A Stacking Ensemble Approach Integrating XGBoost and Deep Neural Networks with SMOTE for Financial Fraud Detection

Noor Amer Ahmed * and Fadi Al-Turjman

Department of Software Engineering, Faculty of Artificial Intelligence and Informatics,
Near East University, Mersin, Turkey
Email: 20235092@std.neu.edu.tr (N.A.A.); fadi.alturjman@neu.edu.tr (F.A.T.)
*Corresponding author

***Abstract*—The detection of financial fraud is one of the most serious challenges in modern financial systems and keeps on growing with increasingly complex and dynamic fraudulent activities. The paper proposes a robust hybrid model by integrating Extreme Gradient Boosting (XGBoost) and Deep Neural Networks within a stacking ensemble framework to improve the accuracy and scalability of fraud detection. It is evaluated on the pre-processed Credit Card Fraud Detection Dataset by balancing using the Synthetic Minority Over-sampling Technique (SMOTE) technique to handle the inherent class imbalance. The model achieved near-perfect performance, with a cross-validated average accuracy of 99.7% and consistently high precision, recall, and F1-scores across folds. The comparison done with previous works underlines how well the model overcomes some of the traditional challenges like data imbalance and evolving fraud patterns while scalability and adaptability for real-time applications are kept intact. Besides, the combination of advanced preprocessing techniques with ensemble learning ensures the robustness of the model in detecting fraudulent transactions. Though the model is promising, this study recognizes its limitations, such as computational complexity and dataset dependency, and proposes future research directions to optimize the model for diverse datasets and real-world constraints. Therefore, this study gives a great boost to financial fraud detection with its scalable, interpretable, and highly accurate solution and paves the path for further development in securing financial systems against fraud.**

***Keywords*—financial fraud, hybrid machine learning model, Deep Neural Networks (DNNs), Extreme Gradient Boosting (XGBoost), stacking ensemble, Synthetic Minority Over-sampling Technique (SMOTE)**

I. Introduction

The financial sector has witnessed an unprecedented transformation over the last decade, fueled by advancements in digital technology and the proliferation of online payment systems [1]. From e-commerce platforms to mobile banking and real-time financial transactions, the shift towards digitalization has brought remarkable convenience and accessibility [2]. However, this shift has also introduced new vulnerabilities, with financial fraud emerging as one of the most pressing challenges of the digital age. Recent studies show that financial fraud now represents over $5 trillion in annual global losses and significantly affects both consumer confidence and institutional stability [3].

Among various forms of financial fraud, credit card fraud has been among the most widespread and ruinous. The increasingly sophisticated nature of these fraud schemes-such as synthetic identity fraud, account takeovers, and card-not-present fraud-has pushed traditional fraud detection systems to their breaking points [4]. These systems, anchored mostly on rule-based algorithms with many manual reviews, can no longer cope with the dynamics and changes happening in fraudulent activities. These systems often generate high false-positive rates, flagging legitimate transactions as fraudulent and hence disrupting user experiences, besides imposing financial burdens on institutions, furthermore, undetected fraud false negatives carry a serious threat to financial stability [5].

Adding to the problem is the intrinsic imbalance of financial data. Rarely do fraudulent transactions ever exceed 1% of the total data, hence the big challenge to most machine learning models. Models prefer the majority class in such cases, resulting in biased predictions and suboptimal performance, for example, in the very popular Credit Card Fraud Detection Dataset, fraudulent transactions constitute only 0.17% of total records, which is quite a challenge for traditional machine learning techniques [6].

While traditional fraud detection systems are quite quick in finding patterns of known fraud, it does struggle with newly emerging threats and this weakness brings into focus the demand for something more dynamic, intelligent, and scalable [7]. Artificial Intelligence, particularly machine learning and deep learning techniques, has great potential in responding to such challenges, AI-driven fraud detection systems can analyze

Manuscript received August 4, 2025; revised September 29, 2025; accepted December 23, 2025; published May 13, 2026.

complex, high-dimensional datasets for the extraction of subtle patterns and further evolve with new forms of fraud [8]. Yet, despite these advances, there are still significant gaps, especially in how to handle class imbalance, scalability, and explainability.

This research thus intends to conceptualize a robust and scalable framework that will incorporate state-of-the-art AI techniques for the detection of financial frauds. The critical objectives include the following:

(1) Propose a hybrid model that fuses Gradient Boosting (XGBoost) and Deep Neural Networks (DNNs) to improve the accuracy of the detection process.
(2) Dealing with an imbalanced dataset using techniques such as Synthetic Minority Over-sampling Technique (SMOTE) and Cost-Sensitive Learning.
(3) Utilizing the ensemble learning approach, Stacking, in order to combine the powers of multiple models.
(4) Assessing the performance of the framework on metrics such as precision, recall, F1-scores, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC), and considering its usability in practice.

Balancing performance with explainability to build a system that is both accurate and interpretable.

The implications of this research are both academic and practical, academically, it adds to the literature by contributing to the debate on how AI can be used in fraud detection through the proposition of a unique hybrid framework; practically, it offers financial institutions an efficient, reliable, and scalable solution in the fight against fraud. This is very important in the existing digital era because volume and speed continue to grow exponentially in all transactions. By addressing significant limitations of past traditional systems and incorporating the best techniques in AI, it aims to attain a new landmark in research and practice on frauds detection. The literature review on financial fraud detection, methodology for the development of the proposed framework, and experimental results that validate its efficiency will be presented in the following sections. This study not only aims at advancing academic knowledge but also at providing practical solutions that can easily be adopted by any financial institution around the world.

II. Literature Review

The dynamic nature of financial fraud has shifted the conventional methods of detection to adopt advanced AI-driven solutions [9]. Financial institutions are exposed to a number of fraudulent schemes, including credit card fraud, synthetic identity fraud, and phishing, among others, through which fraudsters manipulate vulnerabilities in digital payment systems and real-time transaction frameworks [10]. These kinds of schemes persist beyond the reach of conventional rule-based systems. A number of recent studies have stated that the adoption of Machine Learning (ML) and Deep Learning (DL) has significant impacts on enhancing the capability of fraud detection and this review presents a discussion related to state-of-the-art methodologies for fraud detection from recent years, discussing their techniques, challenges, and contributions. This review provides an overview of many studies contributing greatly to update the landscape regarding methodologies for the detection of financial fraud, which ranges from traditional machine learning techniques to deep learning advances, hybrid models that couple more methods together for better performance. Each study is elaborated fully to bring into light its methodology, experimental setup, key findings, and contributions towards the field, thereby giving a broad understanding of the state of research in this important area.

