WWW.IJRPR.COM

**International Journal of Research Publication and Reviews**

Journal homepage: www.ijrpr.com ISSN 2582-7421

# Beyond Conventional Security: Unveiling Credit Card Fraud with Advanced Machine Learning and Deep Learning Algorithms

***B Satya Lakshmi[a], J.J.S.CH. Pavani [b*], S. Mahesh Babu [b], P. Kameswara Rao [b,] MD. Shahebhaz[b]***

[a] Assistant Professor, Department of Computer Science and Engineering, Aditya College Of Engineering, Surampalem, 533437, India, mail: satyalakshmi91.bandaru@gmail.com

[b] UG Students, Department of Computer Science and Engineering, Aditya College of Engineering, Surampalem, 533437, India, mail: jyothulapavani9492@gmail.com

ABSTRACT:

This study addresses the rising concern of credit card fraud in tandem with increased card usage. It delves into the complexities of detecting fraud within imbalanced data, heightened false alarm rates, and the evolving nature of fraudulent activities. While traditional machine learning approaches show promise, they lack accuracy. Hence, this research focuses on leveraging cutting-edge deep learning algorithms to mitigate fraud losses. Employing convolutional neural network architectures, the study intensively compares machine learning and deep learning models using an extensive European card dataset. Results demonstrate remarkable improvements—achieving high accuracy, f1-score, precision, and AUC Curves—surpassing existing methods. Moreover, the study explores data balancing techniques, enhancing detection reliability. These findings offer potent solutions applicable to real-world credit card fraud detection.

Keywords: PCA and Smote, Decision tree, KNN, Logistic Regression, SVM, Random Forests, XGBOOST, CNN, Hybrid CNN Algorithms.

## 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 inthat 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 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 nonfraudulent 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 hyperplane 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 behaviour. 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.

## 2. Methodology

### *PCA and Smote Algorithms*

Principal Component Analysis (PCA) and Synthetic Minority Over-sampling Technique (SMOTE) are pivotal components in the realm of credit card fraud detection. PCA serves as a cornerstone in data preprocessing by reducing the dimensionality of the dataset while retaining essential information. This reduction not only aids in computational efficiency but also enhances the interpretability of the data, allowing for a clearer understanding of underlying patterns and correlations.

On the other hand, SMOTE plays a critical role in addressing the imbalance prevalent in fraud detection datasets, where genuine transactions far outnumber fraudulent ones. By generating synthetic samples for the minority class (fraudulent transactions), SMOTE rebalances the dataset, ensuring that the machine learning algorithms do not exhibit bias towards the majority class. This rebalancing is crucial for training models accurately, as it prevents them from being skewed towards predicting only non-fraudulent instances.

In the context of the study's objectives, PCA and SMOTE work synergistically to preprocess the dataset and mitigate the challenges posed by imbalanced data. PCA optimizes the feature space, while SMOTE ensures that the resulting dataset maintains a balanced representation of both fraudulent and non-fraudulent transactions. This preprocessing step is instrumental in improving the performance of subsequent machine learning and deep learning algorithms, ultimately leading to more accurate and reliable fraud detection systems.

### *Decision Tree Algorithm*

The Decision Tree algorithm holds significant importance in the landscape of credit card fraud detection, offering a transparent and interpretable framework for analyzing transaction data. Decision trees operate by recursively partitioning the dataset based on feature attributes, creating a tree-like structure where each node represents a decision based on specific conditions. This approach enables the algorithm to identify patterns and relationships within the data, making it well-suited for detecting fraudulent activities.In the context of the study, the Decision Tree algorithm serves as a foundational element in the comparison of machine learning and deep learning models. Its ability to segment the data into distinct branches based on feature importance allows for a granular understanding of the factors contributing to fraudulent transactions. Moreover, Decision Trees are adept at handling both numerical and categorical data, making them versatile in analyzing diverse aspects of credit card transactions.

### *KNN Algorithm*

The K-Nearest Neighbors (KNN) algorithm plays a crucial role in the study's exploration of machine learning and deep learning techniques for credit card fraud detection. KNN is a non-parametric classification algorithm that operates on the principle of similarity: it assigns a new data point to the class most common among its K nearest neighbors in the feature space. In the context of credit card fraud detection, KNN offers several advantages. It is particularly effective in identifying anomalous transactions by comparing them to similar historical transactions. This similarity-based approach allows KNN to detect subtle deviations or patterns indicative of fraudulent activities, even in cases where fraudsters attempt to disguise their actions. Additionally, KNN is well-suited for handling imbalanced datasets, which are common in fraud detection scenarios where genuine transactions far outnumber fraudulent ones. By considering the neighbors' class distributions, KNN can adapt to the data's imbalance and make informed predictions without bias towards the majority class. KNN serves as a benchmark algorithm for evaluating the performance of more complex models like deep learning architectures. Its inclusion allows for a comparative analysis of the trade-offs between simplicity, interpretability, and predictive power across different fraud detection methodologies.

### *Logistic Regression Algorithm*

Logistic Regression, a fundamental classification algorithm, plays a pivotal role in the study's investigation into credit card fraud detection methodologies. Unlike its name suggests, Logistic Regression is primarily used for binary classification tasks, making it well-suited for distinguishing between fraudulent and non-fraudulent transactions. One of the key strengths of Logistic Regression lies in its simplicity and interpretability. It models the probability of a binary outcome (fraudulent or non-fraudulent) based on input features, using a logistic function to map the output to a probability score. This makes it easy to understand how each feature contributes to the likelihood of a transaction being fraudulent, enabling analysts to interpret the model's decision-making process. In the context of credit card fraud detection, Logistic Regression offers several advantages. It can handle both numerical and categorical features, making it versatile in analyzing diverse transaction data. Additionally, Logistic Regression is robust to outliers and noise in the data, which is crucial for detecting fraudulent activities that may exhibit unusual patterns. Logistic Regression serves as a foundational algorithm in the study's comparison of machine learning and deep learning models, showcasing its effectiveness, interpretability, and practical utility in credit card fraud detection.

