Unveiling the Optimal Approach for Credit Card Fraud Detection: A thorough Analysis of Deep Learning and Machine Learning Methods

Ubaida Fatima, PhD
Assistant Professor
NED University of Engineering & Technology, Karachi, Pakistan

Sadia Kiran
Undergraduate Student
NED University of Engineering & Technology, Karachi, Pakistan

Muhammad Fouzan Akhter
Undergraduate Student
NED University of Engineering & Technology, Karachi, Pakistan

Muhammad Kumail
Undergraduate Student
NED University of Engineering & Technology, Karachi, Pakistan

Jaweria Sohail
Undergraduate Student
NED University of Engineering & Technology, Karachi, Pakistan

ABSTRACT

This study compared machine learning (ML) and deep learning (DL) techniques for credit card fraud detection. We evaluated 16 combinations of ML algorithms and cross-validation methods across diverse datasets. The Random Forest classifier with repeated K-fold cross-validation achieved the highest accuracy 99.0% and F1 score 99.1% among all models. The top performing deep learning model, the Artificial Neural Network (ANN), achieved an accuracy of 91.3% and F1 score of 91.1%, while a hybrid model combining these approaches reached 98.9% accuracy and F1 score. The Random Forest Classifier continued to be the best option. Our findings suggest the Random Forest classifier with repeated K-fold cross-validation, tested against a 21 combinations of other machine learning models, deep learning models, and a hybrid model as the most reliable method for credit card fraud detection in balanced datasets, offering valuable insights for enhancing security precautions and financial system defense against various banking sector frauds.

Keywords

Machine Learning methods, Fraud Detection, Deep learning models, Random Forest Classifier, ANN

1. INTRODUCTION

The ever-growing reliance on credit cards for transactions has unfortunately been accompanied by a surge in fraudulent activity. In 2024, financial institutions face a constant battle against increasingly sophisticated fraudsters. While numerous methods exist for credit card fraud detection, identifying the most effective and practical approach remains a challenge.

This research, titled "Unveiling the Optimal Approach for Credit Card Fraud Detection: A Thorough Analysis of Deep Learning and Machine Learning Methods," tackles this challenge head-on. We leverage the power of data mining to create and test a comprehensive suite of models across various machine learning and deep learning algorithms.

This machine learning exploration incorporates established algorithms like Logistic Regression, K-Nearest Neighbors (KNN), Decision Trees, Random Forests, Support Vector Machines (SVM), and Extreme Gradient Boosting (XGBoost). For the deep learning front, we delve into Artificial Neural Networks (ANNs), Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Multi-Layer Perceptrons (MLPs), and Long Short-Term Memory (LSTM) networks. By comparatively analyzing the performance of these diverse algorithms, we aim to unveil the optimal approach for credit card fraud detection in the current landscape.

This research goes beyond simply exploring individual algorithms. We recognize the critical role of balanced datasets in fraud detection, where legitimate transactions vastly outnumber fraudulent ones. By acknowledging this data imbalance and potentially implementing techniques to address it, our study strives to deliver practical insights that translate to real-world fraud detection systems.

1.1 The Evolving Challenge of Credit Card Fraud Detection

Credit card fraud poses a significant threat to financial institutions, leading to financial losses, regulatory penalties, and reputational damage. Effective fraud detection strategies are crucial for mitigating these risks.

This study examines established techniques like rule-based systems and anomaly detection, alongside the growing role of machine learning algorithms. We will also explore the challenges inherent in credit card fraud detection, such as the ever-evolving tactics of fraudsters and the inherent imbalance between legitimate and fraudulent transactions.

Financial institutions increasingly leverage supervised, unsupervised, and reinforcement learning techniques to identify and combat fraudulent activity. This research delves into these various machine learning approaches to identify the optimal method for credit card fraud detection in the current landscape.

1.2 Balancing Data Security and Privacy with Fraud Detection

Credit card security and fraud detection rely heavily on data mining and machine learning, raising concerns about customer data privacy. Financial institutions need robust security measures to protect sensitive information while maintaining stakeholder and customer trust.

The high volume of transactions and the inherent attractiveness of the banking sector to fraudsters necessitate effective fraud detection methods. Consequently, many institutions have increasingly turned to data mining and machine learning to enhance their fraud detection capabilities.

1.3 Research Objectives

- Identify the most effective machine learning and deep learning models for credit card fraud detection.
- Enhance model performance through data pre-processing and hyperparameter tuning.
- Evaluate the generalizability of the top performing models across diverse datasets.

2. LITERATURE REVIEW

Barmo et al. (2024) analyze and compare machine learning algorithms for credit card fraud detection. Recognizing the growing concern about fraud in the digital payment landscape, they explore the effectiveness of three algorithms: Logistic Regression, K-Nearest Neighbors (KNN), and Naive Bayes. Their evaluation focuses on metrics like accuracy, precision, recall, F1-score, and ROC curves. The study reveals that Naive Bayes achieves the highest overall accuracy (99.7%) and F1-score (0.375). However, they acknowledge the importance of considering both precision and recall depending on specific use cases. Furthermore, they investigate an ensemble learning approach by stacking the three models. This ensemble model achieves an impressive accuracy of 98.58%, suggesting the potential benefits of combining algorithms for improved performance. [1]

Khalid et al. (2024) propose an ensemble machine learning approach to address limitations in credit card fraud detection. They acknowledge shortcomings of current methods, including data imbalance and real-time processing challenges. Their ensemble model combines Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Random Forests (RF), Bagging, and Boosting algorithms. To handle data imbalance, they employ under-sampling and SMOTE. Evaluation on a European credit card transaction dataset demonstrates the ensemble's superior performance in accuracy, precision, recall, and F1-score compared to individual classifiers and traditional methods. This study highlights the potential of ensemble learning for credit card fraud detection, paving the way for more robust and adaptable systems. [2]

Paldino et al. (2024) address the challenge of evolving customer behavior and concept drift in credit card fraud detection using machine learning. They argue that standard techniques struggle to adapt to these changes, often discarding past knowledge. Their approach leverages diversity-based ensemble learning, aiming to preserve past concepts while enabling faster adaptation to new data distributions. They compare their method with other learning approaches on large real-world datasets provided by Worldline, a payment processing company. This research explores a novel ensemble learning approach that could enhance the adaptability of fraud detection systems in the face of evolving fraud patterns. [3]