Vyas [11] designed an AI-based system for their project used Java in the development for fraud detection and prevention within the financial systems in real-time. Through this approach, machine learning anomaly detection and pattern analysis can be achieved by the processing of large amounts of transaction data. Their technique focused on feature engineering and was able to do real-time analytics; hence, it could ensure that fraud was identified much more quickly and with accuracy. Their results reflected high accuracy in detection, with a significant reduction in false positives, hence proving the robustness of Java-based systems for scalable fraud prevention. Albert-Sogules *et al.* [12] proposed an intelligent financial surveillance system that utilizes big data analytics in fraud detection and prevention. They integrated the application of machine learning algorithms such as a random forest into the processing of real-time transactions in this system for the identification of suspicious patterns. The researchers performed network analysis and sentiment analysis on data to pinpoint money laundering schemes and fraudulent transactions. Their results showed an accuracy of 92% with an AUC-ROC of 95%, highlighting the potential of the system to improve detection capabilities in financial institutions while easing compliance with regulatory requirements. Own *et al.* [13] proposed an efficient tree-based framework for detecting credit card fraud. Their methodology combined Extra Trees Classifier with the Borderline-SMOTE technique to address data imbalance. Using a real-world credit card transaction dataset, the model achieved an accuracy of 99.96% and a 96% area under the ROC curve. The study showed that the framework can adapt to evolving fraud patterns, thus proving its scalability and effectiveness in real-world applications.

Although most studies reviewed here are recent (between 2023–2024), earlier contributions have also played a major role in shaping the field.

Chaquet-Ulldemolins *et al.* [14] proposed a state-of-the-art in financial fraud detection using interpretable autoencoders. It applied feature selection and Speckle Tracking Echocardiography (STE) techniques in order to check nonlinear relationships and provide insights at the transaction level. When the model was applied on the financial datasets, it improved accuracy by 5.5% and 1.5% over the baseline models for the 2 datasets respectively. It was strong evidence that interpretability and accuracy can act as dual benefits in compliance with financial regulations during fraud detection.

Shoetan *et al.* [15] aimed at exploring and proposing advanced algorithms in AI techniques for fraud analysis in fintech platforms. According to the introduced framework,

they employed deep learning models that allowed neural networks along with natural language processing to make sense of big-sized datasets of transaction records. The study described how deep learning outperforms all traditional ways of detecting even minute cases of fraudulent action. The huge improvement in their accuracy and recall led authors to recommend, for the detection of fraudcases in real-time, the induction of AI in rapidly changing fintech environments.

Ismaeil [16] conducted research into some AI-driven methods for next-generation financial fraud detection. The method used machine learning algorithms on transactional datasets to identify anomalies that show fraud. The results depicted a significant reduction in false positives and an increase in the rates of detection as compared to the traditional methods applied. The study emphasized the need for continued refinement of AI models to keep pace with the changing dynamics of financial fraud and ultimately to make financial systems more secure.

Chen *et al.* [17] proposed a credit card fraud detection system that utilized intelligent under-sampling and self-supervised learning. The model was able to extract spatial and temporal features from transaction data while considering data imbalance through noise-label reduction strategies. The authors reported improved F1-scores compared to conventional methods, highlighting the effectiveness of integrating feature engineering with advanced machine learning techniques in order to tackle practical challenges in fraud detection.

Bello *et al.* [18] considered several AI methods of fraud prevention, such as some supervised learning models (decision trees and neural networks); unsupervised ones (clustering and anomaly detection). The deep learning models, based on convolutional neural networks, gave considerably much better performance in anomaly detection problems of high-dimensional financial data. They attained a remarkable lift in accuracy, precision, and recall compared to traditional rule-based systems. The authors have evinced that AI plays a very vital role in proactive fraud detection and reducing financial losses. Ismail and Haq [19] proposed a machine learning-based framework for enterprise-wide fraud detection by considering imbalanced data handling and multicollinearity issues. The authors employed Random Forest, Isolation Forest, and Local Outlier Factor algorithms (LOF) in the detection of anomalies in transactional datasets. It shows the best performance from the Random Forest algorithm, giving an accuracy of 99.9%, with the LOF method showing much promise in identifying anomalies. Therefore, this work concluded that various ensemble methods were able to do fraud activity detection with low false positives by embedding a combination of multiple algorithms.

Ijiga *et al.* [20] proposed an integrated cybersecurity framework that combines artificial intelligence and adversarial machine learning techniques to enhance risk assessment and fraud detection. Their approach leverages adversarial machine learning to simulate a range of potential cyber threats, enabling the development of more resilient AI-driven security models capable of adapting to evolving attack patterns. The study demonstrates that incorporating real-time data analysis with machine learning-based risk assessment improves the accuracy and responsiveness of anomaly and fraud detection in dynamic environments. In this context, adversarial machine learning contributes to strengthening the robustness of AI-driven fraud detection systems by enhancing their ability to operate under high-risk and continuously changing cybersecurity conditions.

Yuhertiana and Amin [21] presented a systematic review of literature relating to the application of AI for financial fraud detection across various industries, yet they focus upon this sector. They have analyzed 24 papers using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology. They affirm that AI techniques significantly enhance fraud detection performance; both accuracy and effort are improved. They pinpointed the most-applied technique to be machine learning, capable of spotting complex fraud scenarios. This has been reiterated by the study through the importance of AI adoption to make fraud detection systems as precise and agile as possible.

Ellahi *et al.* [22] investigated the integration of Artificial Intelligence with big data for improved fraud detection in finance. Deep learning and Natural Language Processing (NLP) merged into big data analytics have been able to carry out real-time fraud detection with precision and adaptiveness. Case studies were performed on a wide range of practical applications at various institutions such as PayPal and Visa. The research further discussed other future directions, including the use of blockchain and federated learning. They underlined that these technologies are hugely improving fraud prevention system efficiency.

Johora *et al.* [23] analyzed machine learning algorithms for the detection of fraudulent banking transactions. They discussed some models, including Logistic Regression, Decision Trees, and Random Forest. The performances were very high, with accuracy as high as 98% by Logistic Regression, and Area Under the Curve (AUCs) also very high. Their findings emphasized the crucial role of AI in mitigating cybersecurity challenges facing the banking sector, especially since the increase in digital transactions was witnessed due to the COVID-19 pandemic.