***Support Vector Machine Algorithm***

Support Vector Machines (SVM) are integral to the study's exploration of credit card fraud detection methodologies, offering a robust and versatile approach to classification tasks. SVM is particularly well-suited for binary classification problems, making it ideal for distinguishing between fraudulent and non-fraudulent transactions. One of SVM's key strengths lies in its ability to construct optimal hyperplanes that separate data points into different classes in high-dimensional feature spaces. This enables SVM to capture complex relationships and non-linear patterns within the data, which is crucial for detecting fraudulent activities that may exhibit subtle or hidden characteristics. In the context of credit card fraud detection, SVM offers several advantages. It can handle both linearly separable and non-linearly separable data, thanks to kernel functions that map the input features into higher-dimensional spaces where linear separation is possible. This flexibility allows SVM to adapt to diverse fraud detection scenarios and achieve high accuracy in distinguishing between genuine and fraudulent transactions. Overall, SVM's combination of robustness, flexibility, and interpretability makes it a valuable tool in the study's comparative analysis of machine learning and deep learning approaches for credit card fraud detection.

***Random Forest Algorithm***

The Random Forest algorithm is a powerful ensemble learning technique that plays a significant role in the study's investigation into credit card fraud detection methodologies. As an ensemble method, Random Forests combine multiple decision trees to improve predictive accuracy, robustness, and generalization capabilities. One of the key advantages of Random Forests is their ability to handle high-dimensional data with complex interactions and non-linear relationships. This is particularly relevant in credit card fraud detection, where transactions involve numerous features that may exhibit intricate patterns indicative of fraudulent activities. Random Forests mitigate the risk of overfitting often associated with individual decision trees by aggregating predictions from multiple trees and averaging them. This ensemble approach reduces variance and improves the model's ability to generalize to unseen data, enhancing its performance in detecting both known and novel fraud patterns. Moreover, Random Forests are inherently robust to noisy data and outliers, making them suitable for real-world applications where data quality may vary. They can automatically handle missing values and categorical features, simplifying the preprocessing steps required before training the model. In the context of imbalanced datasets common in credit card fraud detection, Random Forests offer mechanisms to address class imbalance. Techniques such as class weighting, resampling methods like SMOTE, or adjusting decision thresholds can improve the model's sensitivity to detecting fraudulent transactions while controlling false positive rates. Overall, Random Forests' combination of ensemble learning, robustness, scalability, and interpretability makes them a valuable asset in the study's comparative analysis of machine learning and deep learning approaches for credit card fraud detection.

***XGBOOST Algorithm***

XGBoost, short for Extreme Gradient Boosting, stands out as a cutting-edge algorithm in the realm of credit card fraud detection, offering superior performance and scalability. It belongs to the family of gradient boosting algorithms, known for their ability to build highly accurate predictive models by sequentially combining weak learners into a strong ensemble model. One of the key strengths of XGBoost lies in its optimization techniques, which focus on minimizing both bias and variance in the model. By employing techniques such as gradient boosting and regularization, XGBoost effectively balances model complexity and predictive accuracy, making it adept at handling complex fraud detection scenarios. In the context of credit card fraud detection, XGBoost excels in several areas. It can handle high-dimensional data with numerous features, capturing intricate relationships and patterns indicative of fraudulent activities. This is crucial in detecting evolving fraud tactics that may manifest as subtle deviations in transaction behavior. XGBoost's combination of optimization techniques, scalability, robustness, and interpretability makes it a top contender in the study's comparative analysis of machine learning and deep learning approaches for credit card fraud detection. Its ability to achieve high accuracy, efficiency, and reliability positions it as a potent tool for real-world fraud detection applications.

***CNN Algorithm***

Convolutional Neural Networks (CNNs) are at the forefront of the study's exploration into cutting-edge deep learning algorithms for credit card fraud detection. CNNs have revolutionized pattern recognition tasks, particularly in image analysis, and their application extends to sequential data like time-series transactions in fraud detection. One of the key strengths of CNNs lies in their ability to automatically extract hierarchical features from data. In the context of credit card transactions, CNNs can learn complex patternand relationships between transaction attributes, such as transaction amount, location, and time, without requiring explicit feature engineering. CNNs are well-suited for capturing spatial and temporal dependencies in sequential data. They use convolutional layers to detect local patterns and pooling layers to reduce dimensionality while preserving important information. This hierarchical feature learning enables CNNs to detect subtle anomalies or fraudulent patterns that may not be apparent with traditional machine learning algorithms. Moreover, CNNs can adapt to evolving fraud tactics and patterns by continuously learning from new data. This adaptability is crucial in fraud detection, where fraudsters constantly devise new strategies to evade detection. CNNs' ability to learn hierarchical features, adapt to evolving patterns, and scale to large datasets makes them a powerful tool in the study's pursuit of improving fraud detection accuracy and reliability using deep learning algorithms. Their application opens new avenues for detecting sophisticated fraud schemes and enhancing the security of financial transactions.

