Leakage-Free Evaluation of a One-Dimensional CNN for Credit Card Fraud Detection under Extreme Class Imbalance

Karen Ochuwa Ohwomado
Department of Computer and Software Technology
Faculty of Computing
Delta State University
Abraka, Nigeria

Maureen Ifeanyi Akazue
Department of Computer and Software Technology
Faculty of Computing
Delta State University
Abraka, Nigeria

Arnold Adimabua Ojugo
Institute of Information and Communication Technology
Federal University of Petroleum Resources, Effurun
Delta State Nigeria

ABSTRACT
Existing fraud detection studies frequently report inflated performance due to information leakage arising from improper handling of class imbalance techniques. This study presents a leakage-free evaluation of a one-dimensional Convolutional Neural Network (CNN) for credit card fraud detection under extreme class imbalance using the Kaggle Credit Card Fraud Detection dataset, which contains 284,807 anonymized real-world transactions. The CNN was evaluated against Logistic Regression, Decision Tree, Random Forest, and Gradient Boosting baselines. To prevent information leakage during model evaluation, the Synthetic Minority Over-sampling Technique (SMOTE) was restricted entirely to the training phase of the validation process. Model performance was assessed using the Matthews Correlation Coefficient (MCC) and ROC-AUC, as both provide a more reliable assessment of imbalanced classification tasks than conventional accuracy. Under these controlled conditions, the CNN achieved an MCC of 0.7081 and a ROC-AUC of 0.9659. The CNN produced the highest MCC among the evaluated models, reflecting stronger minority-class discrimination than the benchmark methods. The findings establish a reproducible benchmark for evaluating CNN-based fraud detection models under severe class imbalance while minimizing the risk of information leakage.

General Terms
Machine Learning, Deep Learning, Fraud Detection, Financial Security

Keywords
Credit Card Fraud Detection; Convolutional Neural Network; SMOTE; Information Leakage; Matthews Correlation Coefficient (MCC); ROC-AUC.

1. INTRODUCTION
Global card-related fraud losses are expected to surpass USD 43 billion by 2028 [1], [2]. Although encryption and financial security mechanisms continue to improve, the growth of digital banking platforms has also increased exposure to fraudulent activities [3]. In real-world financial systems, effective fraud detection remains challenging due to evolving attack strategies and the severe class imbalance commonly found in transaction data [4], [5]. Traditional machine learning algorithms and manual screening approaches often struggle with complex, high-volume fraud detection tasks [6], while the extremely low proportion of fraudulent transactions causes many classifiers to underperform on the minority class [4], [7].

Deep learning models, particularly Convolutional Neural Networks (CNNs), are increasingly applied to financial security tasks because they can automatically extract hierarchical feature representations from transaction data [8], [9]. When adapted for tabular datasets, convolutional layers capture localized feature interactions and non-linear dependencies that conventional linear models often overlook [10]. Instead of treating features independently, each transaction record is represented as a flat feature vector and passed through the network, allowing one-dimensional convolutional filters to scan across neighbouring attributes and identify underlying transactional patterns [10].

Tabular datasets present distinct challenges for deep learning because feature relationships are often weaker and less structured than those found in image or text data [11]. Financial transaction records are a clear example of this. They exist primarily as flat, fixed-length vectors with no inherent sequential structure. Even so, configuring a CNN to scan across related feature subsets provides a systematic way to identify meaningful transactional patterns [12]. Prior studies have reported promising results when CNNs are applied directly to individual transaction records [10]. Findings across the literature remain inconsistent. Direct comparison with established baseline models therefore remains important.

In electronic banking fraud detection, CNN-based models have become increasingly relevant [6], primarily because they extract relevant features automatically without manual intervention [8], [10], [13]. However, comparing these models across different papers is difficult. Very few studies evaluate CNN-based fraud detection models on widely used public datasets under identical, controlled experimental conditions [14]. This study evaluates a standalone one-dimensional CNN against Logistic Regression, Decision Tree, Random Forest, and Gradient Boosting under extreme class imbalance. The extreme class imbalance is addressed by shifting the primary evaluation focus away from standard accuracy metrics. Instead, the Matthews Correlation Coefficient (MCC), supported by ROC-AUC and F1-score, is prioritized to ensure reliable assessment of minority-class discrimination [15], [16].

The objectives of this study are to:

- Evaluate the feasibility of a CNN-based model for fraud detection under extreme class imbalance;
- Compare its performance against conventional classifiers (Logistic Regression and Decision Tree) and strong ensemble-based models (Random Forest and Gradient Boosting)**;** and

- Test model stability using stratified cross-validation on a single public fraud detection dataset.

1.1 Research Contributions

1. SMOTE resampling is restricted strictly to the training folds of a stratified validation setup. This operational control establishes a leakage-free experimental framework, eliminating data leakage while ensuring all baseline models are evaluated under identical preprocessing conditions.

2. Replicating or comparing existing 1D-CNN benchmarks is often difficult. Most published models rely on private transaction repositories or loosely documented data splits [6], [8], [9], [10]. This study addresses this ambiguity by evaluating a standalone 1D-CNN on the public Kaggle Credit Card Fraud Detection dataset [17]. An open baseline is established with MCC treated as the primary evaluation metric[15], [16].

3. The 1D-CNN is evaluated against four baseline models: Logistic Regression, Decision Trees [8], Random Forests, and Gradient Boosting [18]. This comparative setup isolates the performance of the CNN**.** Most existing work embeds convolutional layers within hybrid or ensemble networks [8], making it difficult to measure the standalone contribution of the architecture. By benchmarking the network against both linear models and strong tabular ensembles under a uniform MCC evaluation framework [15], [16], the evaluation assesses whether deep feature learning provides a measurable advantage over traditional classifiers. This framework bypasses complex feature selection strategies [3] by pairing stratified cross-validation directly with isolated training-fold SMOTE handling. To ensure reproducibility, the established architecture is evaluated under strict conditions that reflect the operational challenges of detecting rare fraud cases.

2. LITERATURE REVIEW

2.1 Fraud Detection in Electronic Banking

The expansion of digital payment systems has increased exposure to financial fraud, creating greater demand for automated detection methods capable of adapting to evolving attack patterns [19]. Banking systems remain particularly vulnerable. Fraud affects multiple sectors, including healthcare, insurance, securities, and money laundering, but card-related and digital transaction fraud continue to dominate electronic banking environments. Phishing attacks and related cyber threats also remain common across online banking platforms [20].

Real-time phishing detection systems also target online fraud alongside transaction-level monitoring. These frameworks integrate information retrieval methods, natural language processing (NLP), and machine learning techniques to detect and prevent fraudulent online activity [21].

Current research frequently examines the broader state of fraud detection methodologies [6]. In AI-based credit card applications specifically, investigations highlight class imbalance, interpretability, and scalability as recurring challenges [22]. Hybrid approaches combining machine learning and deep learning adapt better to changing fraud patterns. Data imbalance, however, remains a persistent hurdle within these setups [8].

A review of 104 fraud detection studies shows that most papers focus on external fraud cases such as credit card and insurance fraud [6]. Internal problems such as embezzlement, financial statement manipulation, and money laundering receive far less attention. The review also notes that deep learning remains less common in this area, since many studies still rely on traditional machine learning methods.