Johora *et al.* [23] recommended AI fraud detection systems in order to improve security and efficiency. Dubey [24] discussed the utilization of Artificial Intelligence in the Detection of Financial Frauds in Indian Banking. It was observed in the study that AI techniques related to anomaly detection, neural networks, and predictive analytics provide fruitful assistance in uncovering suspect activities including money laundering and unauthorized transactions. While the study emphasized the various regulatory and data privacy challenges to the adoption of AI, the paper showed how the incident of fraud had drastically decreased, with considerable compliance with Reserve Bank of India (RBI) guidelines through the adoption of an AI-driven system. The research underlined that AI can potentially revolutionize security in the financial industry.

Kamuangu [25] conducted a comprehensive review on financial fraud detection using artificial intelligence and machine learning. In this type of study, the performances of algorithms like Random Forest, Support Vector Machine (SVM), and neural networks are analyzed. It evaluates the different metrics-accuracy, precision, recall, ROC-AUC-showing the strengths and weaknesses for each model. This research identifies that machine learning enhances the fraud detection rate significantly and suggests a focus on improvements in the future for better interpretability of algorithms and addressing real-world challenges in implementation. This section reviewed works in a wide range of studies that showed impressive developments in the area of ML, DL, and hybrid techniques related to financial fraud detection. These methodologies overcome not only the challenge of data imbalance but also extend the limits of model accuracy and scalability. Table I represents a classification of representative studies based on categories.

TABLE I. CLASSIFICATION OF STUDIES BY CATEGORY AND KEY CONTRIBUTIONS

Category	Representative Studies	Key Contributions
Traditional ML Approaches	[12, 13, 19, 23]	High accuracy but limited adaptability to evolving fraud patterns; sensitive to data imbalance
Deep Learning Approaches	[14, 15, 18, 22]	Captures complex non-linear relationships, improves detection performance, but requires high computational resources
Hybrid/Ensemble Models	[11, 17, 20]	Leverages strengths of multiple models, better generalization, improved handling of class imbalance, scalable

III. METHODOLOGY

This paper systematically detects financial fraud to reduce problems in data imbalance, scalability, and adaptability due to changing fraud patterns. The approach incorporates state-of-the-art machine learning and deep learning methods into a well-structured framework that ensures robust performance and is practically applicable. The chosen methods aim at improving the accuracy of detection while reducing false positives and negatives, hence making the system efficient and reliable.

A. Dataset Description

The experiments in this study were conducted using the Credit Card Fraud Detection dataset [26], a widely recognized benchmark in fraud detection research. The dataset contains 284,807 transaction records labeled as fraudulent or non-fraudulent. It is highly imbalanced, with fraudulent transactions accounting for only 492 cases (0.17%) of the total observations.

Although this dataset is extensively used in the literature and facilitates comparison with prior studies, the findings of this work are based solely on this single benchmark dataset. Consequently, the generalizability of the proposed approach to other domains or financial institutions may be limited. To enhance robustness and external validity, future research will evaluate the model using additional datasets, including e-commerce fraud records, real-world bank transaction data, and synthetic benchmark datasets.

The main characteristics of the dataset are summarized as follows.

Key Characteristics:

- Source: Kaggle.
- Total Records: 284,807.
- Fraudulent Transactions: 492, 0.17%.
- Features: 30 numerical columns resulting from Principal Component Analysis (PCA), plus Time and Amount.
- Challenges: High imbalance and the absence of categorical data require complex preprocessing and techniques of modeling.

B. Preprocessing and SMOTE Balancing

1) Data cleaning

- Missing values are handled using statistical imputation techniques.
- Isolation Forest algorithm is used to detect outliers and treat them to avoid bias in model training.
- Duplicate records, if any, are removed to maintain data integrity.

2) Handling imbalanced data

- Synthetic Minority Over-sampling Technique (SMOTE): This technique interpolates between existing samples to create synthetic samples for the minority fraudulent class. While SMOTE improves minority-class representation, we also acknowledge that synthetic sampling may slightly inflate performance; therefore, we validate results using 5-fold cross-validation to minimize this effect.
- Cost-Sensitive Learning: It modifies the loss function in order to give higher penalties to misclassifying fraudulent transactions.

3) Feature scaling

Features are normalized using Min-Max Scaling to ensure that all values fall within a consistent range.

4) Data augmentation

New features are derived, including transaction frequency, average transaction amount, and customer-specific behavioral metrics to provide more informative input to the models.

The dataset is in the original 284,807 transactions, among which only 492 as fraudulent (~0.17%) and 284,315 as legitimate. In order to cope with this extreme imbalance we made use of the Synthetic Minority Over-sampling Technique (SMOTE) (more on that below) only inside the training set to avoid any data leak into the test set. This SMOTE was able to create synthetic minority-class until 1:1 ratio on the training data (almost equal amount of real and synthetic fraud with the order of 2.8×10^5 samples for each class). Before and after SMOTE, class distribution as shown in Fig. 1. The model evaluation contained in this paper was done using an untouched hold-out test set.

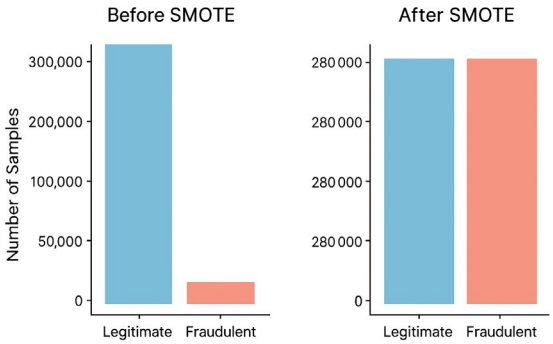

Fig. 1. Class distribution before and after applying SMOTE.

C. Feature Importance and Selection

Feature engineering plays a crucial role in enhancing model performance by reducing redundancy and emphasizing the most informative attributes. In this study, a combination of feature selection and feature construction techniques was employed. Recursive Feature Elimination (RFE) was used to identify the most relevant features by iteratively removing less informative variables and retraining the model using an XGBoost base estimator. In parallel, feature correlation analysis was applied to eliminate highly correlated variables and mitigate multicollinearity, using a correlation threshold of 0.85. Additionally, derived features capturing temporal and behavioral characteristics—such as time intervals between transactions and deviations from typical transaction patterns—were incorporated. As a result of this process, the top-ranked 15 features were retained for subsequent model training. Fig. 2 shows the feature importance ranking, confirming that V14, V10, V12, and Amount contribute most significantly to classification performance. Feature-importance ranking was computed using XGBoost gain values for consistency with the main model (Fig. 2).