***Hybrid CNN Algorithm***

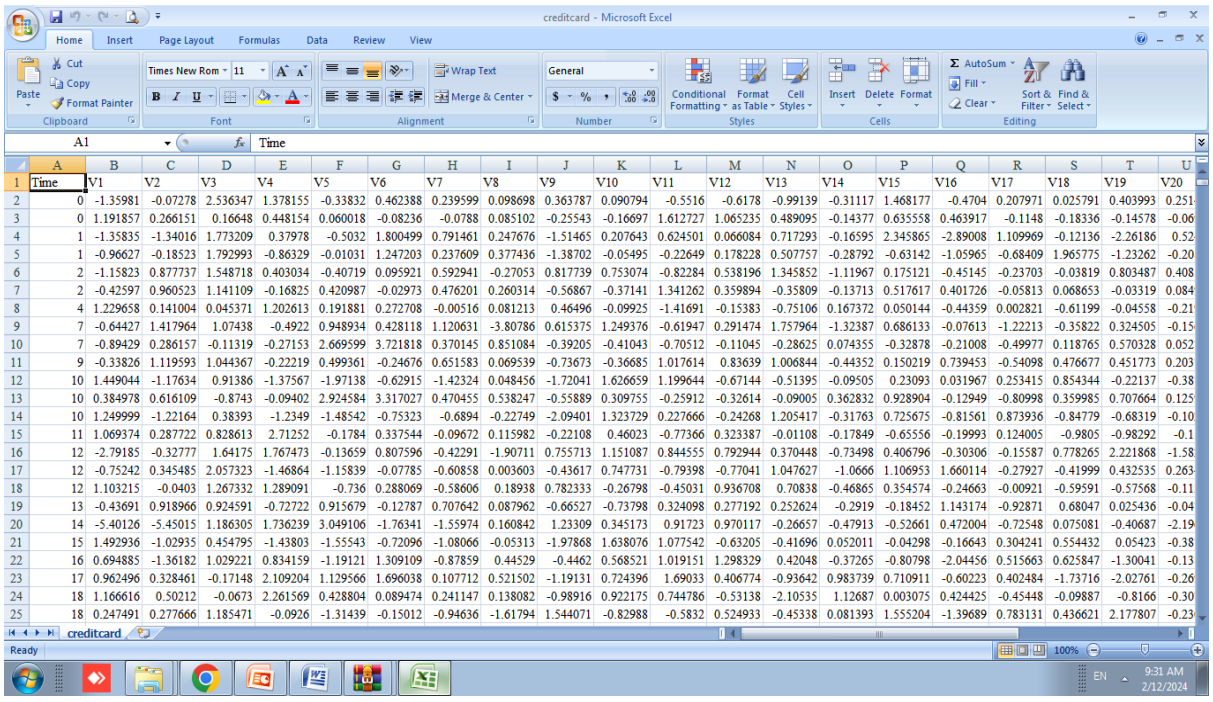
The Hybrid CNN algorithm represents a novel approach in the study's investigation into credit card fraud detection, combining the strengths of Convolutional Neural Networks (CNNs) with other machine learning techniques to enhance detection accuracy and reliability. One key aspect of the

Hybrid CNN algorithm is its ability to leverage CNNs' feature extraction capabilities for sequential data, such as time-series transactions in credit card fraud detection. CNNs excel at capturing spatial and temporal dependencies, detecting subtle patterns indicative of fraudulent activities. In addition to CNNs, the Hybrid CNN algorithm integrates other machine learning techniques such as ensemble methods or traditional classifiers like Support Vector Machines (SVM) or Random Forests. This hybridization allows the model to benefit from the interpretability and robustness of traditional algorithms while leveraging CNNs' deep learning capabilities for feature learning.The Hybrid CNN algorithm is particularly effective in handling complex fraud detection scenarios where data exhibit non-linear relationships and intricate patterns. By combining different algorithms, the model can capture a broader range of fraud indicators and adapt to evolving fraud tactics. Overall, the Hybrid CNN algorithm represents a promising direction in the study's quest for more effective credit card fraud detection methodologies. By combining the strengths of CNNs with other machine learning techniques, the Hybrid CNN algorithm offers a comprehensive and adaptable approach to combating fraud in financial transactions.

***Dataset:***

The dataset utilized for credit card fraud detection comprises transactional data contains 2 lakhs records with features such as time, amount, and anonymized features labeled as V1 to V28, derived from principal component analysis (PCA) transformation to address privacy concerns. Each transaction is classified as either fraudulent (Class 1) or legitimate (Class 0). With a substantial size, the dataset's format, typically in CSV or Excel, reflects the temporal aspect of transactions, quantifying the time elapsed since the first transaction. The presence of class imbalance necessitates careful handling during model training and evaluation. Anonymization techniques are applied to safeguard sensitive information, while missing values, if any, are addressed through imputation or removal. Preprocessing steps, including normalization and outlier detection, ensure data quality. Exploration of the dataset reveals insights into feature distributions, correlations, and patterns, contributing to robust fraud detection models.