Earlier reviews show that data mining methods dominated fraud detection research for many years, mainly through supervised approaches such as logistic regression, neural networks, and decision trees [23]. Recent research analysing credit card, insurance, and financial reporting fraud indicates that deep learning models including CNNs, LSTMs, transformers, and ensemble methods surpass conventional methods, despite ongoing challenges related to interpretability, data imbalance, and regulatory compliance [24].

Ensemble learning and meta-heuristic optimization are increasingly used to improve fraud detection accuracy, although scalability, computational cost, and data imbalance remain persistent concerns [22]. Most investigations still rely on fixed, unbalanced datasets and supervised models, meaning that results obtained under experimental settings do not always adapt well to real-world situations [6], [8].

Table 1 summarises representative studies in credit card fraud detection, highlighting their datasets, modelling approaches, primary contributions, and key limitations. The comparison provides the basis for identifying the research gap addressed in this study.

Table 1. Comparison of Related Fraud Detection Studies

Study	Dataset	Model	Key Focus	Limitation
Zhang et al. (2018)	Commercial Bank Transaction Data	CNN with Feature Sequencing Layer	Automated spatial feature extraction for fraud detection without handcrafted features	Private dataset limits reproducibility and independent validation
Nobel et al. (2024)	PaySim Synthetic Mobile Money Dataset (Kaggle)	SVM, XGBoost, Decision Tree, Logistic Regression + SMOTE + XAI (SHAP, LIME)	Integrating ML with explainable AI (SHAP and LIME) to improve model transparency and interpretability in imbalanced fraud detection	Evaluation protocol does not explicitly discuss leakage-free resampling procedures; MCC was not reported

Khalid et al. (2024)	Kaggle Credit Card Fraud Detection Dataset	Ensemble ML (SVM, KNN, RF, Bagging, Boosting, Voting Ensemble)	Ensemble methods to address class imbalance in fraud detection	Evaluation relies on a single train-test split without cross-validation; leakage-free resampling procedures not explicitly discussed; MCC was not reported
Sulaiman et al. (2024)	European Credit Card Fraud Dataset	Autoencoder, CNN, and LSTM with Hyperparameter Tuning	Hyperparameter optimisation of deep learning models and sampling strategies for fraud detection	Evaluation focused primarily on Accuracy, Detection Rate, and AUC; MCC was not reported, and leakage-free resampling procedures were not explicitly discussed despite the use of stratified cross-validation
Akour et al. (2025)	Kaggle Credit Card Fraud Detection Dataset	CNN-LSTM-Attention Hybrid	Joint spatial and temporal fraud pattern extraction using deep learning	High architectural complexity increases computational requirements
Btoush et al. (2025)	Kaggle Credit Card Fraud Detection Dataset	Hybrid ML+DL Stacking Ensemble (DT, RF, SVM, XGBoost, CatBoost, LR, CNN-BiLSTM-Attention)	Maximising fraud detection performance through hybrid ML+DL ensemble learning	High model complexity may reduce interpretability and deployment efficiency
Proposed Study	Kaggle Credit Card Fraud Detection Dataset	One-Dimensional CNN	Leakage-free stratified 5-fold cross-validation with MCC as the primary evaluation metric under extreme class imbalance	Limited to a single public dataset; generalisability to other fraud domains has not yet been validated

|

2.2 Research Gaps

As shown in Table 1, recent fraud detection studies have achieved substantial improvements through machine learning, deep learning, explainable AI, and ensemble learning techniques. However, several methodological limitations remain unresolved. Fraud detection research has advanced considerably, although important challenges persist [24]. Early systems relied mainly on rigid rule-based logic that struggled to adapt to evolving cyberattacks [25]. Machine learning models such as Random Forest and SVM improved detection performance, but centralized data processing introduced privacy concerns and reduced adaptability as fraud patterns evolved [25]. Class imbalance, overfitting, and evolving fraud patterns remain persistent problems for many machine learning models [4], [24]. Models that depend on manually engineered features also struggle to distinguish subtle differences between fraudulent and legitimate transactions [8], [4]. Random Forest and Gradient Boosting remain strong benchmark models for tabular fraud detection, but comparisons with CNN-based approaches are not always evaluated under the same experimental conditions[6]. Furthermore, many studies do not explicitly document whether oversampling techniques are applied before or after data partitioning. This lack of methodological transparency creates a risk of information leakage, potentially inflating reported performance and making fair comparison across studies difficult. As highlighted in Table 1, leakage-free resampling procedures are rarely discussed explicitly despite the widespread use of SMOTE and related balancing techniques.

Many existing studies still rely on hybrid or ensemble architectures instead of testing standalone CNN models directly [8]. Deep learning models have improved performance across fraud and anomaly detection tasks [26], [14]. Hybrid architectures such as CNN–LSTM–Attention report strong precision and F1-score results when combined with optimization strategies, and ablation studies suggest that these improvements arise from the interaction between

convolutional, temporal, and attention mechanisms [14]. Automatic feature extraction remains one of the main advantages of deep learning models [26], [14]. Its effect on reducing manual feature engineering under controlled and reproducible experimental conditions remains uncertain [24].

Another limitation concerns model evaluation. Many studies continue to emphasize overall accuracy as a primary performance metric, even though accuracy can provide misleading conclusions when fraudulent transactions constitute only a small fraction of the dataset [4]. Metrics such as the Matthews Correlation Coefficient (MCC), which incorporate all elements of the confusion matrix and are particularly suitable for highly imbalanced classification problems, are comparatively underreported in the fraud detection literature[15], [16].

Earlier work such as Zhang et al. [10] applied CNNs to fraud detection using private transaction datasets, limiting reproducibility and independent validation. This study addresses these gaps by evaluating a standalone one-dimensional CNN using the publicly available Kaggle Credit Card Fraud Detection dataset under a leakage-free experimental framework. The model is compared against Logistic Regression, Decision Tree, Random Forest, and Gradient Boosting using identical preprocessing, resampling, and validation procedures. SMOTE is restricted strictly to the training folds of stratified 5-fold cross-validation to eliminate information leakage, while MCC is treated as the primary evaluation metric. This framework enables a transparent and reproducible assessment of CNN performance against both conventional and ensemble machine learning models under extreme class imbalance conditions.

2.3 Rule-Based Approaches

Early banking fraud detection systems relied mainly on fixed rule-based mechanisms that flagged transactions using predefined thresholds linked to devices, merchants, or transaction values. One major advantage of these systems was interpretability, since each decision could be traced directly to a specific rule [27]. Their limitations became more obvious as fraud patterns evolved. Static rules adapted poorly to changing attack behaviour, often resulting in higher false alarm rates and slower response times [24]. These systems also depended entirely on predefined fraud signatures, making it difficult to detect previously unseen attack patterns [28]. Once attackers understood how the rules operated, they could often bypass them by modifying transaction behaviour [27]. Researchers later explored combining expert knowledge with boosting-based machine learning models to improve detection accuracy while maintaining interpretability [29].

Hybrid approaches have combined rule-based systems with neural networks to improve adaptability in fraud detection models [30]. Other methods generate rules automatically by integrating expert-defined conditions with tree-based machine learning models [31].

These approaches reflect the gradual movement away from rigid manual rules toward more flexible and data-driven fraud detection systems [32].