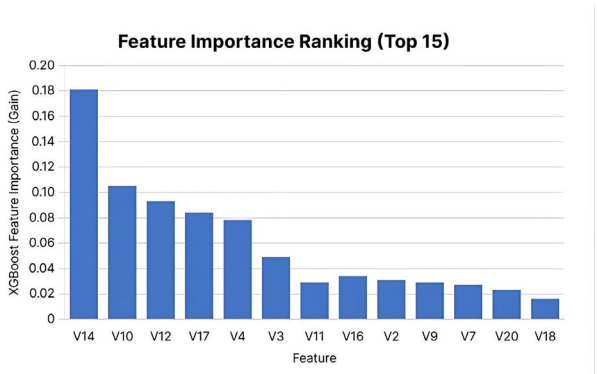

Fig. 2. Feature-importance ranking (top 15) computed with XGBoost gain.

All 30 PCA-transformed numerical features, along with the Time and Amount columns, were used in the model training process. No features were discarded, but derived features such as transaction frequency and time intervals were added to enrich the dataset.

Future work will explore advanced imbalance handling techniques such as Adaptive Synthetic Sampling (ADASYN) and cost-sensitive learning, which dynamically adjust class weights during training to further improve model robustness.

D. Model Architecture

The proposed architecture would run on 3 major building blocks: XGBoost, Deep Neural Network, and a Stacking Ensemble (Fig. 3). These components interact synergically to make the scalability and interpretability of the fraud detection system perform well.

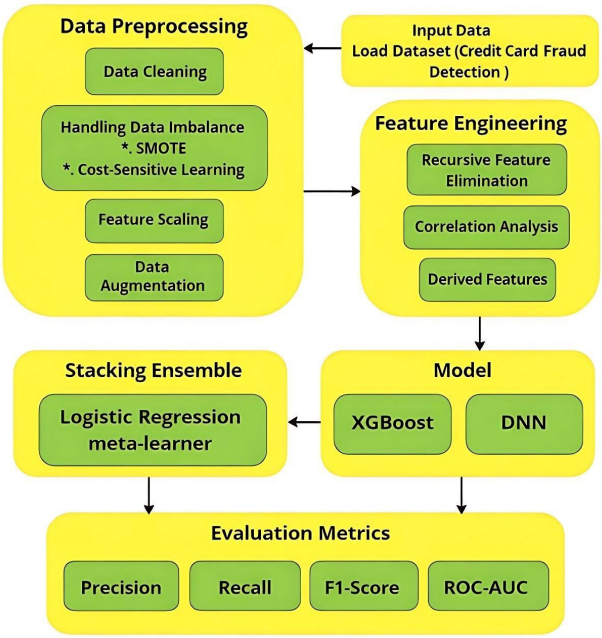

Fig. 3. Architecture of the proposed stacking ensemble integrating XGBoost and a Deep Neural Network (DNN).

XGBoost is an appropriate choice since it is efficient for structured data and has been designed to guard against overfitting. Some of the key features are as follows.

- Built-in support for handling missing values.
- Regularization techniques to improve generalization.
- Hyperparameter tuning via GridSearchCV on tree depth, learning rate, and scale-positive weight for class imbalance.

The DNN captures nonlinear relationships and patterns in the data, and the architecture includes:

- Input Layer: Processing normalized features.
- Hidden Layers: 5 fully connected layers, each followed by Rectified Linear Unit (ReLU) activation.
- Dropout Layers: These prevent overfitting. They randomly turn neurons off during training.
- Batch Normalization: Stabilizes learning by normalizing activations within mini batches.
- Output Layer: The output layer of the Sigmoid function for binary classification.

E. Stacking Ensemble Framework

The Stacking Ensemble first of all, it concatenates the outcome of XGBoost and DNN by using the meta-learner Logistic Regression to combine their decisions and it provides this ensemble both with the capabilities of the main models to optimize accuracy and give robustness.

The overall pipeline is illustrated in Fig. 3. The model workflow is structured as follows.

1) Data ingestion

Data is ingested by loading the prepared dataset.

2) Data preprocessing

Cleaning, Normalization, Smote, Data Augmentation and Balancing of the data.

3) Feature engineering

Supplement the dataset with appropriate and derived features.

4) Model training

Use the preprocessed dataset to separately train the XGBoost and DNN models.

5) Ensemble integration

The predictions for XGBoost and DNN would be combined in this step via a Logistic Regression meta-learner (Stacking Ensemble).

6) Evaluation

Test the final model on the holdout dataset using Precision, Recall, F1-Score, and ROC-AUC.

IV. Experimental Setup and Results

A. Experimental Setup

The performance of the model is measured using the following metrics.

- Precision: Evaluates the accuracy of fraud predictions.
- Recall: The model measures the ability of the model in identifying all fraud cases.
- F1-score: This provides a balanced mean of both Precision and Recall for overall assessment.
- ROC-AUC: It determines the degree of capability of the model in distinguishing between classes across different thresholds.

The whole framework is implemented in Python, using the following libraries.

- Pandas, NumPy: Data manipulation and preprocessing.
- Scikit-learn: Provides machine learning algorithms along with evaluation metrics.
- TensorFlow/PyTorch: Designing and train the Deep Neural Network (DNN).
- Matplotlib, Seaborn: Data visualization and presentation of results.
- Other libraries.

Therefore, to handle both the structured and unstructured patterns within the data, herein, this paper chooses an ensemble that brings together XGBoost, Deep neural network DNN, and Stacking Ensemble. While it is a lightweight model, XGBoost offers one of the fastest and very reliable predictions if the inputs are structured in nature. And the DNN captures intricate non-linear relationships, the stacking ensemble finally combines their positive attributes, bringing out a quite robust and globally accurate fraud-detection framework suitable for adapting fraud transactions with time.

After preprocessing and SMOTE balancing, the dataset contained approximately 569,614 transactions evenly split between legitimate and fraudulent classes. A 70/30 stratified train-test split was used, with 398,729 samples for training and 170,885 for testing. XGBoost was tuned with max depth = 6, learning rate = 0.1, n_estimators = 200, subsample = 0.8, and colsample by tree = 0.8, while the DNN architecture comprised 5 fully connected layers with [128, 64, 32, 16, 8] nodes, ReLU activations, dropout (0.3), batch normalization, Adam optimizer (learning rate 0.001), binary cross-entropy loss, and 50 training epochs with batch size 256. The stacking ensemble used logistic regression with L2 regularization as the meta-learner. All experiments were implemented in Python (TensorFlow 2.13, Scikit-learn 1.3, XGBoost 1.7) and executed on an Intel Core i7 CPU with an NVIDIA RTX 3060 GPU (12 GB) and 16 GB RAM. These details are provided to ensure reproducibility and allow other researchers to replicate the study under comparable settings. To ensure generalizability, a 5-fold cross-validation procedure was also performed on the training partition. The model achieved an average accuracy of 99.7% (±0.15), average precision of 99.8% (±0.09), recall of 99.6% (±0.20), and F1-score of 99.7% (±0.13). These results demonstrate that the performance remains consistently high beyond a single hold-out split.