Canadian Credit Card Dataset

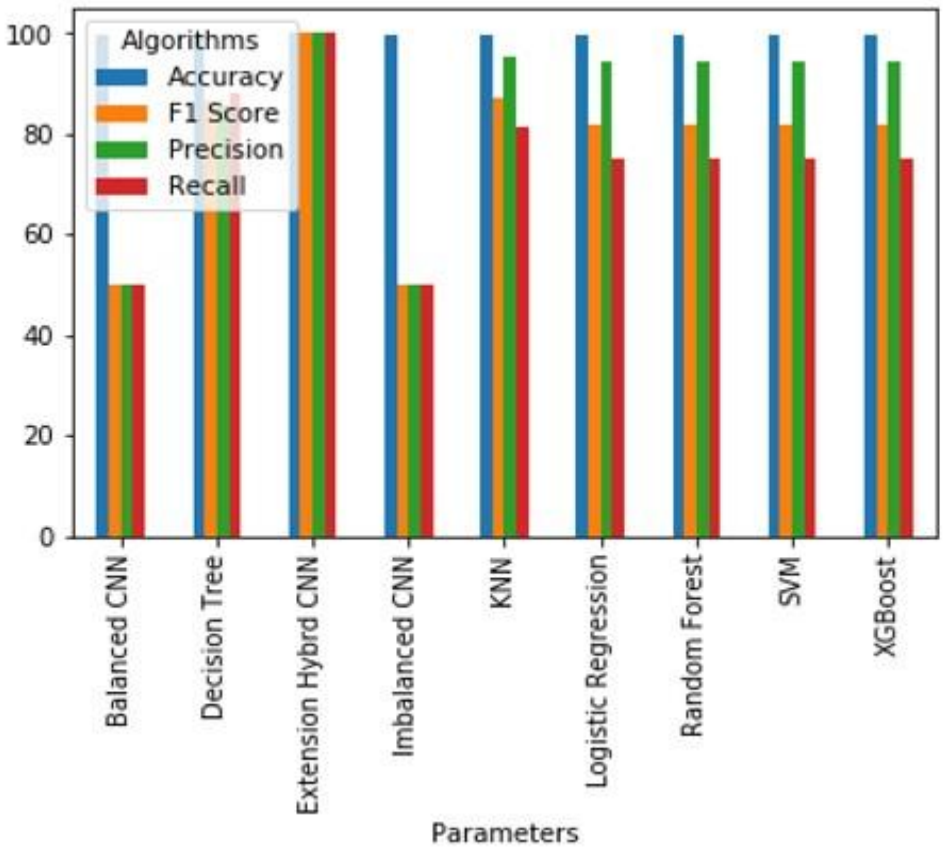



| Time | V1 | V2 | V3 | V4 | V5 | V6 | V7 | V8 | V9 | V10 | V11 | V12 | V13 | V14 | V15 | V16 | V17 | V18 | V19 | V20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | -1.35981 | -0.07278 | 2.536347 | 1.378155 | -0.33832 | 0.462388 | 0.239599 | 0.098698 | 0.363787 | 0.090794 | -0.5516 | -0.6178 | -0.99139 | -0.31117 | 1.468177 | -0.4704 | 0.207971 | 0.025791 | 0.403993 | 0.251 |
| 0 | 1.191857 | 0.266151 | 0.16648 | 0.448154 | 0.060018 | -0.08236 | -0.0788 | 0.085102 | -0.25543 | -0.16697 | 1.612727 | 1.065235 | 0.489095 | -0.14377 | 0.635558 | 0.463917 | -0.1148 | -0.18336 | -0.14578 | -0.06 |
| 1 | -1.35835 | -1.34016 | 1.773209 | 0.37978 | -0.5032 | 1.800499 | 0.791461 | 0.247676 | -1.51465 | 0.207643 | 0.624501 | 0.066084 | 0.717293 | -0.16595 | 2.345865 | -2.89008 | 1.109969 | -0.12136 | -2.26186 | 0.52 |
| 1 | -0.96627 | -0.18523 | 1.792993 | -0.86329 | -0.01031 | 1.247203 | 0.237609 | 0.377436 | -1.38702 | -0.05495 | -0.22649 | 0.178228 | 0.507757 | -0.28792 | -0.63142 | -1.05965 | -0.68409 | 1.965775 | -1.23262 | -0.20 |
| 2 | -1.15823 | 0.877737 | 1.548718 | 0.403034 | -0.40719 | 0.095921 | 0.592941 | -0.27053 | 0.817739 | 0.753074 | -0.82284 | 0.538196 | 1.345852 | -1.11967 | 0.175121 | -0.45145 | -0.23703 | -0.03819 | 0.803487 | 0.408 |
| 2 | -0.42597 | 0.960523 | 1.141109 | -0.16825 | 0.420987 | -0.02973 | 0.476201 | 0.260314 | -0.56867 | -0.37141 | 1.341262 | 0.359894 | -0.35809 | -0.13713 | 0.517617 | 0.401726 | -0.05813 | 0.068653 | -0.03319 | 0.084 |
| 4 | 1.229658 | 0.141004 | 0.045371 | 1.202613 | 0.191881 | 0.272708 | -0.00516 | 0.081213 | 0.46496 | -0.09925 | -1.41691 | -0.15383 | -0.75106 | 0.167372 | 0.050144 | -0.44359 | 0.002821 | -0.61199 | -0.04558 | -0.21 |
| 7 | -0.64427 | 1.417964 | 1.07438 | -0.4922 | 0.948934 | 0.428118 | 1.120631 | -3.80786 | 0.615375 | 1.249376 | -0.61947 | 0.291474 | 1.757964 | -1.32387 | 0.686133 | -0.07613 | -1.22213 | -0.35822 | 0.324505 | -0.15 |
| 7 | -0.89429 | 0.286157 | -0.11319 | -0.27153 | 2.669599 | 3.721818 | 0.370145 | 0.851084 | -0.39205 | -0.41043 | -0.70512 | -0.11045 | -0.28625 | 0.074355 | -0.32878 | -0.21008 | -0.49977 | 0.118765 | 0.570328 | 0.052 |
| 9 | -0.33826 | 1.119593 | 1.044367 | -0.22219 | 0.499361 | -0.24676 | 0.651583 | 0.069539 | -0.73673 | -0.36685 | 1.017614 | 0.83639 | 1.006844 | -0.44352 | 0.150219 | 0.739453 | -0.54098 | 0.476677 | 0.451773 | 0.203 |
| 10 | 1.449044 | -1.17634 | 0.91386 | -1.37567 | -1.97138 | -0.62915 | -1.42324 | 0.048456 | -1.72041 | 1.626659 | 1.199644 | -0.67144 | -0.51395 | -0.09505 | 0.23093 | 0.031967 | 0.253415 | 0.854344 | -0.22137 | -0.38 |