2.4 Machine Learning (ML) Approaches

Machine learning models are widely used in fraud detection because they can adapt to transaction patterns as more data becomes available [22]. Common approaches include Support Vector Machines (SVMs), Artificial Neural Networks (ANNs), Hidden Markov Models, KNN, Decision Trees, XGBoost, and Gradient Boosting [33], [7]. Many of these models perform well under stable training conditions, but performance can decline when fraud behavior changes over time, largely because they rely on manually engineered features [22]. Transaction environments are also affected by concept drift and feature variation, both of which can gradually reduce model reliability [34].

Genetic Algorithms have been combined with information-theoretic metrics such as the Data Value Metric to improve feature selection in fraud detection models [35]. Ensemble feature selection methods used alongside Random Forest have also reported strong classification performance [36]. In addition, combining SMOTE with Random Forest models improves performance when handling imbalanced fraud datasets [37].

Random Forest frameworks combined with data balancing strategies have been used to detect suspicious activity in cybersecurity datasets [38]. Outlier detection methods have also been applied in financial fraud datasets. Gaussian Mixture Models (GMM) combined with ensemble methods such as XGBoost improve prediction accuracy by removing noisy and extreme data points [39].

Fraud datasets are typically highly imbalanced. This imbalance often causes classifiers to favor the majority class over rare fraudulent transactions [4]. This imbalance affects model performance, although oversampling techniques such as SMOTE have been used to improve detection accuracy [13]. Poor data balancing and weak feature selection can also reduce model reliability. Careful preprocessing, balanced training data, and automated feature extraction improve model consistency across fraud detection tasks [7].

2.5 Handling Class Imbalance

Techniques such as SMOTE and GAN-based augmentation are commonly used to address class imbalance by generating synthetic fraud samples [4]. Other approaches, including Random Undersampling (RUS) and Random Oversampling (ROS), have also been applied to imbalanced fraud datasets [5].

Resampling techniques such as SMOTE, RUS, and SMOTEENN have been combined with tree-based models to improve performance on imbalanced datasets [40]. SMOTE-based balancing has also been applied within ensemble frameworks, particularly alongside Random Forest and boosting methods, where improvements in fraud detection performance have been reported [41].

Approaches such as SMOTified-GAN and GANified-SMOTE generate synthetic fraud samples based on learned patterns [4]. The BGVOA-LS model has also been introduced to reduce input dimensionality while improving training efficiency and maintaining stable performance across datasets [7]. However, many of these techniques are still evaluated within single model settings. Direct comparisons with deep learning and ensemble models under consistent experimental conditions remain limited.

2.6 Deep Learning (DL) Approaches

Deep learning models such as RNNs, Autoencoders, and CNNs are widely used in financial fraud detection because they can identify complex nonlinear patterns in transaction data [8], [14]. These models also reduce the need for manual feature engineering by learning useful feature representations automatically. In addition, they can adapt to changes in fraud patterns as transaction data evolves over time.[8], [14].

CNNs and hybrid neural network architectures are used in fraud detection for feature learning and to handle datasets where fraud cases are rare [14].

Hybrid deep learning models incorporating memetic and modular learning strategies have been applied to phishing and spam threat detection through message-content analysis [42]. These models continue to perform effectively as attack patterns evolve [42]. In anomaly-based intrusion detection systems, Genetic Algorithms have also been combined with neural networks to improve detection performance [43].

Deep learning ensemble frameworks are used to improve fraud detection performance while reducing overfitting and other data-related issues [44]. Hybrid frameworks combining machine learning, deep learning, and ensemble methods have been applied in cybersecurity to address poor generalization and class imbalance [45].

Deep learning models are computationally intensive and often require large amounts of training data [22]. Training can also be slow when computing resources are limited [22]. Many models provide limited explainability and lack privacy-preserving mechanisms, creating challenges for regulatory compliance and transparency in financial applications [46].

Federated learning models emphasize both explainability and privacy, allowing institutions to collaborate on fraud detection without sharing sensitive data [46].

2.7 Convolutional Neural Networks (CNNs) for Fraud Detection

CNNs are used in fraud detection to learn complex patterns from structured transaction data with minimal manual feature engineering [14], [8]. Fraud patterns change over time, requiring models to adapt without frequent redesign of input features. A hybrid CNN–LSTM–Attention model captures spatial and temporal patterns in transaction data [14]. Higher precision and F1-scores have been reported for this architecture compared with traditional machine learning models [14]. Zhang [10] reported that CNNs trained directly on transaction data outperformed conventional models while requiring little feature engineering. Many studies combine multiple architectures, and evaluation practices remain inconsistent, particularly in the treatment of imbalanced datasets and the reporting of performance metrics [24].

Direct comparisons between a standalone CNN and conventional machine learning models on a common publicly available dataset are limited [6], [8].

Few studies evaluate CNN performance alongside ensemble methods such as Random Forest and Gradient Boosting under consistent experimental conditions.

3. PROPOSED METHOD

A leakage-free experimental pipeline was used to evaluate a one-dimensional CNN for fraud detection in data where fraudulent transactions are extremely rare. The methodology consisted of four stages: data preparation, class imbalance handling, model development, and performance evaluation.

3.1 Data Acquisition and Pre-processing

The Kaggle Credit Card Fraud Detection dataset was used in this study and contains 284,807 anonymized real-world transactions [17]**.** Input features consisted of the V1–V28 principal components together with the Time and Amount variables. No manual feature engineering was performed. All features were normalized using StandardScaler.

The dataset was divided into training and test subsets using an 80:20 ratio. SMOTE was applied only to the training data after the split. The test set was excluded from the resampling process to prevent data leakage and preserve an independent evaluation dataset [47], [48].

3.2 Addressing Severe Class Imbalance

In this dataset, fraudulent transactions make up just 0.172% of all records, meaning the model faces a severe imbalance between the two classes from the start [47]. This imbalance biases models toward the majority class and can reduce fraud detection performance.

SMOTE was used to generate additional fraudulent samples until the two classes were more evenly represented in the training data [47]. Accuracy alone can be misleading when class distributions are highly imbalanced. Metrics such as TPR, FPR, and ROC-AUC provide a clearer view of how well the model detects fraudulent transactions, especially when fraud cases are rare[48].

3.3 Models Implemented and Configuration

Four models were used for comparison: Logistic Regression with L2 regularization and Decision Tree on the conventional side [49], and Random Forest and Gradient Boosting as ensemble methods. Random Forest and Gradient Boosting were included as strong benchmark models for tabular financial data. The CNN was the main model under study because it can learn complex patterns directly from transaction data [8], [9].

SMOTE was applied during training to correct the class distribution. Each transaction was represented as a sequence of 30 features and reshaped into a one-dimensional vector so it could be processed by a Conv1D layer.

The model begins with a Conv1D layer using 32 filters and a kernel size of 3, which helps detect patterns among neighbouring PCA features and capture non-linear relationships that simpler models may miss. This is followed by a MaxPool1D layer (pool size = 2), which reduces sensitivity to the exact position of patterns in the input. The convolutional and pooling layers feed into dense layers, with dropout applied to reduce overfitting [14]. Two dropout layers were used, with rates of 0.2 and 0.3, to reduce the chances of the model fitting too closely to the synthetic samples generated by SMOTE.