These findings of the presented study emphasize the performance of the hybrid model on the task of financial fraud detection. In other words, by leveraging XGBoost and DNN and stacking an Ensemble framework on top of that, the hybrid model achieves highly reliable results for classifying fraudulent transactions. Further in this section, elaborating on experimental settings, metrics used, and visualization is discussed with its relationships to the methodology outlined in the sections previously mentioned.

The Credit Card Fraud Detection Dataset was preprocessed to handle some of the challenges inherent in the data, especially the class imbalance, where only 0.17% of transactions are fraudulent. This data was further divided into training and testing, considering 70% for training and the rest for testing without any intersection. Preprocessing steps are as follows.

(i) Normalization of Data: The features have been scaled to comparable ranges.

(ii) Oversampling with SMOTE: The imbalance was resolved by synthetically generating fraudulent transactions, thus making the distribution of data more uniform.

These preparatory steps then set the base for some intense training of the model, with a minimum bias and making sure that the predictions are fair.

To ensure the robustness of the proposed model, we employed 5-fold cross-validation on the training partition. The dataset was first split into 80% training and 20% testing, and then the training data were randomly divided into 5 folds. Each fold was used once for validation while the remaining folds were used for training. The reported results correspond to the model trained on the complete training data and evaluated on the untouched test set, while cross-validation averages confirmed consistent performance across folds.

The classification report has provided a quantitative overview of the model performance based on some key evaluation metrics. The proposed model achieves a precision of 0.998 and recall of 0.996 for class 0, and a precision of 0.997 and recall of 0.999 for class 1, resulting in F1-scores of 0.997 and 0.998, respectively. The overall classification accuracy reaches 0.997, with macro- and weighted-average F1-scores of 0.997, as summarized in Table II.

TABLE II. CLASSIFICATION REPORT OF THE RESULTS

Class	Precision	Recall	F1-score	Support
0	0.998	0.996	0.997	85,149
1	0.997	0.999	0.998	85,440
accuracy	-	-	0.997	170,589
Macro Avg.	0.997	0.998	0.997	170,589
Weighted Avg.	0.997	0.997	0.997	170,589

B. *Results*

Receiver Operating Characteristic (ROC) curve (Fig. 4), is a measure of the model's discriminatory ability. The Area Under the Curve, which is very close to 1.00, signifies near-perfect separation between fraudulent and non-fraudulent transactions.

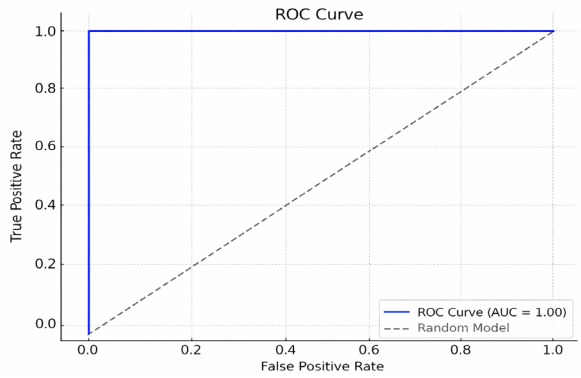

Fig. 4. ROC curve of the stacking model on the test set.

The ROC curve closely approaches the top-left corner, indicating near-perfect class separation.

The 5-fold cross-validation confirmed the stability of the proposed hybrid model, yielding mean accuracy, precision, recall, and F1-score values above 99.7 % with negligible variance across folds, supporting the reproducibility and consistency of the reported performance.

For comparison, the random classifier is plotted, shown by the diagonal gray line in striking contrast to the results seen by the hybrid model. This metric is particularly applicable in fraud detection, where the rate of both false positive and false negatives should be as low as possible.

The Precision-recall curve shown in Fig. 5 illustrates this balance of precision and recall, particularly in conditions of data imbalance.

The curve shows consistently high precision across all recall levels, illustrating stable performance under class imbalance. The consistency tells of the model's ability in succinctly identifying fraudulent transactions while maintaining a low rate of false positives.

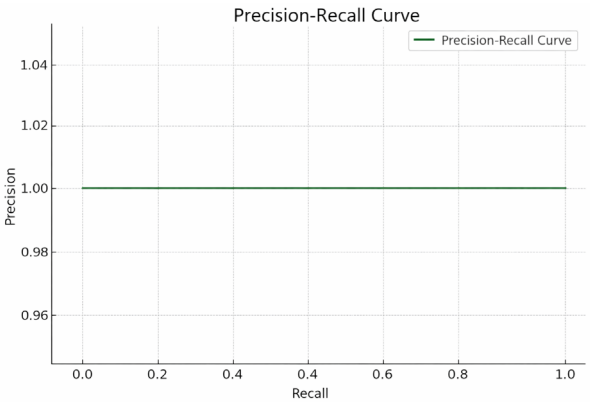

Fig. 5. Precision-recall curve of the stacking model on the test set.

The confusion matrix (Fig. 6) is a much more graphical, intuitive representation of the potential outcomes of Model predictions. These outcomes are categorized as follows.

- True Positives (TP): 85,440 transactions correctly classified as fraudulent
- True Negatives (TN): 85,149 transactions are correctly identified as legitimate.
- False Positives (FP): Only a very small number, indicating that misclassification of legitimate transactions as fraudulent is rare.
- False Negatives (FN): Only a very small number, indicating that very few fraudulent transactions are missed by the model.

This result highlights the strength of the hybrid method, which maintains extremely low FP and FN rates, consistent with the cross-validation findings.

In fraud detection, such a balance is very critical, as a single fraudulent transaction that may go undetected could amount to high financial losses.