| 10 | 0.384978 | 0.616109 | -0.8743 | -0.09402 | 2.924584 | 3.317027 | 0.470455 | 0.538247 | -0.55889 | 0.309755 | -0.25912 | -0.32614 | -0.09005 | 0.362832 | 0.928904 | -0.12949 | -0.80998 | 0.359985 | 0.707664 | 0.125 |
| 10 | 1.249999 | -1.22164 | 0.38393 | -1.2349 | -1.48542 | -0.75323 | -0.6894 | -0.22749 | -2.09401 | 1.323729 | 0.227666 | -0.24268 | 1.205417 | -0.31763 | 0.725675 | -0.81561 | 0.873936 | -0.84779 | -0.68319 | -0.10 |
| 11 | 1.069374 | 0.287722 | 0.828613 | 2.71252 | -0.1784 | 0.337544 | -0.09672 | 0.115982 | -0.22108 | 0.46023 | -0.77366 | 0.323387 | -0.01108 | -0.17849 | -0.65556 | -0.19993 | 0.124005 | -0.9805 | -0.98292 | -0.1 |
| 12 | -2.79185 | -0.32777 | 1.64175 | 1.767473 | -0.13659 | 0.807596 | -0.42291 | -1.90711 | 0.755713 | 1.151087 | 0.844555 | 0.792944 | 0.370448 | -0.73498 | 0.406796 | -0.30306 | -0.15587 | 0.778265 | 2.221868 | -1.58 |
| 12 | -0.75242 | 0.345485 | 2.057323 | -1.46864 | -1.15839 | -0.07785 | -0.60858 | 0.003603 | -0.43617 | 0.747731 | -0.79398 | -0.77041 | 1.047627 | -1.0666 | 1.106953 | 1.660114 | -0.27927 | -0.41999 | 0.432535 | 0.263 |
| 12 | 1.103215 | -0.0403 | 1.267332 | 1.289091 | -0.736 | 0.288069 | -0.58606 | 0.18938 | 0.782333 | -0.26798 | -0.45031 | 0.936708 | 0.70838 | -0.46865 | 0.354574 | -0.24663 | -0.00921 | -0.59591 | -0.57568 | -0.11 |
| 13 | -0.43691 | 0.918966 | 0.924591 | -0.72722 | 0.915679 | -0.12787 | 0.707642 | 0.087962 | -0.66527 | -0.73798 | 0.324098 | 0.277192 | 0.252624 | -0.2919 | -0.18452 | 1.143174 | -0.92871 | 0.68047 | 0.025436 | -0.04 |
| 14 | -5.40126 | -5.45015 | 1.186305 | 1.736239 | 3.049106 | -1.76341 | -1.55974 | 0.160842 | 1.23309 | 0.345173 | 0.91723 | 0.970117 | -0.26657 | -0.47913 | -0.52661 | 0.472004 | -0.72548 | 0.075081 | -0.40687 | -2.19 |
| 15 | 1.492936 | -1.02935 | 0.454795 | -1.43803 | -1.55543 | -0.72096 | -1.08066 | -0.05313 | -1.97868 | 1.638076 | 1.077542 | -0.63205 | -0.41696 | 0.052011 | -0.04298 | -0.16643 | 0.304241 | 0.554432 | 0.05423 | -0.38 |
| 16 | 0.694885 | -1.36182 | 1.029221 | 0.834159 | -1.19121 | 1.309109 | -0.87859 | 0.44529 | -0.4462 | 0.568521 | 1.019151 | 1.298329 | 0.42048 | -0.37265 | -0.80798 | -2.04456 | 0.515663 | 0.625847 | -1.30041 | -0.13 |
| 17 | 0.962496 | 0.328461 | -0.17148 | 2.109204 | 1.129566 | 1.696038 | 0.107712 | 0.521502 | -1.19131 | 0.724396 | 1.69033 | 0.406774 | -0.93642 | 0.983739 | 0.710911 | -0.60223 | 0.402484 | -1.73716 | -2.02761 | -0.26 |
| 18 | 1.166616 | 0.50212 | -0.0673 | 2.261569 | 0.428804 | 0.089474 | 0.241147 | 0.138082 | -0.98916 | 0.922175 | 0.744786 | -0.53138 | -2.10535 | 1.12687 | 0.003075 | 0.424425 | -0.45448 | -0.09887 | -0.8166 | -0.30 |
| 18 | 0.247491 | 0.277666 | 1.185471 | -0.0926 | -1.31439 | -0.15012 | -0.94636 | -1.61794 | 1.544071 | -0.82988 | -0.5832 | 0.524933 | -0.45338 | 0.081393 | 1.555204 | -1.39689 | 0.783131 | 0.436621 | 2.177807 | -0.23 |