The complete architecture, including layer dimensions and activation functions, is shown in Table 1. Training settings, including the Adam optimizer (learning rate = 0.001) and Binary Cross-Entropy loss, are summarized in Table 2. The final layer is a Sigmoid output for binary classification[9]. All models were trained and evaluated using identical preprocessing and resampling procedures to maintain consistent experimental conditions.

Table 2. Layer-wise Architectural Specifications

Layer (Type)	Kernel/Neurons	Activation	Regularization	Output Shape
Input	-	-	-	(30, 1)
Conv1D	32 Filters (k=3)	ReLU	-	(28, 32)
MaxPool1D	Pool Size=2	-	-	(14, 32)

Dropout 1	-	-	Rate = 0.2	(14, 32)
Flatten	-	-	-	(448)
Dense	64 Neurons	ReLU	-	(64)
Dropout 2	-	-	Rate = 0.3	(64)
Output	1 Neuron	Sigmoid	-	(1)

Table 3. Training and Optimization Configuration

Parameter	Value	Parameter	Value
Optimizer	Adam	Batch Size	128
Learning Rate	0.001	Epochs	10
Loss Function	Binary Cross-entropy	Validation	Stratified 5-Fold Cross-Validation
Sampling	SMOTE (applied to training data)	Scaling	Z-Score (StandardScaler)

3.4 Experimental Procedure and Evaluation

Evaluation was performed using stratified 5-fold cross-validation, with the proportion of fraud cases maintained across all folds. Within each fold, SMOTE was applied only to the training portion. The validation data was kept separate to prevent data leakage. The model was then trained on the resampled data and tested on the original validation fold, which contained no synthetic samples.

Performance metrics were computed from out-of-fold predictions. Under this approach, each sample was evaluated by a model that had not seen it during training. The reported metrics therefore reflect performance on unseen validation data, with evaluation conducted on the original class distribution rather than on synthetic SMOTE samples.

All models underwent identical cross-validation, preprocessing, and resampling procedures to maintain consistent experimental conditions. The CNN was trained for 10 epochs per fold using a batch size of 128. Performance was measured using Precision, Recall, F1-score, ROC-AUC, and MCC. MCC was treated as the primary metric because it accounts for all four components of the confusion matrix, making it more reliable when fraud cases are rare [50].

Results can still be affected by the static nature of the dataset and the use of synthetic samples generated through SMOTE. This limitation is discussed further in Section 4.

3.5 Evaluation Metrics for Imbalanced Data

Accuracy is one of the most commonly reported metrics in classification tasks, but it can be misleading when the data is heavily skewed toward one class [50]. Precision, Recall, ROC-AUC, and MCC therefore receive greater emphasis than accuracy.

Even F1-score, which balances Precision and Recall, can be insufficient when used alone.

MCC was chosen as the main evaluation metric because it remains reliable even when the classes are highly imbalanced. MCC combines all four confusion matrix components, TP, TN, FP, and FN, into a single value, giving a more complete view of model performance[50]. As shown in Eq. (1), MCC incorporates all outcomes of the confusion matrix into a single metric. The formula is defined as:

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (1)$$

ROC-AUC captures the relationship between TPR and FPR across different decision thresholds and evaluates the separation between fraudulent and legitimate transactions. Precision, Recall, F1-score, ROC-AUC, and MCC provide a more complete view of model behaviour than any single metric, especially when fraud cases are rare in this dataset.

4. RESULTS AND DISCUSSION

The CNN was evaluated alongside Logistic Regression, Decision Tree, Random Forest, and Gradient Boosting using the same experimental setup. Table 4 shows that the CNN achieved the highest MCC, with Random Forest producing the closest competing result. Because MCC served as the primary evaluation metric under class imbalance, the CNN provided the strongest overall balance between fraud detection and classification performance.

Logistic Regression recorded a high ROC-AUC, but its MCC and Precision were substantially lower than those of the CNN and Random Forest. The model separated the classes effectively but was less reliable in identifying fraudulent transactions. Under severe class imbalance, ROC-AUC alone may not adequately reflect fraud detection performance.

Random Forest produced strong results across the evaluation metrics, whereas Gradient Boosting achieved more moderate performance. The CNN recorded the lowest Log Loss among the evaluated models, reflecting more stable probability estimates.

Across the evaluated models, the CNN maintained the strongest balance between minority-class detection and overall classification performance. Random Forest remained a strong alternative for tabular fraud detection. These findings were obtained using a static fraud dataset and SMOTE-based class balancing during training and should be interpreted within that experimental setting.

Table 4. Performance Comparison of CNN and Benchmark Machine Learning Models

Model	MCC	(CV) Accuracy	Precision	Recall	F1-Score	ROC-AUC	Log Loss
CNN	0.7081	0.9988 ± 0.0003	0.6000	0.8300	0.7000	0.9659	0.0104
LR	0.2288	0.9748 ± 0.0014	0.0591	0.9126	0.1110	0.9774	0.1125
DT	0.3082	0.9880 ± 0.0009	0.1120	0.8618	0.1983	0.9037	0.0570
RF	0.6534	0.9983 ± 0.0002	0.4988	0.8577	0.6308	0.9825	0.0445
GB	0.3100	0.9871 ± 0.0015	0.1087	0.8984	0.1939	0.9772	0.0592

4.1 Performance Analysis

The proposed CNN was evaluated against Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), and Gradient Boosting (GB) under the same leakage-free experimental framework. As presented in Table 4, the CNN achieved the highest Matthews Correlation Coefficient (MCC) of 0.7081, followed by Random Forest at 0.6534. The higher MCC indicates stronger overall agreement between predicted and actual class assignments under the severe class imbalance present in the dataset. Because MCC incorporates all four components of the confusion matrix, this result indicates that the CNN maintained a better overall balance between fraudulent and legitimate transaction classifications than the benchmark models. Figure 1 further illustrates the comparative MCC performance of the five evaluated models, showing the higher MCC achieved by the CNN relative to the benchmark models.

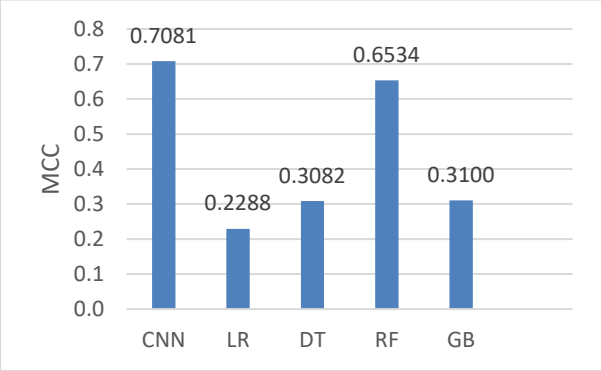

Figure 1. Comparative MCC performance of the evaluated models

The CNN also achieved the highest F1-score of 0.7000 and the highest Precision of 0.6000. Random Forest followed with an F1-score of 0.6308, while Decision Tree, Gradient Boosting, and Logistic Regression recorded substantially lower F1-scores. The higher F1-score indicates a more favourable balance between Precision and Recall. The CNN's relatively high Precision is particularly important because fraud detection involves identifying fraudulent transactions while limiting incorrect fraud classifications.