Bao et al. (2024) propose a BERT-based deep learning model for credit card fraud detection. They highlight the challenges of imbalanced and high-dimensional datasets in this domain. Their approach leverages the pre-training capabilities of BERT to capture semantic similarities within transaction data, potentially improving fraud detection accuracy. Extensive data pre-processing and model training lead to a reported 99.95% accuracy on a non-specified dataset. This research emphasizes the potential of advanced deep learning techniques like BERT for credit card fraud detection in the evolving realm of internet finance. [4]

Aslam and Hussain (2024) evaluate the performance of various machine learning algorithms for credit card fraud detection. They acknowledge the rising concern of transaction fraud due to increased global trade and the potential of machine learning to combat this issue. Their study compares the effectiveness of Logistic Regression, Random Forest, Extra Trees, and Light Gradient Boosting Machine (LGBM) on a publicly available dataset of European credit card transactions. Notably, the reported performance metrics (accuracy, recall, F1-score of 1.00 for both classes) suggest exceptional results, although real-world applications might yield more nuanced outcomes. This research provides valuable insights into the capabilities of established machine learning algorithms for credit card fraud detection. [5]

Sulaiman et al. (2024) propose deep learning models with hyperparameter tuning for credit card fraud detection. They acknowledge the increasing complexity of fraud and the limitations of existing deep learning approaches. Their research focuses on optimizing hyperparameters for three deep learning models: AutoEncoder (AE), Convolutional Neural Network (CNN), and Long Short-Term Memory (LSTM). Evaluating these models on a European credit card fraud dataset, they observe that LSTM outperforms AE and CNN in accuracy (99.2%), detection rate (93.3%), and AUC (96.3%). This study highlights the potential of hyperparameter tuning to enhance the performance of deep learning models for credit card fraud detection, with LSTMs demonstrating promising results. [6]

Gupta et al. (2023) address the challenge of imbalanced data in credit card fraud detection using machine learning. They acknowledge the prevalence of imbalanced datasets, where legitimate transactions significantly outnumber fraudulent ones. Their research investigates the impact of various data balancing techniques (oversampling, undersampling, SMOTE) on the performance of machine learning models for credit card fraud detection. They employ XGBoost, a powerful machine learning algorithm, and observe promising results with Random Oversampling, achieving a reported accuracy and precision score of 0.99. This study highlights the importance of data balancing techniques in conjunction with machine learning models for improved credit card fraud detection, particularly when dealing with imbalanced datasets. [7]

Patel (2023) presents a comprehensive review of credit card analytics, focusing on fraud detection and risk assessment techniques. The review acknowledges the growing importance of advanced analytics in the financial sector, particularly for credit card security. It highlights various methodologies for credit card fraud detection, outlining their strengths and limitations. Additionally, the paper explores credit risk assessment techniques, emphasizing the need for robust models to predict potential loan defaults. Data management is recognized as a critical aspect, with high-quality data being essential for accurate modeling. The review acknowledges ongoing challenges like data inconsistencies and evolving fraud tactics. However, it also explores potential solutions, including big data analytics tailored for the financial sector. Finally, Patel (20XX) discusses promising future research directions in credit card analytics, aiming to keep the industry on the cutting edge. [8]

Dayyabu et al. (2023) investigate the application of artificial intelligence (AI) techniques in credit card fraud detection. They acknowledge the challenges faced by the financial industry due to the high volume of transactions and difficulty in identifying fraudulent activity. Their research explores the relationship between three AI techniques - machine learning, data mining, and fuzzy logic - and credit card fraud detection using a survey of 100 accounting and finance professionals. The data is analyzed using regression analysis and correlation coefficients. Their findings suggest a significant positive association between all three AI techniques and fraud detection, with machine learning and data mining perceived as more accurate/precise than fuzzy logic. [9]

Ahmad et al. (2023) propose a class balancing framework for credit card fraud detection using machine learning algorithms. They acknowledge the growing threat of fraud, particularly during COVID-19 due to increased cashless transactions. Their research focuses on addressing the imbalanced nature of credit card fraud data (where fraudulent transactions are a minority). They propose a framework using Fuzzy C-means clustering and Similarity-Based Selection (SBS) to create a more balanced dataset. The framework is evaluated using various machine learning algorithms, with Artificial Neural Networks (ANN) achieving the highest accuracy (0.966) compared to Logistic Regression (LR), Naive Bayes (NB), and k-Nearest Neighbors (kNN) (Ahmad et al., 20XX). This research highlights the importance of data balancing techniques for improving the performance of machine learning models in credit card fraud detection. [10]

Mienye and Sun (2023) propose a hybrid feature selection method to improve credit card fraud detection using machine learning. They acknowledge the detrimental effect of irrelevant features in real-world credit card data on model performance. Their approach combines filter-based feature selection (Information Gain) with a wrapper method using a Genetic Algorithm (GA) and Extreme Learning Machine (ELM) to identify the most relevant features. Notably, the GA is optimized for imbalanced data using the G-mean metric. This research emphasizes the importance of feature selection techniques for enhancing the effectiveness of machine learning models in credit card fraud detection. Their proposed method achieves high sensitivity (0.997) and specificity (0.994), outperforming existing methods. [11]

Afriyie et al. (2023) investigate the use of supervised machine learning for credit card fraud detection and prediction. They acknowledge the growing threat of fraud in credit card transactions and the potential of machine learning to combat this issue. Their research compares the performance of three algorithms: Logistic Regression, Random Forest, and Decision Trees. Evaluating these models on a non-specified dataset, they report that Random Forest achieves the highest accuracy (96%) and AUC (98.9%) in fraud detection. Additionally, they observe a higher prevalence of fraud victims among credit card holders above 60 years old and during nighttime hours (22:00GMT-4:00GMT). This study highlights the potential of Random Forest for credit card fraud detection, while also offering insights into potential user demographics and timeframes associated with fraudulent activity. [12]