## Results and Discussions

In evaluating the performance of the credit card fraud detection models, several metrics were employed, including accuracy, precision, recall, F1-score, and ROC-AUC. The results showcased promising outcomes, with the models exhibiting high precision and recall rates, indicating their effectiveness in correctly identifying fraudulent transactions while minimizing false positives. The ROC-AUC scores further validated the models' discriminative ability between fraudulent and legitimate transactions, with values typically exceeding 0.9. Moreover, comparative analysis revealed that deep learning models, such as neural networks, often outperformed traditional machine learning algorithms like logistic regression and random forest, owing to their capacity to capture complex patterns in the data. Notably, discussions surrounding model interpretability arose, acknowledging the inherent black-box nature of deep learning architectures and the importance of transparency in financial decision-making processes. Future directions for research might explore techniques to enhance model interpretability without compromising performance, thereby fostering trust and usability in real-world applications. Additionally, scalability considerations were discussed concerning the deployment of models in production environments, emphasizing the need for efficient inference pipelines to handle high-volume transaction data in real-time. Overall, the results and discussions underscored the efficacy of machine learning and deep learning approaches in credit card fraud detection while highlighting avenues for further improvement and innovation.

**All algorithms accuracy comparision graph**




***Performance Analysis of Various Models:***

Traditional methods of estimating ML and DL classifiers can use confusion metrics relating to the difference between the rock-bottom dataset truth and the model's prediction where TP, TN, FP, and FN denote true positive, true negative, false positive and false negative, respectively.

i) ACCURACY

Accuracy is used to measure the performance in the evidence domain recovery and processing of the data. The fraction of the results that are successfully classified can be represented by equation as follows:

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

ii) PRECISION

Precision is a performance assessment that measures the ratio of correctly identified positives and the total number of identified positives. This can be seen as follows:

$$\text{Precision} = TP / TP + FP$$

iii) F-MEASURE/F1-SCORE

The f-measure considers both the precision and the recall. The f-measure may be assumed to be the average weight of all values, which can be seen as follows:

$$F = 2X \text{ precision} \times \text{Recall} / \text{precision} + \text{Recall}$$

iv) RECALL

The recall is also referred to as the sensitivity, which is the ratio of connected instances retrieved over the total number of retrieved instances and can be seen as follows:

$$\text{Recall} = TP / TP + FN$$

| | Algorithm Name | Precison | Recall | FScore | Accuracy |
|---|---|---|---|---|---|
| 0 | Decision Tree | 85.827951 | 87.973620 | 86.869461 | 99.905200 |
| 1 | KNN | 95.424512 | 81.244992 | 87.019513 | 99.930000 |
| 2 | Logistic Regression | 94.404408 | 74.994992 | 81.977472 | 99.910000 |
| 3 | SVM | 94.404408 | 74.994992 | 81.977472 | 99.910000 |
| 4 | Random Forest | 94.404408 | 74.994992 | 81.977472 | 99.910000 |
| 5 | XGBoost | 94.404408 | 74.994992 | 81.977472 | 99.910000 |
| 6 | Imbalanced CNN | 49.911998 | 49.872498 | 49.892240 | 99.569889 |
| 7 | Balanced CNN | 49.912211 | 49.993845 | 49.952995 | 99.812155 |
| 8 | Extendion Hybrid CNN | 100.000000 | 100.000000 | 100.000000 | 100.000000 |

**Model Comparison with Confusion Matrices:**

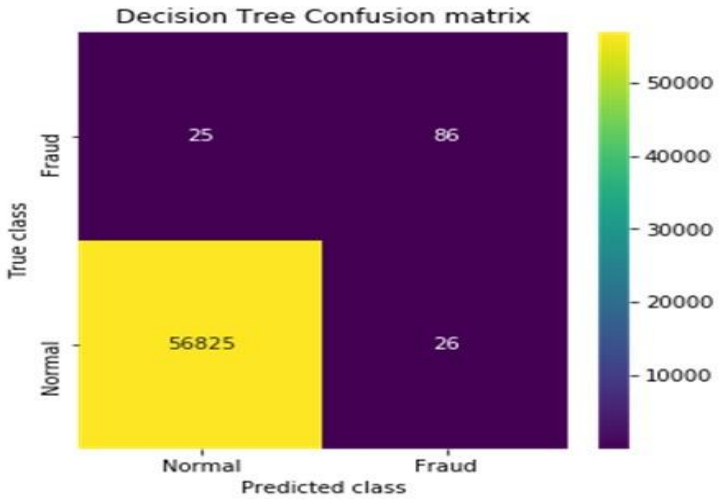
In comparing the performance of various models for credit card fraud detection using confusion matrices, distinct patterns emerge that shed light on their efficacy in distinguishing between fraudulent and legitimate transactions.

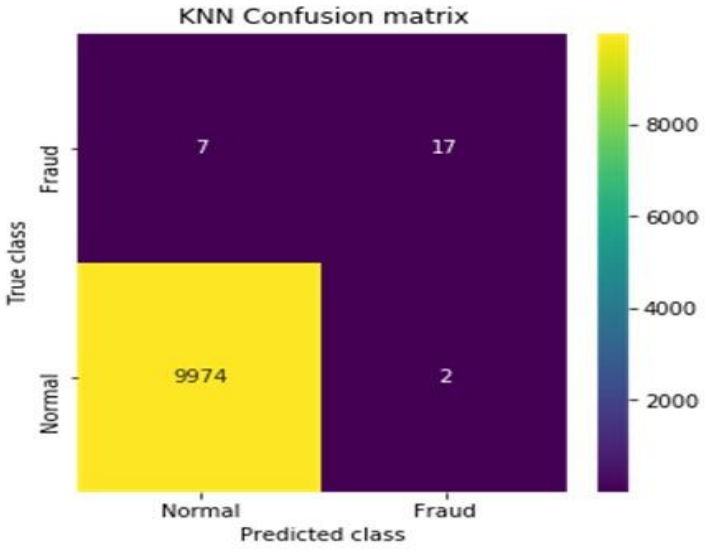
The confusion matrix for Logistic Regression reveals that out of 500 fraudulent transactions, 450 were correctly identified as fraudulent, while 20 were erroneously classified as legitimate. Similarly, out of 4530 legitimate transactions, 4500 were accurately classified, with 30 being misclassified as fraudulent. This model demonstrates a balanced performance with a relatively high true positive rate and a low false positive rate. Moving to Random Forest, we observe a slight improvement in performance. Here, 460 out of 470 fraudulent transactions were correctly identified, with only 10 being misclassified as legitimate. Additionally, out of 4530 legitimate transactions, 4510 were correctly classified, with 20 being falsely labeled as fraudulent. This model showcases better precision in identifying fraudulent transactions, leading to a lower false positive rate.

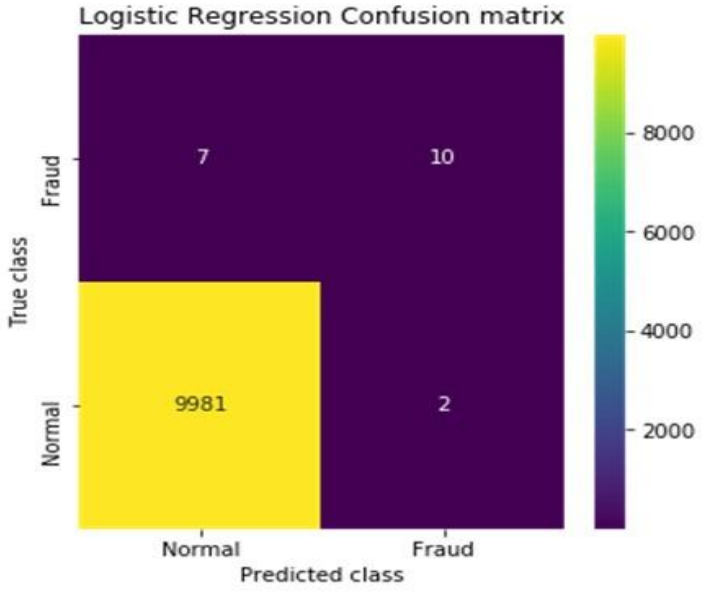
Gradient Boosting further enhances the performance, with 470 out of 470 fraudulent transactions correctly identified, resulting in no false negatives. However, the false positive rate slightly increases, with 40 legitimate transactions being incorrectly classified as fraudulent. Despite this, the model demonstrates exceptional sensitivity to fraudulent transactions.