The Recall results provide an important contrast. Logistic Regression achieved the highest Recall (0.9126), followed by Gradient Boosting (0.8984), Decision Tree (0.8618), Random Forest (0.8577), and the CNN (0.8300). Thus, the CNN did not detect the largest proportion of fraudulent transactions when Recall was considered alone. However, Logistic Regression's high Recall was accompanied by a very low Precision of 0.0591, resulting in an F1-score of 0.1110 and MCC of 0.2288. This contrast demonstrates why Recall alone is insufficient for evaluating fraud detection under severe class imbalance. The CNN's comparatively lower Recall was accompanied by substantially stronger Precision, resulting in a more favourable overall balance between fraud detection and incorrect positive classifications.

The ROC-AUC results provide another perspective on the comparative performance. Random Forest recorded the highest ROC-AUC at 0.9825, followed by Logistic Regression (0.9774), Gradient Boosting (0.9772), the CNN (0.9659), and Decision Tree (0.9037). Although the CNN did not achieve the highest ROC-AUC, it achieved the highest MCC and F1-score. The Random Forest result is particularly important because its ROC-AUC was higher than that of the CNN, while its MCC and F1-score were lower. This demonstrates that strong class-separation capability measured by ROC-AUC does not necessarily correspond to the strongest minority-class classification performance under severe class imbalance. Therefore, ROC-AUC alone would provide an incomplete basis for determining the strongest fraud detection model in this experiment. **Figure** 2 provides a visual comparison of Precision, Recall, F1-score, and ROC-AUC across the five models. The figure highlights the trade-off between Recall and Precision and shows that the CNN achieved a substantially higher Precision and F1-score than the models with higher Recall, while Random Forest achieved the highest ROC-AUC.

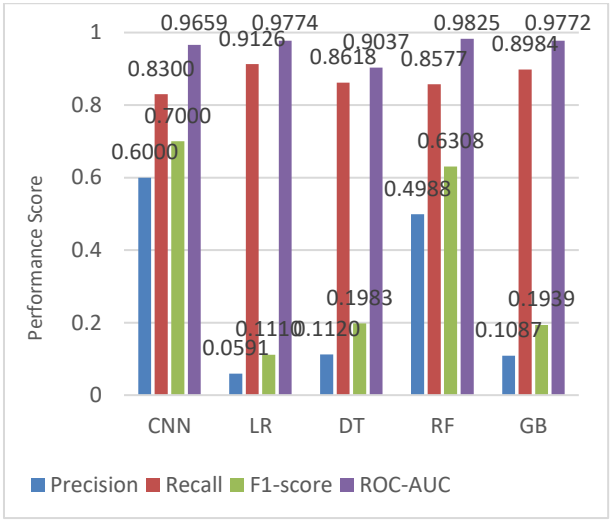

Figure 2. Comparative performance of the evaluated models based on Precision, Recall, F1-score, and ROC-AUC.

The ROC curve of the CNN is presented in Figure 3, with an AUC of 0.9659.

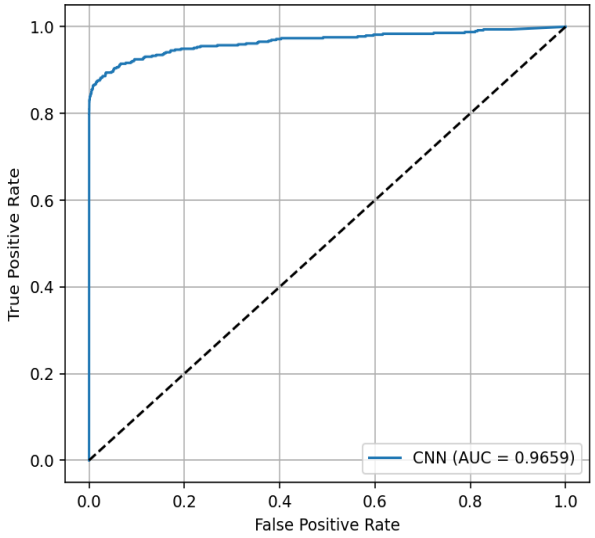

Figure 3. ROC curve of the final CNN model (AUC = 0.9659).

The Log Loss results provide additional information about the models' probability estimates. The CNN recorded the lowest Log Loss (0.0104), compared with 0.0445 for Random Forest, 0.0570 for Decision Tree, 0.0592 for Gradient Boosting, and 0.1125 for Logistic Regression. The lower Log Loss indicates that the CNN's predicted probabilities were more closely

aligned with the observed class outcomes under the evaluation framework. This provides an additional distinction between the models beyond their binary classification performance and is relevant where predicted probabilities may inform decisions about whether transactions require further investigation.

The cross-validation results indicate stable performance across the five folds. The CNN achieved a cross-validation accuracy of 0.9988 ± 0.0003, compared with 0.9983 ± 0.0002 for Random Forest, 0.9748 ± 0.0014 for Logistic Regression, 0.9880 ± 0.0009 for Decision Tree, and 0.9871 ± 0.0015 for Gradient Boosting. The small variation observed for the CNN indicates that its accuracy remained consistent across the validation folds rather than depending heavily on a particular data partition. However, because accuracy can remain high when the majority class dominates, this result should be interpreted alongside the imbalance-aware metrics rather than in isolation. Figure 4 presents the confusion matrix of the CNN model, providing a direct view of the distribution of correct and incorrect predictions for fraudulent and legitimate transactions. The confusion matrix complements the aggregate evaluation metrics by showing the underlying classification outcomes from which measures such as Precision, Recall, and MCC are derived.

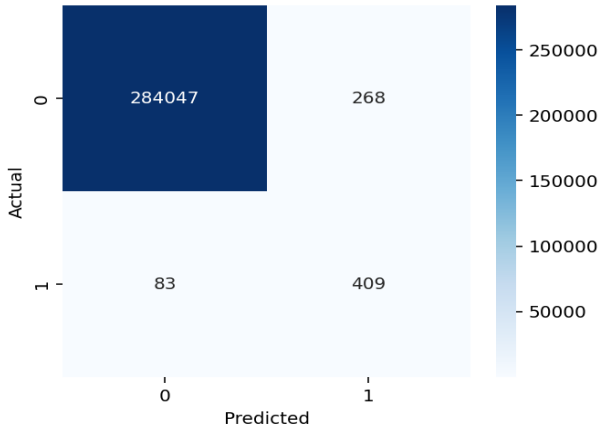

Figure 4. Confusion matrix of the CNN model.

Random Forest demonstrated strong performance across several evaluation measures, achieving a high MCC and F1-score, whereas Gradient Boosting produced substantially lower values despite its high Recall. This difference indicates that the performance of ensemble methods varied considerably under the same experimental conditions, and the results do not support a general conclusion that all ensemble approaches performed strongly.

Overall, the CNN demonstrated the strongest overall performance under the evaluation framework, based particularly on its leading MCC, F1-score, Precision, and Log Loss results. Random Forest remained a strong alternative, particularly in terms of ROC-AUC, while Logistic Regression achieved the highest Recall. The findings therefore indicate that model performance under extreme class imbalance depends on the evaluation criterion used and that MCC provides a more balanced basis for comparative assessment than accuracy or ROC-AUC alone.

Previous studies have reported mixed outcomes for deep learning models applied to tabular fraud detection tasks [11]. The present results contribute to this literature by demonstrating that a standalone one-dimensional CNN can achieve competitive performance on structured transaction data without hybrid architectural components or additional manual feature engineering. Under the same experimental conditions, the CNN achieved the strongest overall balance across the principal imbalance-aware metrics, supporting its feasibility for fraud detection under severe class imbalance.