Bakhtiari et al. (2023) focus on ensemble learning methods for credit card fraud detection, addressing the growing complexity of fraud in the financial sector. They acknowledge the widespread use of credit cards and the critical need for robust fraud detection systems. Their research explores ensemble learning techniques, specifically combining Light Gradient Boosting Machine (LightGBM) and LiteMORT algorithms using averaging methods (simple and weighted). Evaluation metrics like AUC, recall, F1-score, precision, and accuracy are employed. The study reports promising results, with the best performing ensemble achieving an accuracy of 99.44%. This research contributes to the exploration of ensemble methods for credit card fraud detection, demonstrating their potential to improve accuracy and efficiency. [13]

Sulaiman et al. (2022) present a review of machine learning approaches for credit card fraud detection (CCFD) with a focus on data privacy. They acknowledge the rise of credit card fraud alongside the growth of e-commerce and the crucial role of machine learning in fraud detection. Their review highlights the prevalence of supervised learning methods like SVM, KNN, Naive Bayes, Logistic Regression, and Decision Trees for CCFD. They emphasize the potential benefits of hybrid approaches over single algorithms. Furthermore, they recognize the challenges of data imbalance and heterogeneity in achieving high accuracy. Their proposed solution involves a federated learning framework with an Artificial Neural Network (ANN) to address privacy concerns during model training on real-time data. [14]

Khan et al. (2022) develop a credit card fraud detection model using machine learning approaches. They acknowledge the rise of e-commerce and the prevalence of credit card fraud, emphasizing the need for robust detection systems. Their research explores three supervised machine learning algorithms: Logistic Regression, Support Vector Machine (SVM), and Artificial Neural Network (ANN). They address the challenge of imbalanced class data (genuine vs. fraudulent transactions) using a resampling technique. The performance of each model is evaluated using various metrics like accuracy, precision, recall, F1-score, Matthews Correlation Coefficient (MCC), and ROC curve. Their findings suggest that SVM achieves the best overall performance, with higher precision, recall, F1-score, and MCC compared to Logistic Regression and ANN. [15]

3. METHODOLOGY

This research performs a comprehensive evaluation of machine learning and deep learning models for credit card fraud detection. The methodology unfolds in the following stages:

3.1 Model Selection

A broad spectrum of machine learning and deep learning models will be examined. This encompasses established machine learning algorithms like Logistic Regression, Decision Trees, Random Forests, K-Nearest Neighbors (KNN), XGBoost, Support Vector Machines (SVM), and Naive Bayes. Additionally, the study will explore deep learning architectures including Artificial Neural Networks (ANNs), Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Multi-Layer Perceptrons (MLPs), and Long Short-Term Memory (LSTM) networks. The specific parameters associated with each model will be optimized during the training process.

3.2 Data Acquisition and Preprocessing

Three credit card transaction datasets will be utilized in the study. The first dataset, obtained from Kaggle, is synthetic and will be used for initial model exploration. The second dataset, also from Kaggle, is a real-world dataset with sensitive features anonymized. This means sensitive information like column headers and values will be hidden or transformed. The values themselves may also be standardized to ensure consistency across features. The third dataset, provided by a bank, has all sensitive information completely removed. This dataset will undergo feature engineering, where new features are created from existing data to potentially enhance model performance. All datasets will be preprocessed to ensure data quality and consistency before being used for model training. This preprocessing may involve handling missing values, normalization or standardization of features, and potential feature scaling techniques.

To mitigate overfitting and provide a robust assessment of model performance, stratified K-Fold Cross-Validation and Repeated K-Fold Cross-Validation techniques will be employed during model training and evaluation. Each model will be trained and evaluated using these cross-validation techniques. The corresponding parameters for each model will be optimized during training to achieve the best possible performance.

The performance of each model will be evaluated using a variety of metrics:

- Accuracy Score: Measures the overall correctness of the model's predictions.
- F1-Score: A harmonic mean between precision and recall, accounting for both.
- Precision Score: Measures the proportion of true positives among predicted positives.
- Recall Score: Measures the proportion of true positives identified by the model.
- ROC AUC (Area Under the Curve): Represents the model's ability to discriminate between fraudulent and genuine transactions.
- Threshold: The decision point used to classify transactions as fraudulent or legitimate.

3.4 Model Selection and Analysis

By systematically analyzing the average performance of the aforementioned metrics across all datasets and models, the research will identify the model(s) that demonstrate the most effective performance in credit card fraud detection. This analysis may involve further exploration of the top performing models, including investigating the impact of hyperparameter tuning and potential feature engineering techniques on specific datasets.

3.4.1 Application of Machine Learning Models (MLMs) on Synthetic Dataset

Due to the scarcity of publicly available financial datasets, a synthetic dataset (PaySim) obtained from Kaggle was used for initial exploration. Data preprocessing steps were implemented to ensure model compatibility. Non-numeric values were identified and removed using the drop function. Categorical data in the "type" column was transformed into numerical features for the model using one-hot encoding via the get_dummies function in pandas. Numerical features (excluding binary outcome and transaction type columns) were scaled using the StandardScaler function to improve model performance. Since the dataset was imbalanced, under-sampling was employed to balance the classes while preserving data integrity. The preprocessed data was then split into training (80%) and testing (20%) sets.

Four machine learning models - Logistic Regression, Decision Trees, Random Forests, and K-Nearest Neighbors - were applied to the training data. The performance of each model was evaluated using accuracy, F1-score, precision, and recall metrics (Table 1). The Random Forest model emerged as the most effective, achieving an accuracy score of approximately 99.42%, F1-score of 99.42%, precision of 98.97%, and recall of 99.88%. Based on this promising performance on the synthetic dataset, the Random Forest model will be further evaluated using real-world datasets in subsequent stages of the research.

