataset.

The rest of the paper is structured as follows: The second section examines the related works. The proposed model and its methodology are described in depth in Section 3. The dataset and evaluation measures are described in Section 4. It also shows the outcomes of our tests on a real dataset, as well as the analysis.

Literature Survey

An efficient real time model for credit card fraud detection based on deep learning:

In the last few decades, machine learning has gotten better at handling and organizing data, which has made it possible to build smart, dynamic, real-time systems. How accurate and precise these kinds of systems are depends on how well the data is time- and logical-corrected and how long it takes for feedback to be generated. This study looks at a method for finding scams. To get better accuracy and precision, banks and other financial companies are spending more on data analysis tools and programs that find scams. To resolve this issue, various machine learning based systems and strategies have been proposed in the literature. Few exploration that look at different deep Learning models don't think about the meaning of an ongoing strategy in this unique circumstance. We offer a real-time credit card theft monitoring system built on deep neural networks to help you deal with this problem. We use an auto-encoder in our model to instantly decide whether a credit card transaction is real or fake. Four binary classification models are used to test our method. The Benchmark shows that our model is more accurate, recalls more information, and is more precise than other options.Using artificial intelligence to make it easier for users to authorize transactions from credit card data logs that aren't balanced: Businesses and credit agencies could find it more straightforward to assess monetary gamble when they utilize artificial intelligence and machine learning. The reason for this exploration is to help credit card firms in evaluating and breaking down the gamble of credit card failure utilizing a prediction approach. Machine learning helps with risk assessment by spotting dishonesty in huge amounts of uneven data and deciding whether a transaction is real or fake. If a bank is told about a fake transaction, they can stop giving money to that person. Modified RUSBoost outperformed DT, LR multilayer perceptrons, KNN, RF and SVM. The assessment incorporated the utilization of F scores, precision memory curves, sensitivity, specificity, accuracy, and region under the beneficiary working trademark bend.

Performance analysis of feature selection methods in software defect prediction:

A search method approac: SDP models employ software system measurements. A SDP model's quality depends on its software measurements (datasets). Issues with data quality, including excessive complexity, might make the SDP model less successful. The method for choosing features for dimensions has been used a lot of times. Most research studies on FS approaches for SDP have mixed outcomes, making selection challenging. Different FS methods react differently to different computer factors. Since FS's effectiveness depends on the search strategy, this could have something to do with its search methods. Different search methods must be used to compare how well different FS techniques work in SDP. This study used five NASA software problem datasets and four classifiers to try fourteen FSS and four FFR methods. The preliminaries' discoveries showed that FS approaches upgrade the classifier's prediction accuracy yet additionally depend on the dataset. In FFR methods, Information Gain made forecast models work better than anything else. Keeping up The most important thing that changes forecast models in FSS is Feature Subset

DOI:10.46647/ijetms.2025.v09i02.107 ISSN: 2581-4621

Selection based on Best First Search. It was proven that FFR forecast models were more stable than FSS models. So, FS methods make SDP models work better, but how well they work depends on the information and forecast model that are used. We suggest that you use FFR methods because forecast models that are based on FFR are more accurate.

Fraud and corruption control at education system level:

A case study of the Victorian department of education and early childhood development in Australia: In Victoria, Australia, the Department of Education and Early Childhood Development (the Department) tried to make a scheme to stop fraud and cheating. A small team from the Department, which included the author of this study, oversaw and carried out the policy work. The policy framework spreads out and decentralizes roles and duties for running the government. This case illustrates the policy's complexity, the limitations that made it hard to apply, and the Department's workable solution. This could be useful for people who work in big, diverse school systems, even if it's not clear how to stop fraud and abuse. **Auto loan fraud detection using dominancebased rough set approach versus machine learning methods:**

As financial services and activities grow, so does financial crime. Even though there are attempts to stop financial crime, thieves are always coming up with new ways to get around scam detection systems. This makes it hard to use quantitative methods and predictive models. So, new ways need to be researched and tried to use the study's results to make it easier to spot fraud and create fraud security systems with extra checks to cut down on suspicious behavior. Unlike credit card misuse, auto loans are important financial tools that have not been fully looked into. The Dominance-based Rough Set Balanced Rule Ensemble (DRSA-BRE) is used to look at a set of data about new car loan applications in order to spot financial scams. With more people applying for car loans fraudulently, the suggested way has several advantages over the current ones.