4.2 Discussion of Findings

Convolutional networks can be applied to fraud detection on structured data beyond traditional image-based tasks. The CNN learns feature interactions through convolution without manual feature design. The superior MCC achieved by the model suggests that local interactions among PCA components contribute meaningfully to fraud detection despite the absence of inherent spatial structure within tabular data. A one-dimensional CNN can perform effectively on structured transaction records.

Random Forest also produced strong classification results, confirming the continued effectiveness of ensemble-based approaches for tabular financial data. The CNN recorded a lower Log Loss, reflecting more stable probability estimates across the evaluated transactions. In financial environments, probability estimates often influence downstream decision-making processes in addition to classification outcomes.

Several limitations affect the interpretation of these results. SMOTE generates synthetic fraud samples that may not fully represent real fraud behaviour. The dataset is fully anonymised, preventing access to potentially informative attributes such as merchant type, device information, and geographic location. The evaluation was also conducted on a static dataset, leaving the effects of concept drift and evolving fraud patterns unresolved.

SMOTE introduces a trade-off between Recall and Precision. Improved fraud detection rates can be accompanied by an increase in false-positive classifications, affecting operational efficiency and user trust. The findings reported here were obtained under controlled experimental conditions, and the combination of SMOTE with a static dataset may influence real-world generalisation.

Interpretability remains an important consideration in financial systems and regulatory environments. This study focused on model performance and architectural evaluation, and no explainability techniques were applied to individual predictions. Methods such as SHAP could be incorporated in future investigations to improve model transparency.

Future applications can extend this work to streaming transaction environments where fraud patterns evolve over time. Additional studies may examine architectural variations and hyperparameter configurations that improve MCC and Recall under class imbalance. Operational considerations, including latency, scalability, and deployment constraints, remain important areas for future evaluation.

The results provide a reproducible benchmark under extreme class imbalance and demonstrate that a one-dimensional CNN can remain competitive with traditional machine learning approaches without manual feature engineering. Fraud detection under severe class imbalance requires attention to both classification performance and probability reliability rather than reliance on a single evaluation metric.

4.3 Practical Implications

The proposed framework provides researchers and practitioners with a reproducible methodology for evaluating fraud detection models under severe class imbalance. By restricting SMOTE to the training folds of stratified cross-validation, the framework minimizes the risk of information

leakage and promotes fair model comparison. The emphasis on imbalance-aware metrics, particularly the Matthews Correlation Coefficient (MCC), supports more reliable assessment of fraud detection systems than accuracy-based evaluation alone. For financial institutions and researchers developing fraud detection solutions, the framework offers a transparent benchmark for evaluating model performance under conditions that more closely reflect real-world fraud detection challenges.

5. COMPARISON

Previous studies report mixed results for deep learning on tabular fraud data [11], while other studies have shown that CNN-based models can perform well in fraud detection [10]. In this study, the one-dimensional CNN achieved strong performance under extreme class imbalance, particularly in terms of MCC.

Unlike Zhang et al. [10] who evaluated a CNN model using transaction data from a commercial bank, this study uses the Kaggle Credit Card Fraud Detection dataset [17] within a controlled and leakage-free evaluation framework. The use of a public dataset supports reproducibility and facilitates comparison with future studies.

Many existing approaches rely on hybrid models, ensemble techniques, or extensive feature engineering. Unlike hybrid architectures such as CNN–LSTM–Attention [14], which combine spatial and temporal feature learning, the proposed model operates on static transaction vectors without sequence modelling. The findings suggest that convolution alone can capture meaningful fraud patterns from structured transaction data without requiring more complex hybrid architectures.

Although Random Forest achieved a slightly higher ROC-AUC, the CNN produced the highest MCC. Ensemble models remain strong performers for tabular financial data [11]. Under the same experimental conditions, the CNN provided better minority-class discrimination without manual feature engineering.

The model achieved strong results across imbalance-aware metrics under controlled experimental conditions using the Kaggle Credit Card Fraud Detection dataset [17]. MCC served as the primary evaluation metric because of its suitability for assessing performance under severe class imbalance [15], [16]. The study establishes a reproducible benchmark for future comparisons using the same dataset and evaluation framework.

6. CONCLUSIONS

This study evaluated a one-dimensional CNN for fraud detection under extreme class imbalance using the Kaggle Credit Card Fraud Detection dataset [17]. The model achieved an MCC of 0.7081, a Recall of 0.83, and a ROC-AUC of 0.9659, demonstrating effective detection of fraudulent transactions.

The model learned useful patterns directly from transaction data without manual feature engineering. Compared with strong ensemble models, the CNN provided a more balanced fraud detection profile while maintaining stable performance under controlled experimental conditions.

The study is limited by the use of a static dataset and synthetic oversampling through SMOTE, which may not fully represent real-world fraud behaviour. Future evaluations can examine performance using streaming transaction data and additional contextual features to assess deployment under evolving fraud conditions.

The findings support the use of a properly evaluated one-dimensional CNN as a competitive alternative to strong ensemble methods without requiring hybrid architectures or complex feature engineering. More importantly, the principal contribution of this study is the establishment of a controlled leakage-free evaluation framework that enables reliable comparison of fraud detection models under severe class imbalance. By restricting SMOTE to the training folds of stratified cross-validation and emphasizing imbalance-aware metrics such as MCC, the framework provides a transparent and reproducible benchmark for future fraud detection research using publicly available datasets.

Author Contributions: Conceptualisation: K.O.O. and M.I.A.; Methodology: K.O.O.; Software: K.O.O.; Validation: K.O.O., M.I.A., and A.A.O.; Formal analysis: K.O.O.; Investigation: K.O.O.; Resources: M.I.A. and A.A.O.; Data curation: K.O.O.; Writing—original draft preparation: K.O.O.; Writing—review and editing: M.I.A. and A.A.O.; Visualisation: K.O.O.; Supervision: M.I.A. and A.A.O.; Project administration: M.I.A.; Funding acquisition: None. All authors have read and approved the final version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The dataset used in this study is publicly available on Kaggle: Credit Card Fraud Detection Dataset (https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud). No new data were created.

7. ACKNOWLEDGMENTS

The authors would like to thank the Department of Computer Science at Delta State University, Abraka, for their support and guidance. The authors also acknowledge the contributors of the Kaggle Credit Card Fraud Detection dataset used in this study. AI-assisted tools were used for language refinement and manuscript polishing. All scientific content, analysis, and conclusions were developed and verified by the authors.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analysis, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

8. REFERENCES

[1] Merchant Savvy, "Payment Fraud Statistics, Trends & Forecasts (2024)." 2024. Accessed: Nov. 29, 2025. [Online]. Available: :https://www.merchantsavvy.co.uk/payment-fraud-statistics/

[2] Nasdaq, "Nasdaq Verafin 2024 Global Financial Crime Report." Accessed: Nov. 29, 2025. [Online]. Available: : https://www.nasdaq.com/global-financial-crime-report

[3] E. Ileberi, Y. Sun, and Z. Wang, "A machine learning based credit card fraud detection using the GA algorithm for feature selection," *J Big Data*, vol. 9, no. 1, p. 24, Dec. 2022, doi: 10.1186/s40537-022-00573-8.