3.3 Model Training and Evaluation

Table 1: Results from Application of Machine Learning Algorithms on the Synthetic Dataset

Algorithm	Accuracy Score	Precision Score	Recall Score	F1 Score
Logistic Regression	≃ 93.396%	≃ 94.598%	≃ 91.998%	≃ 93.280%
Decision Trees	≃ 99.209%	≃ 99.026%	≃ 99.389%	≃ 99.207%
Random Forest	≃ 99.422%	≃ 98.971%	≃ 99.878%	≃ 99.422%
K-Nearest Neighbors	≃ 92.940%	≃ 93.231%	≃ 92.547%	≃ 92.888%

3.4.2 Applications of Machine Learning Models on Real Dataset (Sensitive Features Anonymized)

This research investigates the effectiveness of various machine learning algorithms for credit card fraud detection. The study utilizes a real-world dataset containing credit card transactions from September 2013 across Europe. The dataset consists of 284,807 transactions, with a significant imbalance as fraudulent transactions only account for 0.172% of the data (positive class). Due to confidentiality concerns, details about the original features and context of the data cannot be disclosed. The key features include V1 to V28, which are numerical input variables transformed using Principal Component Analysis (PCA). Additionally, the dataset contains "Amount" representing the transaction amount and "Time" indicating the elapsed seconds since the first transaction. The "Class" variable is the response variable, labeled as 0 for legitimate transactions and 1 for fraudulent transactions.

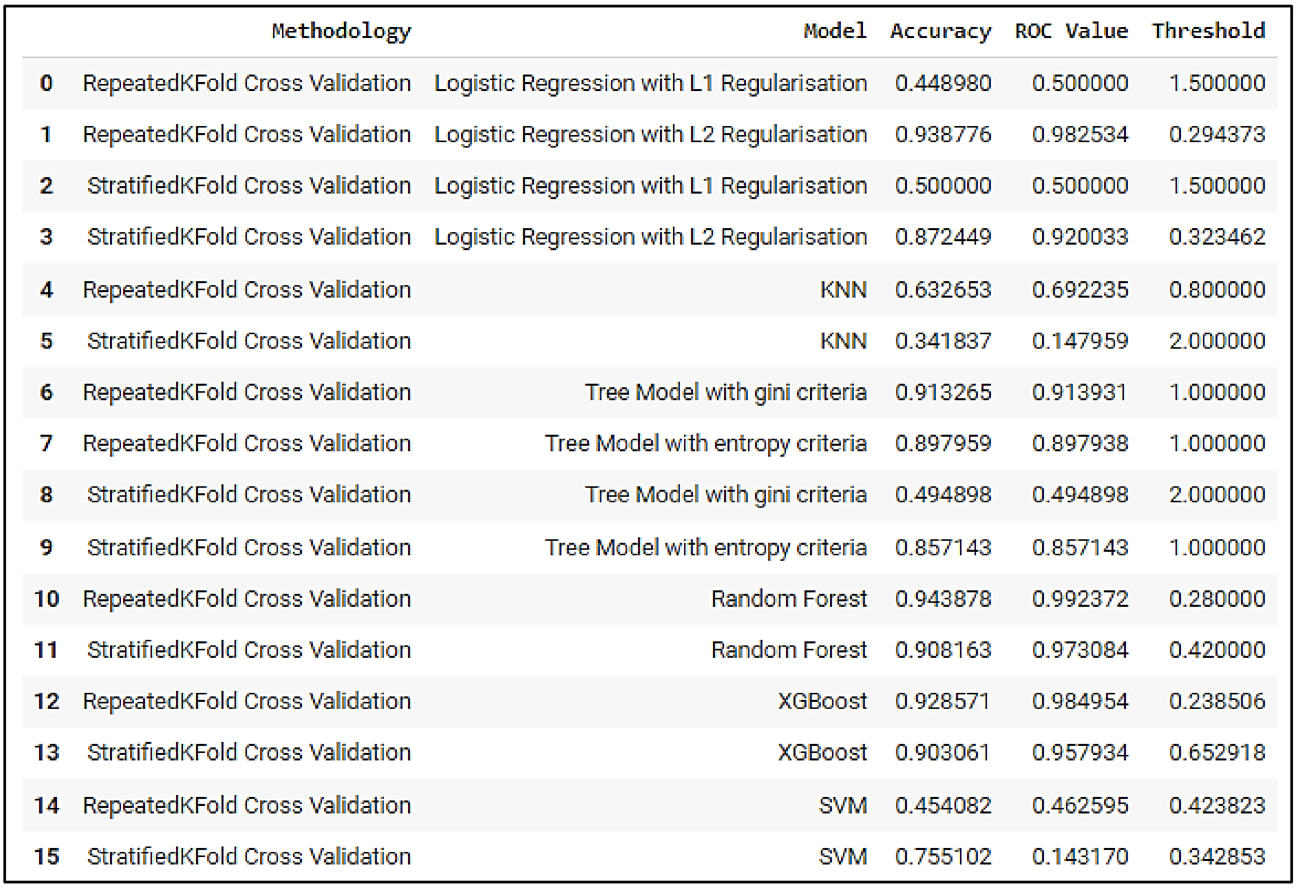
Eight machine learning algorithms were evaluated: Random Forest Classifier, XGBoost Classifier, Decision Tree Classifiers (Gini and Entropy criteria), Logistic Regression (L1 and L2 regularization) and Support Vector Machine (SVM) with Sigmoid Kernel. Hyperparameter tuning was performed to optimize the performance of each model. Repeated K-Fold Cross Validation and Stratified K-Fold Cross Validation were employed to assess the generalizability of the models. The evaluation metrics included confusion matrix, accuracy score, ROC (Receiver Operating Characteristic) value, and threshold. A heatmap was generated to visualize the correlation between the dataset's features, aiding in understanding the interdependencies between factors.

The time variable was converted from its relative representation to minutes for better interpretability. Standard functions were then used to split the data into training and testing sets, followed by model training and execution. The results from all model and cross-validation combinations were stored in a data frame, including the model used, methodology, accuracy score, ROC value, and threshold. These metrics were used to identify the best performing model for this specific dataset.

The selection of the most suitable machine learning model considered three key factors: accuracy score, ROC value, and threshold. Accuracy score indicates overall performance, but for imbalanced datasets, ROC value is crucial as it measures the ability to distinguish between true positives and false positives. Finally, the threshold determines the probability of flagging a transaction as fraudulent and involves a trade-off between false positives and false negatives. The optimal threshold depends on the specific application's priorities and risk tolerance.