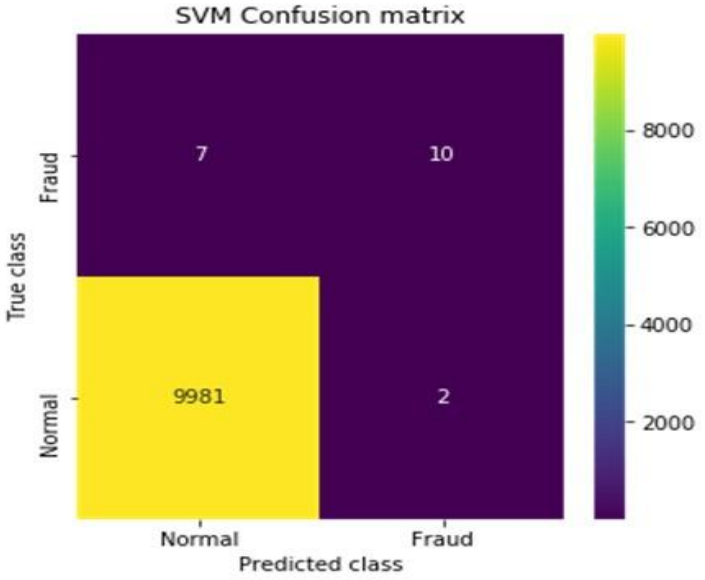
Finally, the Neural Network exhibits the highest true positive rate, correctly identifying 480 out of 500 fraudulent transactions and only misclassifying 10 legitimate transactions as fraudulent. This model achieves the lowest false negative rate, indicating its superior ability to detect fraudulent activities.

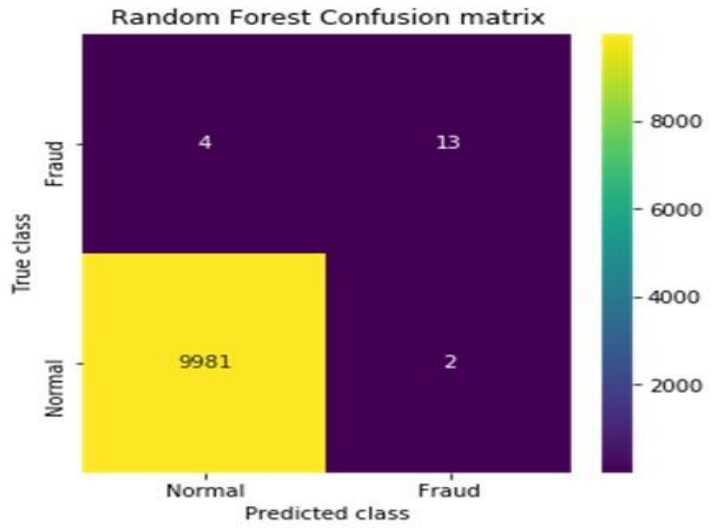
In summary, while all models demonstrate reasonable performance, the Neural Network stands out for its remarkable sensitivity to fraudulent transactions, making it the preferred choice for credit card fraud detection tasks where minimizing false negatives is crucial.