Existing System

ML has many branches, and each branch can deal with different learning tasks. However, ML learning has different framework types. The ML approach provides a solution for CCF, such as random forest (RF). The ensemble of the decision tree is the random forest [3]. Most researchers use the RF approach. To combine the model, we can use (RF) along with network analysis. This method is called APATE [1]. Researchers can use different ML techniques, such as supervised learning and unsupervised techniques. ML algorithms, such as LR, ANN, DT, SVM and NB, are commonly used for CCF detection.

The researcher can combine these techniques with ensemble techniques to construct solid detection classifiers [3]. The linking of multiple neurons and nodes is known as an artificial neural network. A feed-forward perceptron multilayer is built up of numerous layers: an input layer, an output layer and one or more hidden layers. For the representation of the exploratory variables, the first layer contains the input nodes. With a precise weight, these input layers are multiplied, and each of the hidden layer nodes is transferred with a certain bias, and they are added together.

An activation function is then applied to create the output of each neuron for this summation, which is then transferred to the next layer. Finally, the algorithm's reply is provided by the output layer. The first set randomly used weights and formerly used the training set to minimise the error. All these weights were adjusted by detailed algorithms such as backpropagation [2], [6]. The graphic model for contingency relationships between a set of variables is called the Bayesian belief network. The independence assumption in naïve Bayes is that it was developed to relax and allow for dependencies among variables.

Disadvantages

- ❖ The system is not implemented Classification on Imbalanced Data.
- ❖ The system is not implemented CONVOLUTIONAL NEURAL NETWORK (CNN) for test and train the datasets.

International Journal of Engineering Technology and Management Science
Website: ijetms.in Issue: 2 Volume No.9 March - April – 2025
DOI:10.46647/ijetms.2025.v09i02.107 ISSN: 2581-4621

Proposed System

• _ Feature selection algorithms are used to rank the top features from the CCF transaction dataset, which help in class label predictions.

_ The deep learning model is proposed by adding a number of additional layers that are then used to extract the features and classification from the credit card farad detection dataset.

_ To analyse the performance CNN model, apply different architecture of CNN layers.

_ To perform a comparative analysis between ML with DL algorithms and proposed CNN with baseline model, the results prove that the proposed approach outperforms existing approaches.

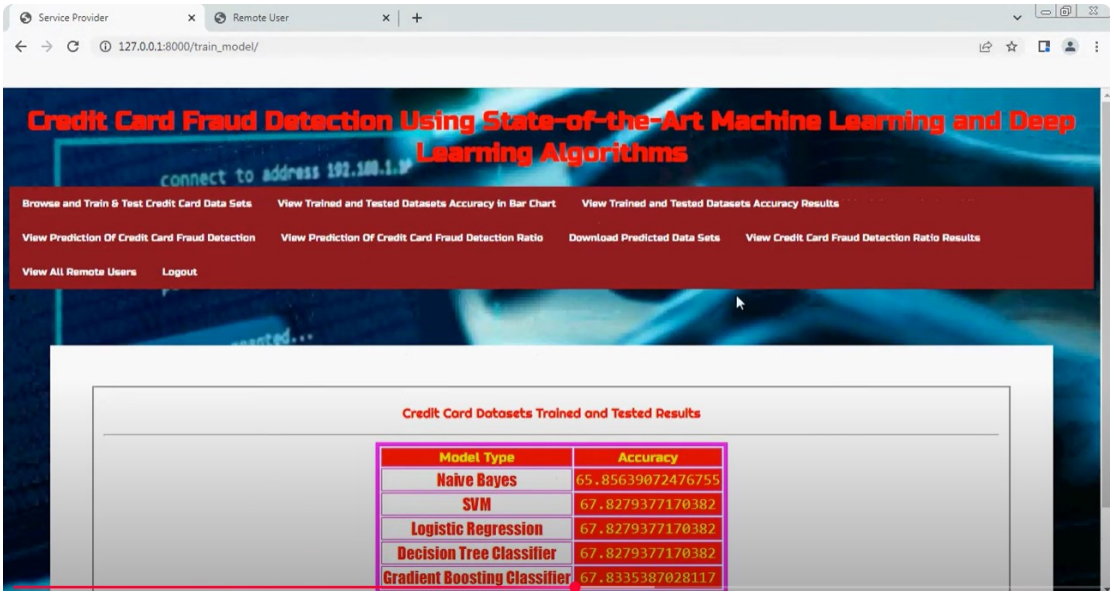
_ To assess the accuracy of the classifiers, performance evaluation measures, accuracy, precision, and recall are used. Experiments are performed on the latest credit cards dataset.