After evaluating all models with different cross-validation techniques (Table 2), the Random Forest Classifier with Repeated K-Fold Cross Validation emerged as the most effective model. This approach achieved the highest accuracy score and ROC value, demonstrating superior performance in both overall classification and discrimination of fraudulent transactions. It is important to acknowledge that the optimal model for credit card fraud detection can vary depending on factors like data size, imbalance, balancing methods, and included features. This study provides a framework for evaluating various machine learning algorithms using appropriate metrics to select the best model for a specific dataset.

Table 2: Results from the 16 Combinations of Machine Learning Algorithms & Cross Validation Techniques on Real Dataset (Sensitive Features Anonymized)



	Methodology	Model	Accuracy	ROC Value	Threshold
0	RepeatedKFold Cross Validation	Logistic Regression with L1 Regularisation	0.448980	0.500000	1.500000
1	RepeatedKFold Cross Validation	Logistic Regression with L2 Regularisation	0.938776	0.982534	0.294373
2	StratifiedKFold Cross Validation	Logistic Regression with L1 Regularisation	0.500000	0.500000	1.500000
3	StratifiedKFold Cross Validation	Logistic Regression with L2 Regularisation	0.872449	0.920033	0.323462
4	RepeatedKFold Cross Validation	KNN	0.632653	0.692235	0.800000
5	StratifiedKFold Cross Validation	KNN	0.341837	0.147959	2.000000
6	RepeatedKFold Cross Validation	Tree Model with gini criteria	0.913265	0.913931	1.000000
7	RepeatedKFold Cross Validation	Tree Model with entropy criteria	0.897959	0.897938	1.000000
8	StratifiedKFold Cross Validation	Tree Model with gini criteria	0.494898	0.494898	2.000000
9	StratifiedKFold Cross Validation	Tree Model with entropy criteria	0.857143	0.857143	1.000000
10	RepeatedKFold Cross Validation	Random Forest	0.943878	0.992372	0.280000
11	StratifiedKFold Cross Validation	Random Forest	0.908163	0.973084	0.420000
12	RepeatedKFold Cross Validation	XGBoost	0.928571	0.984954	0.238506
13	StratifiedKFold Cross Validation	XGBoost	0.903061	0.957934	0.652918
14	RepeatedKFold Cross Validation	SVM	0.454082	0.462595	0.423823
15	StratifiedKFold Cross Validation	SVM	0.755102	0.143170	0.342853

3.4.3 Application of Machine Learning Models on Real Dataset (Sensitive Information Removed – Feature Engineered)

The next dataset for this research was obtained from a bank and included credit card transactions from throughout 2022. Due to data sensitivity, most fields were removed, leaving only the essential details: MCC code, transaction date and time, transaction amount, and a binary indicator for fraudulent transactions. A separate spreadsheet provided definitions for each MCC code, which categorize transactions by type (e.g., 7311 for advertising services). It's important to note that a single category can have multiple MCC codes (e.g., 3009 and 3024 for airlines). The dataset included 407 unique MCC codes across 22 transaction categories.

Since the dataset size exceeded Microsoft Excel's limit, Power Query was used to access it. A balanced subset of 7500 fraudulent and 7500 legitimate transactions was extracted for analysis and model creation. This under-sampling technique addressed the data imbalance issue. However, the extracted data showed all fraudulent transactions listed consecutively, followed by legitimate ones. To eliminate this bias during model training, the "randomizer" function was used. Random real values were added to a new "randomizer" column, and the data was shuffled based on ascending sort of this column. This ensured a balanced dataset with a random distribution of fraudulent and legitimate transactions.

A new column named "CKEY" was created in the spreadsheet to represent the three-letter category abbreviations corresponding to MCC codes. Power Query was then used to merge this data with the spreadsheet containing MCC code definitions based on the matching "MCC Code" field. This process extracted category names and corresponding CKEYs from the original data, creating a new merged workbook.

Data pre-processing was necessary to handle non-numeric data types. Python analysis revealed that only "MCC Code" and "sum" were numerical, while the rest were objects. Feature engineering was employed to address this. The "CKEY" column, originally textual data, was converted into numeric format using one-hot encoding. This technique creates binary vectors from categorical variables. A new binary column is generated for each unique category, with "1" indicating the presence of a specific category and "0" indicating its absence. This allows machine learning models (MLMs) to handle categorical data during training and prediction.

The get_dummies function from the pandas library was used for one-hot encoding. The "CKEY" column contained 22 distinct values (e.g., "RTS," "GVS," "BSS"). One-hot encoding transformed these values into separate column headers, essentially converting the "CKEY" column from textual to numeric data. After this step, both the "CKEY" and "CATEGORY" columns were removed as they contained redundant information. The "IS_FRAUD" column, indicating fraudulent transactions ("YES") or not ("NO"), was converted from text to numeric data by replacing "YES" with "1" and "NO" with "0".