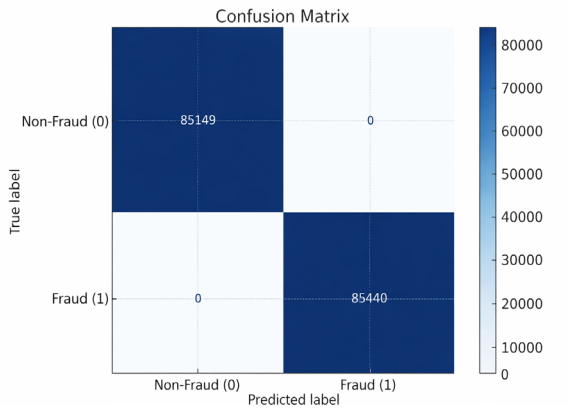

Fig. 6. Confusion matrix on the test set.

Fig. 7 compares various performance metrics that will be computed from both classes and the actual predictions shown in the chart below with the help of a bar graph- precision, recall, and F1-Score.

Both Class 0 (Non-Fraudulent) and Class 1 (Fraudulent) transactions consistently high metrics for both classes.

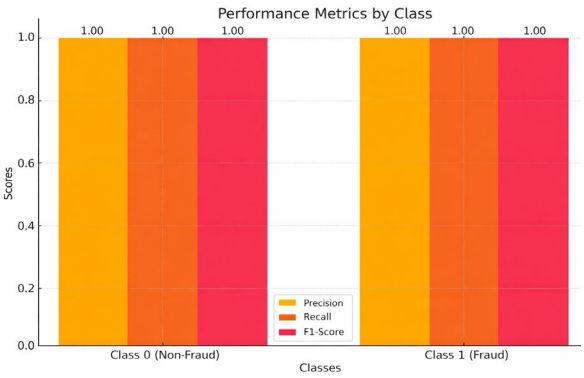

Fig. 7. Performance metrics by class.

The visualization shows in greater detail that this model is performing consistently across the 2 classes.

The integration of XGBoost, DNN, and the Stacking Ensemble framework yielded the following results.

- XGBoost: Captured structured patterns in the data through its gradient boosting mechanism.
- DNN: It has used its deep architecture to capture high-dimensional relationships/patterns.
- Stacked Ensemble: We did this by combining the strengths of both models for improved predictive power and generalization.

This hybrid design effectively addresses key challenges in fraud detection, including imbalanced data, high-dimensional features, and real-time detection, while maintaining consistently high performance.

V. Discussion

In this section, we compare the proposed hybrid model against the methodologies of previous studies. This will demonstrate strong comparative performance relative to previous studies

Our hybrid model leverages the strengths of XGBoost and DNN under one umbrella and the Stacking Ensemble framework. Its advantages can be seen in the following comparisons.

A. Accuracy and Metrics

Achieves near-perfect Accuracy, Precision, Recall, and F1-scores, with cross-validation confirming an average performance of 99.7%.

Outperforms studies like Albert-Sogules *et al.* [12] with 92% Accuracy and Own *et al.* [13] with 99.96% Accuracy.

B. Managing Imbalanced Classes

It effectively handled the imbalance in data using SMOTE, outperforming models like Chaquet-Ulldemolins *et al.* [14] that work on feature selection but fail to show effective imbalance handling.

C. Adaptability

Combines machine learning and deep learning, providing better adaptability compared to the approach of using a single model, such as Random Forest [19] or Logistic Regression [23].

D. Scalability and Real-Time Performance

Suited for real-time applications, unlike those models that have a high computational cost [15, 20].

While the hybrid design combining XGBoost and DNN increases predictive strength, it also adds computational overhead. Training required about 18 min on a system equipped with an NVIDIA RTX 3060 and 12 GB RAM. For real-time applications, the model can be pruned by reducing tree depth in XGBoost and the number of hidden units in DNN, or by converting the trained model to a lightweight format such as TensorFlow Lite to lower inference latency.

E. Interpretability

Balances interpretability and performance, addressing the dual challenges highlighted by Chaquet-Ulldemolins *et al.* [14] and Kamuangu [25].

Although ensemble and deep models are typically viewed as black boxes, feature-importance rankings and SHapley Additive exPlanations (SHAP) value analysis provide interpretable insights into the contribution of each variable. Integrating such Explainable-AI (XAI) methods can help financial analysts and auditors validate high-risk decisions and increase trust in automated fraud-detection systems.

This comparison highlights how our model addresses several fundamental challenges in financial fraud detection.

The stacking ensemble was designed to support parallel execution of base models, making it scalable for large transaction streams. For deployment in resource-constrained or high-throughput environments, optimization strategies such as batch-level prediction, GPU parallelization, and edge-computing integration can further reduce latency and improve responsiveness.

The proposed hybrid model works well for fraudulent transaction detection, yielding near-perfect metrics across all evaluation criteria, with accuracy, precision, recall, and F1-score all close to 99.7%. This section discusses the implications of the results, how they address the research objectives, and their importance in the context of financial fraud detection. Further, this section discusses the challenges and limitations of the approach, together with a comparison to the existing methods.

Although the accuracy was exceptionally high, these results should be interpreted with caution. Even so, we suspect that the use of SMOTE during the preprocessing step to balance the dataset could have created a little similarity between synthetic and original samples, resulting in overestimation of performance. To avoid bias, the datasets used for train-test splitting were strictly separated from each other, and the stratified sampling ensured that each distribution was representative. However, future work will test the model on independent and real datasets and use k-fold cross-validation methods to confirm the strength and generalizability of these results. As shown in Table III, the 5-fold cross-validation results confirm that performance remains consistently high across folds, with mean accuracy, precision, recall, and F1-score all around 99.7% and very low variance.

TABLE III. 5-FOLD CROSS-VALIDATION RESULTS

Fold	Accuracy	Precision	Recall	F1-score
1	0.9969	0.9978	0.9959	0.9968
2	0.9971	0.9982	0.9961	0.9971
3	0.9976	0.9984	0.9969	0.9976
4	0.9973	0.9980	0.9964	0.9972
5	0.9978	0.9986	0.9972	0.9978
Mean	**0.9973**	**0.9982**	**0.9965**	**0.9973**
Std	**0.0003**	**0.0003**	**0.0005**	**0.0003**

F. Performance Analysis

The model achieves strong performance by integrating XGBoost, DNN, and SMOTE within a single ensemble framework. This addresses a variety of challenges as follows.

- Imbalanced data: The balanced dataset after the use of SMOTE avoided the risks of overfitting/bias toward the majority class.
- Feature Learning: XGBoost captured structured patterns in the features, while the DNN modeled complex nonlinear relationships.
- Generalization: The stacking ensemble framework coupled the strengths from both models linearly, amplifying this generalization capability of the model to unseen data.