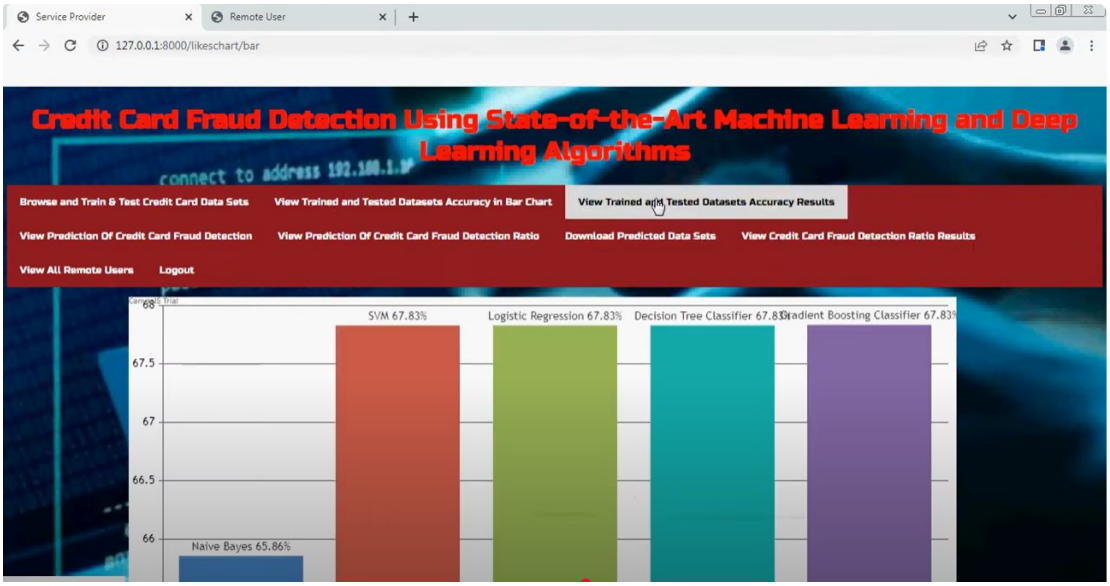
Advantages

➢The proposed system uses SUPERVISED MACHINE LEARNING APPROACHES which are effective for testing and training datasets.

➢The proposed system implemented CNN is to minimise processing without losing key features by reducing the image to make predictions

Results





DOI:10.46647/ijetms.2025.v09i02.107 ISSN: 2581-4621

Conclusion

CCF is an increasing threat to financial institutions. Fraudsters tend to constantly come up with new fraud methods. A robust classifier can handle the changing nature of fraud. Accurately predicting fraud cases and reducing false-positive cases is the foremost priority of a fraud detection system. The performance of ML methods varies for each individual business case. The type of input data is a dominant factor that drives different ML methods. For detecting CCF, the number of features, number of transactions, and correlation between the features are essential factors in determining the model's performance. DL methods, such as CNNs and their layers, are associated with the processing of text and the baseline model. Using these methods for the detection of credit cards yields better performance than traditional algorithms. Comparing all the algorithm performances side to side, the CNN with 20 layers and the baseline model is the top method with an accuracy of 99.72%. Numerous sampling techniques are used to increase the performance of existing examples, but they significantly decrease on the unseen data. The performance on unseen data increased as the class imbalance increased. Future work associated may explore the use of more state of art deep learning methods to improve the performance of the model proposed in this study.