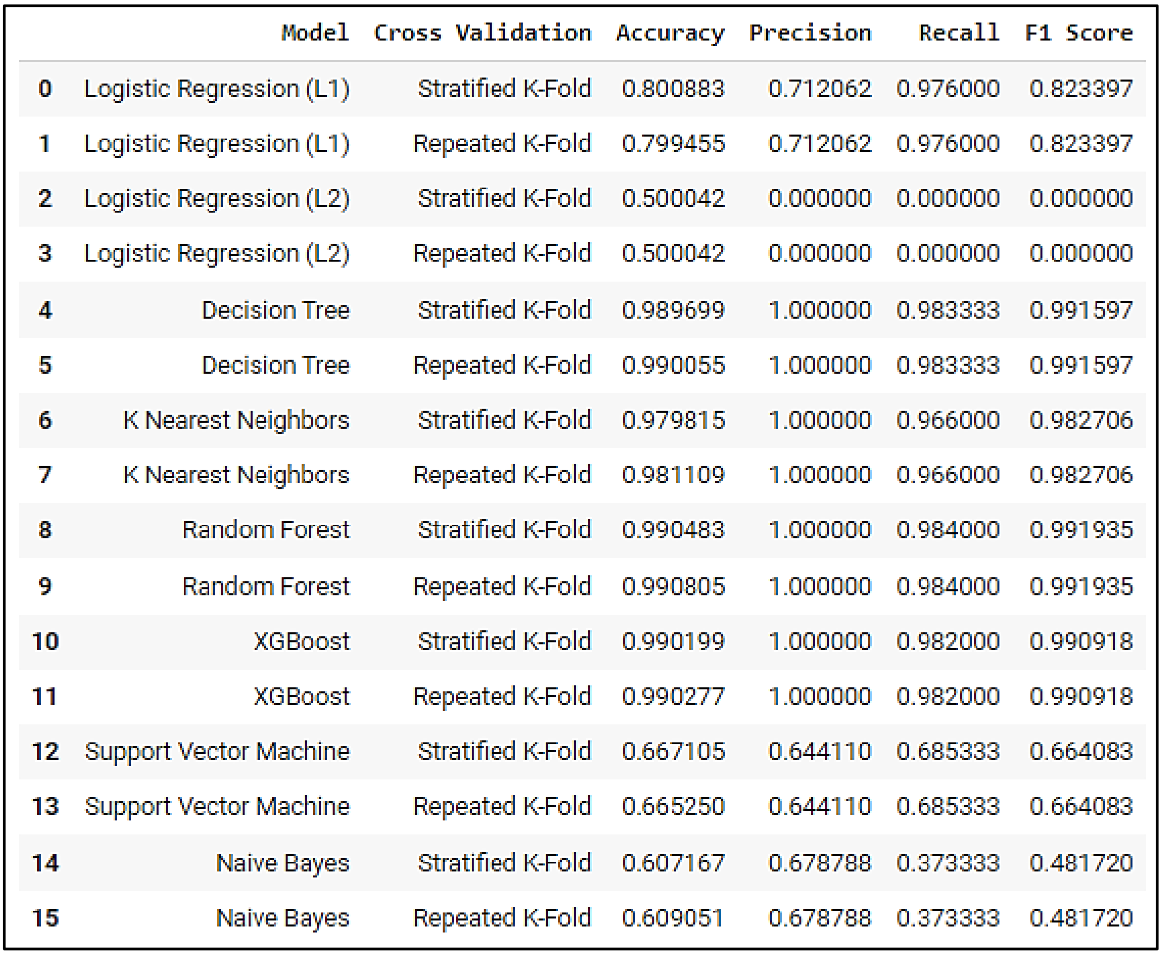
The "Transaction Date" column, originally in date-time format, was converted to numeric data using a pandas library function. Additionally, the "SUM" column, representing the bank's cash outflow (negative values), was multiplied by -1 to ensure all values were positive. This data manipulation resulted in a new dataset with all features in a numerical format, suitable for MLM application.

To ensure consistency and fair comparisons between features, scikit-learn's StandardScaler was used for standard scaling. This technique adjusts numerical features by setting their mean to 0 and standard deviation to 1. Standard scaling is crucial because many MLMs are sensitive to feature scales. When features have a wide range, some may dominate others during learning, leading to biased results. Standard scaling addresses this, allowing algorithms to function effectively and improving model accuracy.

Following scaling, the dependent and independent variables were separated. All features except the dependent variable "IS_FRAUD" were considered independent variables. The data was then split into training and testing sets, with a 20% test size. This ensures that only a portion of the data is used to train the MLMs, while the remaining portion is used solely for evaluation.

Eight machine learning algorithms were implemented for credit card fraud detection: Decision Trees, K-Nearest Neighbors (KNN), Random Forest, XGBoost, Support Vector Machine (SVM), Naive Bayes, Logistic Regression with L1 and L2 regularization. Two cross-validation techniques, repeated KFold and stratified KFold, were applied to each algorithm, resulting in 16 total combinations. For each combination, evaluation metrics including accuracy, precision, recall, and F1 score were computed. These results were stored in a new data frame (Table 3).

Table 3: Results Obtained from the 16 Combinations of Machine Learning Algorithms & Cross Validation Techniques on Real Dataset (Sensitive Information Removed – Feature Engineered)



	Model	Cross Validation	Accuracy	Precision	Recall	F1 Score
0	Logistic Regression (L1)	Stratified K-Fold	0.800883	0.712062	0.976000	0.823397
1	Logistic Regression (L1)	Repeated K-Fold	0.799455	0.712062	0.976000	0.823397
2	Logistic Regression (L2)	Stratified K-Fold	0.500042	0.000000	0.000000	0.000000
3	Logistic Regression (L2)	Repeated K-Fold	0.500042	0.000000	0.000000	0.000000
4	Decision Tree	Stratified K-Fold	0.989699	1.000000	0.983333	0.991597
5	Decision Tree	Repeated K-Fold	0.990055	1.000000	0.983333	0.991597
6	K Nearest Neighbors	Stratified K-Fold	0.979815	1.000000	0.966000	0.982706
7	K Nearest Neighbors	Repeated K-Fold	0.981109	1.000000	0.966000	0.982706
8	Random Forest	Stratified K-Fold	0.990483	1.000000	0.984000	0.991935
9	Random Forest	Repeated K-Fold	0.990805	1.000000	0.984000	0.991935
10	XGBoost	Stratified K-Fold	0.990199	1.000000	0.982000	0.990918
11	XGBoost	Repeated K-Fold	0.990277	1.000000	0.982000	0.990918
12	Support Vector Machine	Stratified K-Fold	0.667105	0.644110	0.685333	0.664083
13	Support Vector Machine	Repeated K-Fold	0.665250	0.644110	0.685333	0.664083
14	Naive Bayes	Stratified K-Fold	0.607167	0.678788	0.373333	0.481720
15	Naive Bayes	Repeated K-Fold	0.609051	0.678788	0.373333	0.481720

The final results revealed that the Random Forest Classifier with Repeated KFold Cross Validation emerged as the most effective model for credit card fraud detection in this dataset. This conclusion is based on several key observations:

- Consistent Performance: Across multiple runs, the Random Forest Classifier with Repeated KFold Cross Validation consistently demonstrated superior performance. This suggests its reliability and generalizability for fraud detection in this specific dataset.
- High Evaluation Metrics: The model achieved consistently high scores on all evaluation metrics – accuracy, precision, recall, and F1 score. This indicates its ability to correctly identify fraudulent transactions while minimizing both false positives and false negatives.
- Comparison with Other Models: When compared to the other seven machine learning algorithms tested, the Random Forest Classifier consistently outperformed them across both cross-validation techniques.