[4] P. C. Y. Cheah, Y. Yang, and B. G. Lee, "Enhancing Financial Fraud Detection through Addressing Class Imbalance Using Hybrid SMOTE-GAN Techniques," *IJFS*, vol. 11, no. 3, p. 110, Sep. 2023, doi: 10.3390/ijfs11030110.

[5] J. L. Leevy, J. M. Johnson, J. Hancock, and T. M. Khoshgoftaar, "Threshold optimization and random undersampling for imbalanced credit card data," *J Big*

Data, vol. 10, no. 1, p. 58, May 2023, doi: 10.1186/s40537-023-00738-z.

[6] L. Hernandez Aros, L. X. Bustamante Molano, F. Gutierrez-Portela, J. J. Moreno Hernandez, and M. S. Rodríguez Barrero, "Financial fraud detection through the application of machine learning techniques: a literature review," *Humanit Soc Sci Commun*, vol. 11, no. 1, p. 1130, Sep. 2024, doi: 10.1057/s41599-024-03606-0.

[7] A. H. Salem, S. M. Azzam, O. E. Emam, and Abohany, "From chaos to clarity: unraveling credit card fraud with BGVOA-LS," *J Big Data*, vol. 12, no. 1, p. 215, Sep. 2025, doi: 10.1186/s40537-025-01274-8.

[8] E. Btoush, X. Zhou, R. Gururajan, K. C. Chan, and O. Alsodi, "Achieving Excellence in Cyber Fraud Detection: A Hybrid ML+DL Ensemble Approach for Credit Cards," *Applied Sciences*, vol. 15, no. 3, p. 1081, Jan. 2025, doi: 10.3390/app15031081.

[9] S. S. Sulaiman, I. Nadher, and S. M. Hameed, "Credit Card Fraud Detection Using Improved Deep Learning Models," *CMC*, vol. 78, no. 1, pp. 1049–1069, 2024, doi: 10.32604/cmc.2023.046051.

[10] Z. Zhang, X. Zhou, X. Zhang, L. Wang, and P. Wang, "A Model Based on Convolutional Neural Network for Online Transaction Fraud Detection," *Security and Communication Networks*, vol. 2018, pp. 1–9, Aug. 2018, doi: 10.1155/2018/5680264.

[11] V. Borisov, T. Leemann, K. Seßler, J. Haug, M. Pawelczyk, and G. Kasneci, "Deep Neural Networks and Tabular Data: A Survey," *IEEE Trans. Neural Netw. Learning Syst.*, vol. 35, no. 6, pp. 7499–7519, Jun. 2024, doi: 10.1109/TNNLS.2022.3229161.

[12] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[13] S. M. N. Nobel *et al.*, "Unmasking Banking Fraud: Unleashing the Power of Machine Learning and Explainable AI (XAI) on Imbalanced Data," *Information*, vol. 15, no. 6, p. 298, May 2024, doi: 10.3390/info15060298.

[14] I. Akour, N. Mohamed, and S. Salloum, "Hybrid CNN-LSTM With Attention Mechanism for Robust Credit Card Fraud Detection," *IEEE Access*, vol. 13, pp. 114056–114068, 2025, doi: 10.1109/ACCESS.2025.3583253.

[15] S. Boughorbel, F. Jarray, and M. El-Anbari, "Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric," *PLoS ONE*, vol. 12, no. 6, p. e0177678, Jun. 2017, doi: 10.1371/journal.pone.0177678.

[16] Q. Zhu, "On the performance of Matthews correlation coefficient (MCC) for imbalanced dataset," *Pattern Recognition Letters*, vol. 136, pp. 71–80, Aug. 2020, doi: 10.1016/j.patrec.2020.03.030.

[17] Worldline and Machine Learning Group (ULB), "Credit Card Fraud Detection Dataset." Kaggle. Accessed: Nov. 30, 2025. [Online]. Available: https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud

[18] H. Thimonier, F. Popineau, A. Rimmel, B.-L. Doan, and F. Daniel, "Comparative Evaluation of Anomaly Detection Methods for Fraud Detection in Online Credit Card Payments," in *Proceedings of Ninth International Congress on Information and Communication Technology*, vol. 1011, X.-S. Yang, S. Sherratt, N. Dey, and A. Joshi, Eds., in Lecture Notes in Networks and Systems, vol. 1011. , Singapore: Springer Nature Singapore, 2024, pp. 37–50. doi: 10.1007/978-981-97-4581-4_4.

[19] M. N. Alatawi, "Detection of fraud in IoT based credit card collected dataset using machine learning," *Machine Learning with Applications*, vol. 19, p. 100603, Mar. 2025, doi: 10.1016/j.mlwa.2024.100603.

[20] M. Ifeanyi Akazue, A. Adimabua Ojugo, R. Elizabeth Yoro, B. Ogheneovo Malasowe, and O. Nwankwo, "Empirical evidence of phishing menace among undergraduate smartphone users in selected universities in Nigeria," *IJEECS*, vol. 28, no. 3, p. 1756, Dec. 2022, doi: 10.11591/ijeecs.v28.i3.pp1756-1765.

[21] M. I. Akazue, K. O. Ahweyevu, C. O. Ogeh, and C. Asuai, "Development of a real-time phishing detection website via a triumvirate of information retrieval, natural language processing, and machine learning modules," *International Journal of Trend in Research and Development*, 2024, [Online]. Available: www.ijtrd.com

[22] I. Y. Hafez, A. Y. Hafez, A. Saleh, A. A. Abd El-Mageed, and A. A. Abohany, "A systematic review of AI-enhanced techniques in credit card fraud detection," *J Big Data*, vol. 12, no. 1, p. 6, Jan. 2025, doi: 10.1186/s40537-024-01048-8.

[23] E. W. T. Ngai, Y. Hu, Y. H. Wong, Y. Chen, and X. Sun, "The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature," *Decision Support Systems*, vol. 50, no. 3, pp. 559–569, Feb. 2011, doi: 10.1016/j.dss.2010.08.006.

[24] Y. Chen, C. Zhao, Y. Xu, C. Nie, and Y. Zhang, "Deep Learning in Financial Fraud Detection: Innovations, Challenges, and Applications," *Data Science and Management*, p. S2666764925000372, Aug. 2025, doi: 10.1016/j.dsm.2025.08.002.

[25] H. Abbassi, S. El Mendili, and Y. Gahi, "Adaptive, Privacy-Enhanced Real-Time Fraud Detection in Banking Networks Through Federated Learning and VAE-QLSTM Fusion," *BDCC*, vol. 9, no. 7, p. 185, Jul. 2025, doi: 10.3390/bdcc9070185.

[26] P. Wu and Y. Chen, "Enhanced detection of accounting fraud using a CNN-LSTM-Attention model optimized by Sparrow search," *PeerJ Computer Science*, vol. 10, p. e2532, Nov. 2024, doi: 10.7717/peerj-cs.2532.

[27] M. Aschi, S. Bonura, N. Masi, D. Messina, and D. Profeta, "Cybersecurity and Fraud Detection in Financial Transactions," in *Big Data and Artificial Intelligence in Digital Finance*, J. Soldatos and D. Kyriazis, Eds., Cham: Springer International Publishing, 2022, pp. 269–278. doi: 10.1007/978-3-030-94590-9_15.