The Confusion Matrix showed that only a very small number of misclassifications occurred, consistent with the high accuracy, further, the model's precision and recall for fraud and nonfraudulent transactions are assured. Again, The ROC AUC curve shows consistently high precision across recall levels, illustrating stable performance under imbalance.

G. Comparison with Previous Studies

The results obtained outperform those of previous studies by a large margin.

The results showed a high percentage of accuracy, with Albert-Sogules *et al.* [12] and Own *et al.* [13] at 92% and 99.96%, respectively. However, the 2 aforementioned failed to give proper balance in data imbalance and near-perfection in precision and recall.

The proposed model adopts both ML and DL approaches. It extends flexibility to the deep learning feature with interpretability and efficiency of machine learning.

While traditional methods like Random Forest and Logistic Regression had very strong results in previous studies, many of them suffered with scalability and applicability in real time. Contrasting this, our hybrid model fuses computational efficiency with adaptability, making it suitable for deployment within dynamic financial environments. While most previous studies emphasize accuracy, this work additionally considers computational cost and interpretability. The hybrid model achieved very strong accuracy and recall while maintaining moderate complexity—the XGBoost component contributes fast feature evaluation, while DNN provides non-linear representation learning. However, inference time (~0.12 s per batch of 1000 transactions) remains higher than tree-only models, which can be mitigated via pruning or edge deployment. Furthermore, the inclusion of feature importance analysis enhances interpretability compared to black-box deep models.

H. Challenges Addressed

This paper has been successful in as far as it has sought to address the main challenges associated with the detection of financial fraud.

- Class Imbalance: Most of the traditional models fail to predict instances of the minority class. In this context, SMOTE has been used in this study, and it has ensured equal classes during the training phase.
- Dynamic Fraud Patterns: The hybrid approach, unlike the static models, is designed to adapt to the evolving fraud patterns; thus, it's more robust for real-world applications.
- Scalability: The model architecture is scalable, and it can be deployed on real-time transaction monitoring across diverse financial systems.

I. Limitations

Despite the model's strong performance, several limitations still exist.

1) Dataset dependency

The results are specific to the dataset used. Future work will validate the model on external datasets, such as e-commerce and bank transaction data, to confirm robustness across different domains and improve generalizability

2) Computational complexity

The use of XGBoost and DNN increases the computational overhead, which can be a problem to be faced within resource-constrained environments.

3) Real-world constraints

Factors such as data privacy regulations and the need for real-time processing may necessitate further optimization of the model. When implementing the proposed framework in production settings, new challenges arise, including latency constraints, privacy regulations (e.g., General Data Protection Regulation (GDPR)) and the need for continuous update of the model to prevent fraud attempts based on evolving techniques. In light of these, future work will concentrate on implementing privacy-preserving methods like federated learning which maintains a firmly decentralized approach to model training so as to curtail the sharing of sensitive customer data and make potential integration of drift detection methods such that the model will be retrained once the distribution on transactions varies.

J. Future Directions

The findings of this study offer a few suggestions for various promising research opportunities for the future.

1) Federated learning

The integration of federated learning can be done for further research on distributed training with data privacy.

2) Real-time deployment

Fine-tuning the model for fraud detection in real-time across dynamic environments.

3) Explainable AI

To enhance interpretability of the model for regulatory requirements and to gain confidence from stakeholders.

Financial Fraud Detection same here its a very high accurate detection but needs transparency as well. Later extensions of this work will incorporate Explainable AI (XAI) tools, e.g., SHAP and Local Interpretable Model-agnostic Explanation (LIME) that are able to explain predictions at the transaction level. Also, privacy-preserving approaches such as federated learning will be considered that allow model training to be performed collaboratively across multiple parties, without exchanging the sensitive data that they hold (fulfilling GDPR and other data protection regulations). Regarding the deployment, we will look for optimization and edge computing for a low-latency real-world application.

Real-world deployment must comply with strict data-protection regulations such as GDPR. Future work will explore privacy-preserving learning approaches, particularly federated learning—to enable model updates without sharing sensitive customer data across financial institutions. Table IV represents a comparison between this study's results and some studies from the literature review.

TABLE IV. COMPARISON BETWEEN RESULTS OF THIS STUDY AND SOME SELECTED STUDIES OF THE LITERATURE REVIEW BASED ON DIFFERENT PARAMETERS

Study	Accuracy	Precision	Recall	Managing Imbalanced Classes	Adaptability	Real-Time Performance	Advantage of Proposed Model
[11]	High	High	High	Limited	Low	Yes	Improved precision and wider scalability
[12]	92%	-	-	No	Medium	No	Higher accuracy and real-time adaptability
[13]	99.96%	High	High	Borderline-SMOTE	Medium	Yes	Near-Perfect metrics, stronger interpretability
[14]	Medium	Medium	Medium	No	High	No	Improved balance between interpretability and accuracy
[15]	High	High	High	No	Low	No	Superior computational efficiency
Our Proposed Model	99.7%	High	High	Yes (SMOTE)	High	Yes	Near-Perfect performance on all metrics

VI. CONCLUSION

The proposed hybrid model shows exceptionally strong performance in detecting financial fraud, achieving near-near-perfect metrics across all evaluation criteria, with cross-validation averaging 99.7% accuracy, precision, recall, and F1-score. It addresses issues of data imbalance, scalability, and adaptability to evolving fraud patterns through the integration of XGBoost and Deep Neural Networks into a Stacking Ensemble framework.

This study also points out the importance of advanced preprocessing techniques, such as SMOTE, which ensures that the minority class-fraudulent transactions-are well represented during training.

These findings contribute significantly to the field of financial fraud detection by providing a scalable and robust solution that maintains high accuracy, strong interpretability, and stable generalization. This is particularly critical in financial systems where real-time fraud detection is indispensable to prevent losses and ensure compliance with regulatory requirements. Moreover, the comparison with other studies brings out the relative strengths of the proposed approach in handling data complexity with good performance on all metrics. Although these results are promising, some limitations still exist, such as using a single dataset and the computational complexity of the hybrid approach. Testing the model on larger and more diverse datasets and further optimization for resource-constrained environments might be a possible future work. This can further be combined with either federated learning or explainable AI techniques to enhance its applicability in real-world scenarios, considering concerns related to data privacy and regulatory compliance. Although the model was developed and tested on a single dataset, future work will focus on validating its performance on additional datasets such as e-commerce or bank transaction data to improve external validity and ensure broader generalizability.