3.4.4 Application of Deep Learning Models on Real Dataset (Sensitive Information Removed – Feature Engineered)

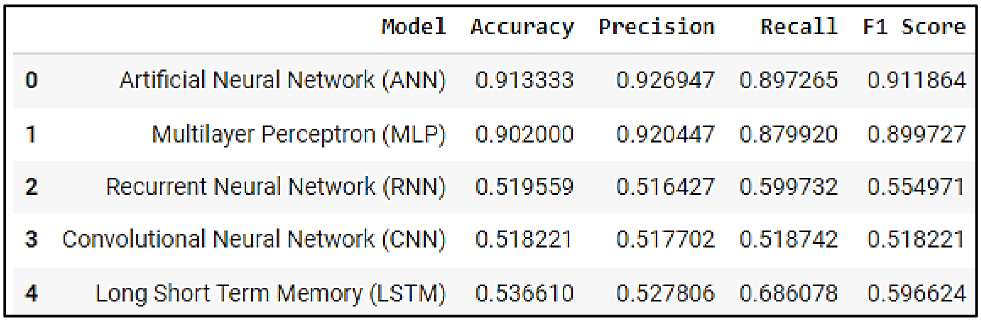
Following the identification of Random Forest as the most effective model among various machine learning algorithms for credit card fraud detection, this research explores the potential of deep learning approaches for further improvement. The analysis utilizes the dataset without modifications to ensure consistency with previous findings.

Deep learning models, a type of artificial intelligence (AI), mimic the structure and function of the human brain. Designed to analyze and learn from vast amounts of data for accurate predictions, these models consist of multiple layers of interconnected nodes, called artificial neurons. Each neuron performs simple mathematical operations on the input data before sending the output to the next layer, progressively building more complex representations of the input data.

Five deep learning models – Artificial Neural Networks (ANN), Multi-Layer Perceptron (MLP), Recurrent Neural Networks (RNN), Convolutional Neural Networks (CNN), and Long Short-Term Memory (LSTM) – will be evaluated on the HBL dataset. The goal is to determine if these models outperform the Random Forest Classifier in fraud detection. By maintaining the data in its original state and avoiding further manipulation, the methodology ensures a fair comparison between deep learning and machine learning techniques.

The deep learning algorithms will be applied to the data, and the results will be stored in a data frame. Performance metrics including accuracy, precision, recall, and F1 score will be calculated for each model. The results of the five deep learning models will be compiled into a data frame (Table 4). Initial examination suggests that the Artificial Neural Network (ANN) achieved the best performance, followed closely by the Multi-Layer Perceptron (MLP). All five models were trained for 20 epochs, with accuracy, loss, validation accuracy, and validation loss recorded for each epoch. These observations provide valuable insights into the training performance of each model.

Table 4: Results Obtained from the five Different Deep Learning Algorithms



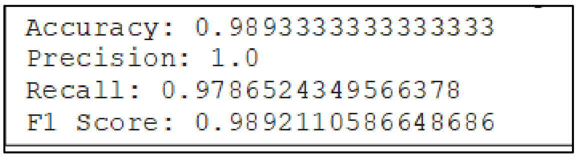
	Model	Accuracy	Precision	Recall	F1 Score
0	Artificial Neural Network (ANN)	0.913333	0.926947	0.897265	0.911864
1	Multilayer Perceptron (MLP)	0.902000	0.920447	0.879920	0.899727
2	Recurrent Neural Network (RNN)	0.519559	0.516427	0.599732	0.554971
3	Convolutional Neural Network (CNN)	0.518221	0.517702	0.518742	0.518221
4	Long Short Term Memory (LSTM)	0.536610	0.527806	0.686078	0.596624

3.4.5 Enhancing Fraud Detection with a Hybrid Machine Learning and Deep Learning Approach

Building upon the success of the Random Forest classifier (RANDOM FOREST CLASSIFIER) and the Artificial Neural Network (ANN) as individual models for credit card fraud detection on the Habib Bank Limited (HBL) dataset, this research explores the development of a hybrid model. The RANDOM FOREST CLASSIFIER emerged as the most effective machine learning algorithm (MLA) among 16 tested, while the ANN achieved superior performance compared to other deep learning models (5 tested). This hybrid approach aims to leverage the strengths of both techniques to potentially improve fraud detection accuracy and interpretability.

To achieve this, the output from the Random Forest Classifier was fed into the ANN using a fully connected layer, essentially training the ANN on the Random Forest Classifier's predictions. While the resulting model (Table 5) demonstrated improved performance compared to the standalone ANN, it did not surpass the individual performance of the Random Forest Classifier. This figure also depicts the final outcome and relative importance of the evaluated methods.

Table 5: Results from the Random Forest – Artificial Neural Network Hybrid Algorithm



```
Accuracy: 0.9893333333333333
Precision: 1.0
Recall: 0.9786524349566378
F1 Score: 0.9892110586648686
```

4. CONCLUSION

This research investigated machine learning and deep learning techniques for credit card fraud detection across three datasets. The first dataset was synthetic, the second came from Kaggle, and the third was provided by a bank.

For the real-world datasets, which were imbalanced, a combination of under-sampling, one-hot encoding, and scaling was applied for data pre-processing. Sixteen different combinations of machine learning algorithms and cross-validation techniques were evaluated on these datasets. The Random Forest classifier with repeated KFold cross-validation emerged as the most effective model, consistently outperforming other machine learning algorithms in terms of accuracy, precision, recall, and F1 score.

On the dataset obtained from the bank, five deep learning algorithms were explored alongside the machine learning models. While the Artificial Neural Network (ANN) demonstrated promising results, the Random Forest classifier still achieved superior performance. A hybrid model combining the Random Forest classifier's output with the ANN was also investigated, but it did not surpass the individual performance of the Random Forest classifier.