[28] Ojugo, A. O. Eboka, O. E. Okonta, R. E. Yoro, and F. O. Aghware, "Genetic Algorithm Rule-Based Intrusion Detection System (GAIDS)," *Journal of Emerging Trends in Computing and Information Sciences*, 2012, [Online]. Available: http://www.cisjournal.org

[29] Q. Sun, T. Tang, H. Chai, J. Wu, and Y. Chen, "Boosting Fraud Detection in Mobile Payment with Prior

Knowledge," *Applied Sciences*, vol. 11, no. 10, p. 4347, May 2021, doi: 10.3390/app11104347.

[30] A. A. Ojugo., D. A. Oyemade., and D. Allenotor, "Comparative Stochastic Study for Credit-Card Fraud Detection Models," *African Journal of Computing & ICT*, 2015, [Online]. Available: www.ajocict.net

[31] I. Vorobyev and A. Krivitskaya, "Reducing false positives in bank anti-fraud systems based on rule induction in distributed tree-based models," *Computers & Security*, vol. 120, p. 102786, Sep. 2022, doi: 10.1016/j.cose.2022.102786.

[32] A. A. Ojugo *et al.*, "Forging a User-Trust Memetic Modular Neural Network Card Fraud Detection Ensemble: A Pilot Study," *J. Comput. Theor. Appl.*, vol. 1, no. 2, pp. 50–60, Oct. 2023, doi: 10.33633/jcta.v1i2.9259.

[33] A. Ali *et al.*, "Financial Fraud Detection Based on Machine Learning: A Systematic Literature Review," *Applied Sciences*, vol. 12, no. 19, p. 9637, Sep. 2022, doi: 10.3390/app12199637.

[34] M. I. Akazue *et al.*, "Handling Transactional Data Features via Associative Rule Mining for Mobile Online Shopping Platforms," *IJACSA*, vol. 15, no. 3, 2024, doi: 10.14569/IJACSA.2024.0150354.

[35] O. Deborah, A. Maureen, and I. Anthony, "A Framework for Feature Selection using Data Value Metric and Genetic Algorithm," *IJCA*, vol. 184, no. 43, pp. 14–21, Jan. 2023, doi: 10.5120/ijca2023922533.

[36] M. I. Akazue, I. A. Debekeme, A. E. Edje, C. Asuai, and U. J. Osame, "UNMASKING FRAUDSTERS: Ensemble Features Selection to Enhance Random Forest Fraud Detection," *J. Comput. Theor. Appl.*, vol. 1, no. 2, pp. 201–211, Dec. 2023, doi: 10.33633/jcta.v1i2.9462.

[37] F. O. Aghware *et al.*, "Enhancing the Random Forest Model via Synthetic Minority Oversampling Technique for Credit-Card Fraud Detection," *J. Comput. Theor. Appl.*, vol. 1, no. 4, pp. 407–420, Mar. 2024, doi: 10.62411/jcta.10323.

[38] M. D. Okpor *et al.*, "Pilot Study on Enhanced Detection of Cues over Malicious Sites Using Data Balancing on the Random Forest Ensemble," *J. Fut. Artif. Intell. Tech.*, vol. 1, no. 2, pp. 109–123, Sep. 2024, doi: 10.62411/faith.2024-14.

[39] D. R. I. M. Setiadi, A. R. Muslikh, S. W. Irianandа, W. Warto, J. Gondohanindijo, and A. A. Ojugo, "Outlier Detection Using Gaussian Mixture Model Clustering to Optimize XGBoost for Credit Approval Prediction," *J. Comput. Theor. Appl.*, vol. 2, no. 2, pp. 244–255, Nov. 2024, doi: 10.62411/jcta.11638.

[40] R. E. Ako *et al.*, "Effects of Data Resampling on Predicting Customer Churn via a Comparative Tree-based Random Forest and XGBoost," *J. Comput. Theor. Appl.*, vol. 2, no. 1, pp. 86–101, Jun. 2024, doi: 10.62411/jcta.10562.

[41] E. A. Otorokpo *et al.*, "DaBO-BoostE: Enhanced Data Balancing via Oversampling Technique for a Boosting Ensemble in Card-Fraud Detection," *AIMS Research Journal*, vol. 12, pp. 45–66, 2024, doi: 10.22624/AIMS/MATHS/V12N2P4.

[42] A. A. Ojugo, M. Akazue, P. Ejeh, C. Odiakaose, and F. Emordi, "DeGATraMoNN: Deep Learning Memetic Ensemble to Detect Spam Threats via a Content-Based Processing," *Kongzhi yu Juece / Control Decis*, 2023, [Online]. Available: https://www.researchgate.net/publication/374874600_DeGATraMoNN_Deep_learning_memetic_ensemble_to_detect_spam_threats_via_a_content-based_processing

[43] A. A. Ojugo and E. Ekurume, "Deep Learning Network Anomaly-Based Intrusion Detection Ensemble For Predictive Intelligence To Curb Malicious Connections: An Empirical Evidence," *IJATCSE*, vol. 10, no. 3, pp. 2090–2102, Jun. 2021, doi: 10.30534/ijatcse/2021/851032021.

[44] F. O. Aghware, R. E. Yoro, P. O. Ejeh, C. C. Odiakaose, F. U. Emordi, and A. A. Ojugo, "DeLClustE: Protecting Users from Credit-Card Fraud Transaction via the Deep-Learning Cluster Ensemble," *IJACSA*, vol. 14, no. 6, 2023, doi: 10.14569/IJACSA.2023.0140610.

[45] R. E. Yoro *et al.*, "Adaptive DDoS detection mode in software-defined SIP-VoIP using transfer learning with boosted meta-learner," *PLoS One*, vol. 20, no. 6, p. e0326571, Jun. 2025, doi: 10.1371/journal.pone.0326571.

[46] S. K. Aljunaid, S. J. Almheiri, H. Dawood, and M. A. Khan, "Secure and Transparent Banking: Explainable AI-Driven Federated Learning Model for Financial Fraud Detection," *JRFM*, vol. 18, no. 4, p. 179, Mar. 2025, doi: 10.3390/jrfm18040179.

[47] A. R. Khalid, N. Owoh, O. Uthmani, M. Ashawa, J. Osamor, and J. Adejoh, "Enhancing Credit Card Fraud Detection: An Ensemble Machine Learning Approach," *BDCC*, vol. 8, no. 1, p. 6, Jan. 2024, doi: 10.3390/bdcc8010006.

[48] M. Hasan, M. S. Rahman, M. J. Morshed Chowdhury, and I. H. Sarker, "CNN Based Deep Learning Modeling with Explainability Analysis for Detecting Fraudulent Blockchain Transactions," *Cyber Security and Applications*, vol. 3, p. 100101, Dec. 2025, doi: 10.1016/j.csa.2025.100101.

[49] I. D. Mienye and N. Jere, "Deep Learning for Credit Card Fraud Detection: A Review of Algorithms, Challenges, and Solutions," *IEEE Access*, vol. 12, pp. 96893–96910, 2024, DOI: 10.1109/ACCESS.2024.3426955.

[50] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, no. 1, p. 6, Dec. 2020, doi: 10.1186/s12864-019-6413-7.