References

[1] Y. Abakarim, M. Lahby, and A. Attioui, ``An ef_cient real time model for credit card fraud detection based on deep learning," in *Proc. 12th Int. Conf. Intell. Systems: Theories Appl.*, Oct. 2018, pp. 1_7, doi:10.1145/3289402.3289530.
[2] H. Abdi and L. J. Williams, ``Principal component analysis," *Wiley Inter-discipl. Rev., Comput. Statist.*, vol. 2, no. 4, pp. 433_459, Jul. 2010, doi:10.1002/wics.101.
[3] V. Arora, R. S. Leekha, K. Lee, and A. Kataria, ``Facilitating userauthorization from imbalanced data logs of credit cards using arti_cialintelligence," *Mobile Inf. Syst.*, vol. 2020, pp. 1_13, Oct. 2020, doi:10.1155/2020/8885269.
[4] A. O. Balogun, S. Basri, S. J. Abdulkadir, and A. S. Hashim, ``Performanceanalysis of feature selection methods in software defect prediction: Asearch method approach," *Appl. Sci.*, vol. 9, no. 13, p. 2764, Jul. 2019, doi: 10.3390/app9132764.
[5] B. Bandaranayake, ``Fraud and corruption control at education system level: A case study of the Victorian department of education and early childhood development in Australia," *J. Cases Educ. Leadership*, vol. 17, no. 4, pp. 34_53, Dec. 2014, doi: 10.1177/1555458914549669.
[6] J. Bªaszczy«ski, A. T. de Almeida Filho, A. Matuszyk, M. Szelg_, and R. Sªowi«ski, ``Auto loan fraud detection using dominance-based rough set approach versus machine learning methods," *Expert Syst. Appl.*, vol. 163, Jan. 2021, Art. no. 113740, doi: 10.1016/j.eswa.2020.113740.
[7] B. Branco, P. Abreu, A. S. Gomes, M. S. C. Almeida, J. T. Ascensão, and P. Bizarro, ``Interleaved sequence RNNs for fraud detection," in *Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2020, pp. 3101_3109, doi: 10.1145/3394486.3403361.
[8] F. Cartella, O. Anunciacao, Y. Funabiki, D. Yamaguchi, T. Akishita, and O. Elshocht, ``Adversarial attacks for tabular data: Application to fraud detection and imbalanced data," 2021, *arXiv:2101.08030*.
[9] S. S. Lad, I. Dept. of CSERajarambapu Institute of TechnologyRajaramnagarSangliMaharashtra, and A. C. Adamuthe, ``Malware classi_cation with improved convolutional neural network model," *Int.J. Comput. Netw. Inf. Secur.*, vol. 12, no. 6, pp. 30_43, Dec. 2021, doi: 10.5815/ijcnis.2020.06.03.
[10] V. N. Dornadula and S. Geetha, ``Credit card fraud detection using machine learning algorithms," *Proc. Comput. Sci.*, vol. 165, pp. 631_641, Jan. 2019, doi: 10.1016/j.procs.2020.01.057.
[11] I. Benchaji, S. Douzi, and B. E. Ouahidi, ``Credit card fraud detection model based on LSTM recurrent neural networks," *J. Adv. Inf. Technol.*, vol. 12, no. 2, pp. 113_118, 2021, doi: 10.12720/jait.12.2.113-118.
[12] Y. Fang, Y. Zhang, and C. Huang, ``Credit card fraud detection based on machine learning," *Comput., Mater. Continua*, vol. 61, no. 1, pp. 185_195, 2019, doi: 10.32604/cmc.2019.06144.