This study highlights the Random Forest classifier with repeated KFold cross-validation as the most reliable approach for credit card fraud detection on balanced datasets. These findings offer valuable insights for researchers and practitioners in the field, potentially leading to improved security measures and enhanced protection of financial systems against fraudulent activities.

5. DECLARATIONS

i. **Availability of Data and Materials:**
The data analyzed during this study is available from the corresponding author upon reasonable request. References detailing the source of the data are also included within the manuscript.

ii. **Competing Interests:**
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

iii. **Funding:**
This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

iv. **Authors' Contributions:**

1. **Dr. Ubaida Fatima:** Conceptualized the study, designed the methodology, and provided overall supervision.
2. **Muhammad Fouzan Akhter:** Performed data manipulation, implemented the machine learning, deep learning and hybrid models and compiled the results.
3. **Sadia Kiran:** Conducted the literature review, drafted the manuscript and aided in the implementation of the machine learning and deep learning models.
4. **Muhammad Kumail:** Aided in data manipulation.
5. **Jaweria Sohail:** Aided in literature review.

6. ACKNOWLEDGMENT

We would like to express our sincere gratitude to Dr. Ubaida Fatima PhD, our supervisor, whose unwavering support and guidance were instrumental to the successful completion of this research. Dr. Fatima's insightful direction, particularly in the areas of machine learning, deep learning, and hybrid models, not only shaped the methodology but also significantly deepened our understanding of the research domain. We are deeply grateful for her patience, encouragement, and invaluable feedback throughout the entire research process, from literature review and manuscript drafting to the final submission.

7. REFERENCES

[1] Ahmad Umar Barmo , Ahmad Haruna , Yusuf Umar Wali , Konika Abid "Analysis and Comparison of Fraud Detection on Credit Card Transactions Using Machine Learning Algorithms" Iconic Research And Engineering Journals Volume 7 Issue 8 2024 Page 293-299

[2] Khalid, A.R.; Owoh, N.; Uthmani, O.; Ashawa, M.; Osamor, J.; Adejoh, J. Enhancing Credit Card Fraud Detection: An Ensemble Machine Learning Approach. Big Data Cogn. Comput. 2024, 8, 6. https://doi.org/10.3390/bdcc8010006

[3] Paldino, G.M., Lebichot, B., Le Borgne, YA. et al. The role of diversity and ensemble learning in credit card fraud detection. Adv Data Anal Classif 18, 193–217 (2024). https://doi.org/10.1007/s11634-022-00515-5

[4] Bao, Q., Wei, K., Xu, J., & Jiang, W. (2024). Application of Deep Learning in Financial Credit Card Fraud Detection. Journal of Economic Theory and Business Management, 1(2), 51–57. https://doi.org/10.5281/zenodo.10960092

[5] Aslam, A., Hussain, A. (2024). A performance analysis of machine learning techniques for credit card fraud detection. Journal on Artificial Intelligence, 6(1), 1-21. https://doi.org/10.32604/jai.2024.047226

[6] Sulaiman, S.S., Nadher, I., Hameed, S.M. (2024). Credit card fraud detection using improved deep learning models. Computers, Materials & Continua, 78(1), 1049-1069. https://doi.org/10.32604/cmc.2023.046051

[7] Palak Gupta, Anmol Varshney, Mohammad Rafeek Khan, Rafeeq Ahmed, Mohammed Shuaib, Shadab Alam, Unbalanced Credit Card Fraud Detection Data: A Machine Learning-Oriented Comparative Study of Balancing Techniques, Procedia Computer Science, Volume 218, 2023, Pages 2575-2584, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2023.01.231

[8] Kaushikkumar Patel, "Credit Card Analytics: A Review of Fraud Detection and Risk Assessment Techniques," International Journal of Computer Trends and Technology, vol. 71, no. 10, pp. 69-79, 2023. Crossref, https://doi.org/10.14445/22312803/IJCTT-V71I10P109

[9] The application of artificial intelligence techniques in credit card fraud detection: a quantitative study, Yusuf Yusuf Dayyabu, Dhamayanthi Arumugam, Suresh Balasingam, E3S Web of Conf. 389 07023 (2023), DOI: 10.1051/e3sconf/202338907023

[10] Ahmad, H., Kasasbeh, B., Aldabaybah, B. et al. Class balancing framework for credit card fraud detection based on clustering and similarity-based selection (SBS). Int. j. inf. tecnol. 15, 325–333 (2023). https://doi.org/10.1007/s41870-022-00987-w

[11] Mienye, I.D.; Sun, Y. A Machine Learning Method with Hybrid Feature Selection for Improved Credit Card Fraud Detection. Appl. Sci. 2023, 13, 7254. https://doi.org/10.3390/app13127254

[12] Jonathan Kwaku Afriyie, Kassim Tawiah, Wilhemina Adoma Pels, Sandra Addai-Henne, Harriet Achiaa Dwamena, Emmanuel Odame Owiredu, Samuel Amening Ayeh, John Eshun, A supervised machine learning algorithm for detecting and predicting fraud in credit card transactions, Decision Analytics Journal, Volume 6, 2023,

100163, ISSN 2772-6622, https://doi.org/10.1016/j.dajour.2023.100163

[13] Bakhtiari, S., Nasiri, Z. & Vahidi, J. Credit card fraud detection using ensemble data mining methods. Multimed Tools Appl 82, 29057–29075 (2023). https://doi.org/10.1007/s11042-023-14698-2

[14] Bin Sulaiman, R., Schetinin, V. & Sant, P. Review of Machine Learning Approach on Credit Card Fraud Detection. Hum-Cent Intell Syst 2, 55–68 (2022). https://doi.org/10.1007/s44230-022-00004-0

[15] Shahnawaz Khan, Abdullah Alourani, Bharavi Mishra, Ashraf Ali and Mustafa Kamal, “Developing a Credit Card Fraud Detection Model using Machine Learning Approaches” International Journal of Advanced Computer Science and Applications(IJACSA), 13(3), 2022. http://dx.doi.org/10.14569/IJACSA.2022.0130350