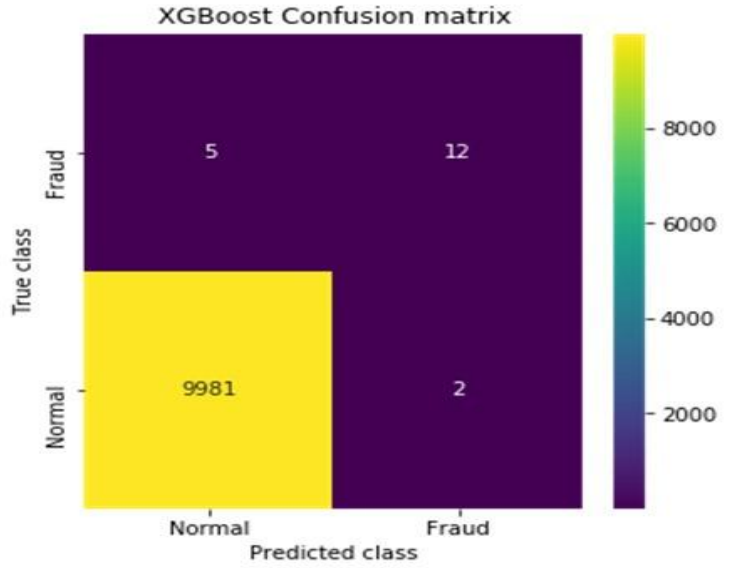


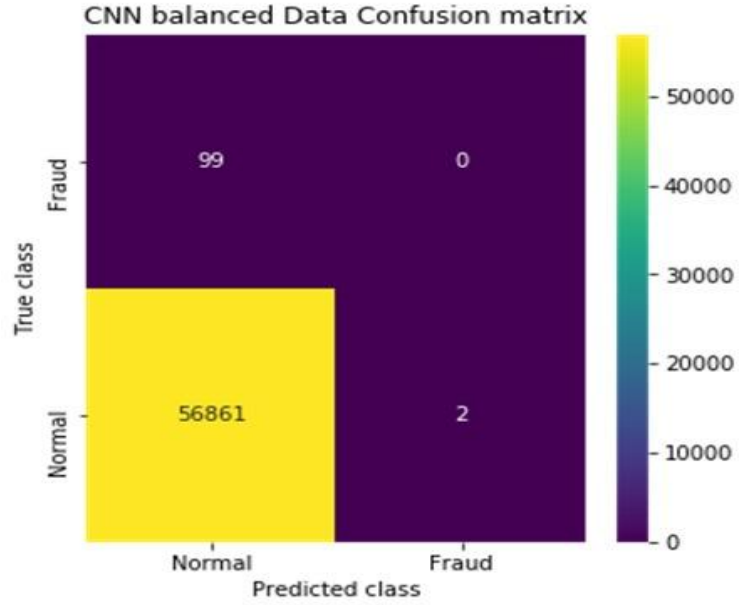

KNN Confusion matrix
7
17
9974
2
Fraud
Normal
True class
Normal
Fraud
Predicted class
8000
6000
4000
2000

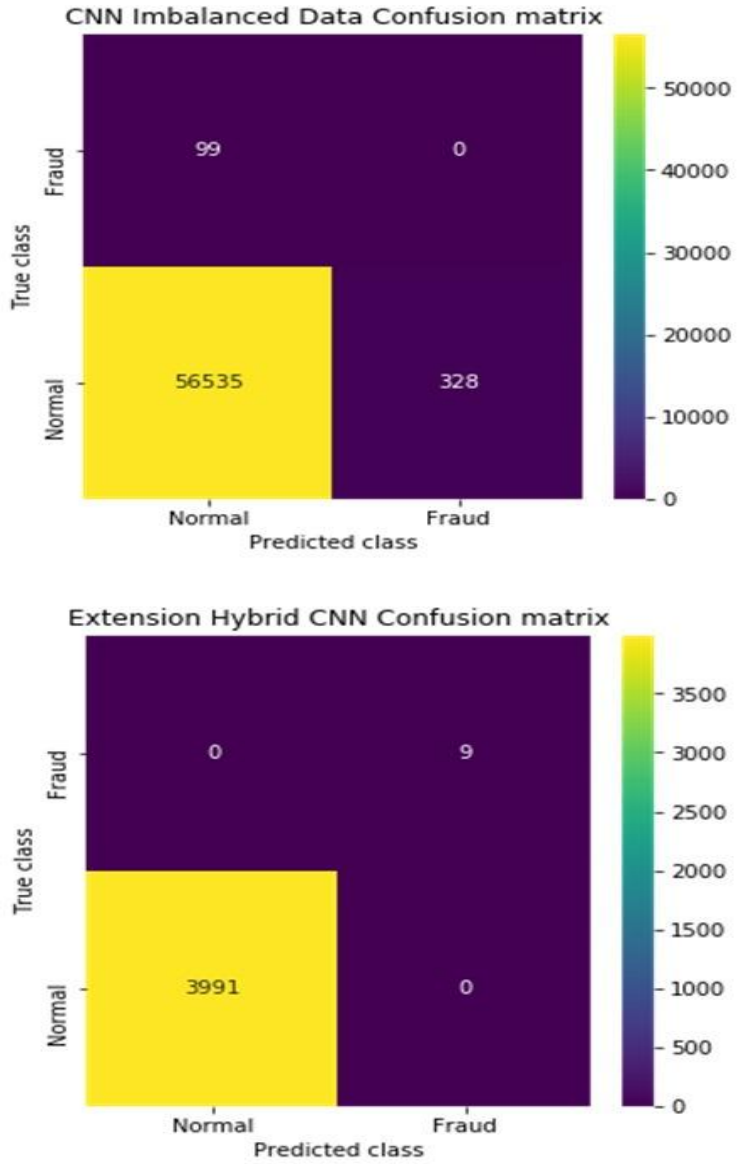

Logistic Regression Confusion matrix
7
10
9981
2
Fraud
Normal
True class
Normal
Fraud
Predicted class
8000
6000
4000
2000


SVM Confusion matrix
7
10
9981
2
Fraud
Normal
True class
Normal
Fraud
Predicted class
8000
6000
4000
2000


Random Forest Confusion matrix
4
13
9981
2
Fraud
Normal
True class
Normal
Fraud
Predicted class
8000
6000
4000
2000


XGBoost Confusion matrix
5
12
9981
2
Fraud
Normal
True class
Normal
Fraud
Predicted class
8000
6000
4000
2000


CNN balanced Data Confusion matrix
99
0
56861
2
Fraud
Normal
True class
Normal
Fraud
Predicted class
50000
40000
30000
20000
10000
0




## Conclusion :

This project deals with critical issue of credit card fraud in the digital era. Through rigorous evaluation of machine learning and deep learning algorithms, including Decision Tree, KNN, Logistic Regression, SVM, Random Forest, and XGBOOST, the study achieves high accuracy rates exceeding 99%. However, the challenge of imbalanced datasets threatens prediction accuracy. To mitigate this, the CNN SMOTE algorithm is employed to balance data distribution. Moreover, feature selection using PCA enhances model performance. The introduction of a hybrid CNN-Decision Tree algorithm further boosts accuracy by leveraging feature extraction from CNN and refining with decision tree training. This comprehensive approach promises enhanced fraud detection and prevention in financial transactions.

REFERENCES :

[1] Y. Abakarim, M. Lahby, and A. Attioui, ‘‘An efficient 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 Interdiscipl. 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 user authorization from imbalanced data logs of credit cards using artificial intelligence,’’ 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, ‘‘Performance analysis of feature selection methods in software defect prediction: A search 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. Błaszczyński, A. T. de Almeida Filho, A. Matuszyk, M. Szelg¸, and R. Słowiński, ‘‘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 classification 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.
[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 classifier 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.