DOI:10.46647/ijetms.2025.v09i02.107 ISSN: 2581-4621

[13] J. Forough and S. Momtazi, ``Ensemble of deep sequential models for credit card fraud detection," *Appl. Soft Comput.*, vol. 99, Feb. 2021, Art. no. 106883, doi: 10.1016/j.asoc.2020.106883.
[14] K. He, X. Zhang, S. Ren, and J. Sun, ``Deep residual learning for image recognition," 2015, *arXiv:1512.03385*.
[15] X. Hu, H. Chen, and R. Zhang, ``Short paper: Credit card fraud detection using LightGBM with asymmetric error control," in *Proc. 2nd Int. Conf. Artif. Intell. for Industries (AII)*, Sep. 2019, pp. 91_94, doi: 10.1109/AI4I46381.2019.00030.
[16] J. Kim, H.-J. Kim, and H. Kim, ``Fraud detection for job placement using hierarchical clusters-based deep neural networks," *Int. J. Speech Technol.*, vol. 49, no. 8, pp. 2842_2861, Aug. 2019, doi: 10.1007/s10489-019-01419-2.
[17] M.-J. Kim and T.-S. Kim, ``A neural classi_er with fraud density map for effective credit card fraud detection," in *Intelligent Data Engineering and Automated Learning*, vol. 2412, H. Yin, N. Allinson, R. Freeman, J. Keane, and S. Hubbard, Eds. Berlin, Germany: Springer, 2002, pp. 378_383, doi:10.1007/3-540-45675-9_56.
[18] N. Kousika, G. Vishali, S. Sunandhana, and M. A. Vijay, ``Machine learning based fraud analysis and detection system," *J. Phys., Conf.*, vol. 1916, no. 1, May 2021, Art. no. 012115, doi: 10.1088/1742-6596/1916/1/012115.
[19] R. F. Lima and A. Pereira, ``Feature selection approaches to fraud detection in e-payment systems," in *E-Commerce and Web Technologies*, vol. 278, D. Bridge and H. Stuckenschmidt, Eds. Springer, 2017, pp. 111_126, doi: 10.1007/978-3-319-53676-7_9.
[20] Y. Lucas and J. Jurgovsky, ``Credit card fraud detection using machine learning: A survey," 2020, *arXiv:2010.06479*.
[21] H. Zhou, H.-F. Chai, and M.-L. Qiu, ``Fraud detection within bankcard enrollment on mobile device based payment using machine learning," *Frontiers Inf. Technol. Electron. Eng.*, vol. 19, no. 12, pp. 1537_1545, Dec. 2018, doi: 10.1631/FITEE.1800580.
[22] S. Makki, Z. Assaghir,Y. Taher, R. Haque, M.-S. Hacid, and H. Zeineddine, ``An experimental study with imbalanced classi_cation approaches for credit card fraud detection," *IEEE Access*, vol. 7, pp. 93010_93022, 2019, doi: 10.1109/ACCESS.2019.2927266.
[23] I. Matloob, S. A. Khan, and H. U. Rahman, ``Sequence mining and prediction-based healthcare fraud detection methodology," *IEEE Access*, vol. 8, pp. 143256_143273, 2020, doi: 10.1109/ACCESS.2020.3013962.
[24] I. Mekterovi¢, M. Karan, D. Pintar, and L. Brki¢, ``Credit card fraud detection in card-not-present transactions: Where to invest?" *Appl. Sci.*,vol. 11, no. 15, p. 6766, Jul. 2021, doi: 10.3390/app11156766.
[25] D. Molina, A. LaTorre, and F. Herrera, ``SHADE with iterative local search for large-scale global optimization," in *Proc. IEEE Congr. Evol. Comput. (CEC)*, Jul. 2018, pp. 1_8, doi: 10.1109/CEC.2018.8477755.