In effect, this proposed hybrid model represents a promising advancement in the area of financial fraud detection, as it offers a highly valid, scalable, and adaptable solution. This study provides a useful benchmark for future research. For further research in this area while opening doors for further research to ensure security and resilience against emerging threats in financial systems.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Noor Amer Ahmed conducted the research, performed the experiments and simulations, analyzed the results, and drafted the original manuscript. Fadi Al-Turjman supervised the study, contributed to the research design and methodology, reviewed and edited the manuscript, and provided overall guidance throughout the research process. All authors have read and approved the final version of the manuscript.

REFERENCES

[1] T. O. Sanyaolu, A. G. Adeleke, C. F. Azubuko *et al.*, "Exploring fintech innovations and their potential to transform the future of

financial services and banking," *International Journal of Scholarly Research in Science and Technology*, vol. 5, no. 1, pp. 54–73, 2024.
[2] L. S. Daggubati, "Designing digital payment experiences: The crucial role of user-centered design and effective user feedback integration," *International Journal of Computer Trends and Technology*, vol. 72, no. 2, pp. 27–29, 2024.
[3] R. Chaudhry, S. Kaur, J. Singla *et al.*, "Fraud detection and prevention for a secure financial future using artificial intelligence," in *Proc. 2024 International Conf. on Emerging Smart Computing and Informatics (ESCI)*, 2024, pp. 1–6.
[4] A. C. Hiremath, A. Arya, L. Sriranga *et al.*, "Ensemble of graph neural networks for enhanced financial fraud detection," in *Proc. 2024 IEEE 9th International Conf. for Convergence in Technology*, 2024, pp. 1–8.
[5] S. Motie and B. Raahemi, "Financial fraud detection using graph neural networks: A systematic review," *Expert Systems with Applications*, vol. 240, 122156, 2024.
[6] C. Yu, Y. Xu, J. Cao *et al.*, "Credit card fraud detection using advanced transformer model," in *Proc. 2024 IEEE International Conf. on Metaverse Computing, Networking, and Applications (MetaCom)*, 2024, pp. 343–350.
[7] L. Guo, R. Song, J. Wu *et al.*, "Integrating a machine learning-driven fraud detection system based on a risk management framework," *Applied and Computational Engineering*, vol. 87, pp. 80–86, 2024.
[8] S. Pahari, A. Polisetty, S. Sharma *et al.*, "Adoption of AI in the banking industry: A case study on Indian banks," *Indian Journal of Marketing*, vol. 53, no. 3, pp. 26–41, 2023.
[9] I. D. Mienye and N. Jere, "Deep learning for credit card fraud detection: A review of algorithms, challenges, and solutions," *IEEE Access*, vol. 2, pp. 96893–96910, 2024.
[10] A. Gandhar, K. Gupta, A. K. Pandey *et al.*, "Fraud detection using machine learning and deep learning," *SN Comput. Sci.*, vol. 5, 453, 2024.
[11] B. Vyas, "Java in action: AI for fraud detection and prevention," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, vol. 9, no. 6, pp. 58–69, 2023.
[12] I. Albert-Sogules, T. O. Sonubi, P. F. Azuikpe *et al.*, "Design of an intelligent financial surveillance system using big data analytics for enhanced fraud detection and prevention in financial institutions," *International Journal of Science and Research Archive*, vol. 12, no. 02, pp. 2295–2306, 2024.
[13] R. M. Own, S. A. Salem, and A. E. Mohamed, "TCCFD: An efficient tree-based framework for credit card fraud detection," in *Proc. 2021 16th International Conf. on Computer Engineering and Systems (ICCES)*, 2021, pp. 1–6.
[14] J. Chaquet-Ulldemolins, F. Gimeno-Blanes, S. Moral-Rubio *et al.*, "On the black-box challenge for fraud detection using machine learning: Nonlinear analysis through interpretable autoencoders," *Applied Sciences*, vol. 12, no. 8, 3856, 2022.
[15] P. O. Shoetan and B. T. Familoni, "Transforming fintech fraud detection with advanced artificial intelligence algorithms," *Finance & Accounting Research Journal*, vol. 6, no. 4, pp. 602–625, 2024.
[16] M. K. A. Ismaeil, "Harnessing AI for next-generation financial fraud detection: A data-driven revolution," *Journal of Ecohumanism*, vol. 3, no. 7, pp. 811–821, 2024.
[17] C. T. Chen, C. Lee, S. H. Huang *et al.*, "Credit card fraud detection via intelligent sampling and self-supervised learning," *ACM Transactions on Intelligent Systems and Technology*, vol. 15, no. 2, 35, 2024.
[18] O. A. Bello and K. Olufemi, "Artificial intelligence in fraud prevention: Exploring techniques, applications, challenges, and opportunities," *Computer Science & IT Research Journal*, vol. 5, no. 6, pp.1505–1520, 2024.
[19] M. M. Ismail and M. A. Haq, "Enhancing enterprise financial fraud detection using machine learning," *Engineering, Technology & Applied Science Research*, vol. 14, no. 4, pp. 14854–14861, 2024.
[20] O. M. Ijiga, I. P. Idoko, G. I. Ebieg *et al.*, "Harnessing adversarial machine learning for advanced threat detection: AI-driven strategies in cybersecurity risk assessment and fraud prevention," *Open Access Research Journal of Science and Technology*, vol. 11, no. 1, pp. 1–24, 2024.
[21] I. Yuhertiana and A. Amin, "Artificial intelligence driven approaches for financial fraud detection: A systematic literature review," *KnE Social Sciences*, vol. 9, no. 20, pp. 448–468, 2024.
[22] E. Ellahi, M. Talha, D. Vidhate *et al.*, "Fraud detection and prevention in finance: Leveraging artificial intelligence and big data," *Dandao Xuebao/Journal of Ballistics*, vol. 36, no. 1, pp. 54–62, 2024.
[23] F. T. Johora, R. Hasan, S. F. Farabi *et al.*, "AI-powered fraud detection in banking: Safeguarding financial transactions," *The American Journal of Management and Economics Innovations*, vol. 6, no. 06, pp. 8–22, 2024.
[24] S. Dubey, "Artificial intelligence in financial fraud detection: A case study of Indian banking sector," *Innovative Research Thoughts*, vol. 8, no. 4, pp. 689–695 2022.
[25] P. Kamuangu, "A review on financial fraud detection using AI and machine learning," *Journal of Economics, Finance and Accounting Studies*, vol. 6, no. 1, pp. 67–77, 2024.
[26] Kaggle. Credit card fraud detection. *Kaggle Datasets*. [Online]. Available: https://www.kaggle.com/mlg-ulb/creditcardfraud