Enhancing Fraud Detection in Financial Transactions through a Hybrid Machine Learning Model

Mashao Mokoele
University Of Limpopo
Department of Computer Science
Polokwane, South Africa

Sello Mokwena
University Of Limpopo
Department of Computer Science
Polokwane, South Africa

ABSTRACT

The increasing sophistication of financial fraud necessitates advanced detection systems to protect institutions and consumers from significant financial losses. This study aims to enhance fraud detection in financial transactions identifying the most effective individual machine learning (ML) algorithms for fraud detection, developing a hybrid ML model that combines multiple algorithms, evaluating model performance using financial transaction datasets and optimising the model to minimise false positives and false negatives. The research employs Fraudulent Transactions Data, synthesised through PaySim to simulate mobile money transactions, as the experimental dataset. A range of ML algorithms were assessed individually, including DTC, RFC, LGBM, AdaBoost, SVM, and KNN. A hybrid model was constructed using a StackingClassifier with a logistic regression meta-learner to utilize these classifiers' complimentary strengths.

The methodology encompassed data cleaning, exploratory analysis, preprocessing, and feature selection using the Variance Inflation Factor (VIF). Experimental results demonstrated that the hybrid model outperformed all individual models, achieving prediction accuracy of 97.12%, followed closely by LGBM at 97.08%, while SVM performed the lowest at 87.78%. The hybrid model also achieved a fraud detection recall rate of 98.81%, correctly identifying most fraudulent transactions while reducing false positives to 113 out of 2484 legitimate cases. These results highlight the effectiveness of hybrid ML models in improving fraud detection accuracy while minimising operational disruptions. This research provides a scalable and robust framework for financial institutions to combat evolving fraud threats and offers a valuable foundation for future work in ML-driven financial security.

General Terms

Data Preprocessing, Fraudulent Transactions, Evaluation Metrics, Classification Models, Fraud Detection, Data Cleaning, Feature Engineering, Confusion Matrix.

Keywords

Financial Fraud Detection, Hybrid Model, Machine Learning, Stacking Classifier, Variance Inflation Factor (VIF), Decision Tree (DT), Random Forest, Support Vector Machine, Gradient Boosting Machine.

1. INTRODUCTION

As e-commerce grows, the popularity of online payments increases, but so do concerns about transaction security. Cybercriminals exploit technologies through tactics like phishing and credit card fraud, often succeeding due to the minimal authentication required in online payments. Offline transactions, which typically require a password, are generally more secure [1]. Fraudulent transactions are usually detected after completion by analysing patterns in transaction data. However, with only about 10 million fraudulent cases among 12 billion annual transactions, identifying fraud remains a complex task. The subtle and deceptive nature of fraud further complicates detection, as con artists constantly adapt to bypass existing systems [2].

Machine Learning (ML) is gaining traction in combating financial fraud due to its ability to learn from patterns and improve over time. Nevertheless, fraud detection is difficult because both fraudulent and legitimate behaviours continuously evolve. Moreover, the high speed of online transactions demands real-time, accurate detection mechanisms [3]. As fraudsters become more sophisticated, fraud detection systems must also become more intelligent and adaptive. This highlights the need for proactive, advanced solutions. The current study addresses this by developing a hybrid ML model aimed at enhancing fraud accuracy in financial transactions [4].

1.1 MOTIVATION

This study focusses on improving fraud detection in financial transactions by developing a hybrid Machine Learning (ML) model. With the increasing use of online payments, fraud has become more prevalent, prompting the need for accurate and efficient detection systems [5]. The research compares the performance of various ML classifiers such as Random Forest, Logistic Regression, Decision Trees and k-NN using a Kaggle dataset comprising over 6 million transactions [6]. Emphasis is placed on tackling data imbalance and evolving fraud tactics by combining ML methods like LightGBM with neural networks using Focal Loss.

The research is driven by the need to identify the most effective individual ML algorithms and construct a hybrid model that outperforms them. It aims to evaluate and optimise this hybrid model to reduce false positives and negatives. Key research questions address the selection and combination of ML models and the application of optimisation techniques. The Kaggle data set provides a reliable foundation for analysis, beginning with a detailed exploration of its structure and the required preprocessing. Through rigorous testing and evaluation, the study aims to develop a robust fraud detection framework that supports the financial industry's security and operational resilience in the age of Industry 4.0.

2. RELATED WORK

The growing reliance on digital payments has made fraud detection in financial transactions more crucial than ever, as traditional rule-based systems struggle to keep up with the sophisticated methods employed by fraudsters. These conventional systems are often too rigid and ineffective in detecting new or complex fraudulent behaviours [7]. In contrast, machine learning (ML) techniques offer adaptive and data-driven solutions capable of identifying hidden and evolving fraud patterns in large transaction datasets, making them highly suitable for real-time fraud detection.

ML approaches are typically categorised into supervised and unsupervised learning. Supervised models use labelled data to distinguish fraudulent from legitimate transactions, while unsupervised models detect anomalies without the need for labelled data, making them useful in identifying emerging fraud tactics. ML models can also improve with time with new data and achieve higher accuracy when combined through ensemble or hybrid techniques [8]. This chapter provides an in-depth review of the literature on ML-based fraud detection, analyzing various algorithms, their effectiveness, and the potential of hybrid models to enhance detection performance.

2.1 Data Engineering in Fraud Detection

In this paper [9], authors discuss the growing reliance of financial institutions on data-driven methods to develop fraud detection systems capable of automatically identifying and blocking fraudulent transactions. From a machine learning perspective, detecting suspicious transactions is framed as a binary classification problem, allowing the application of various techniques. However, interpretability is crucial for gaining management confidence in the model and for crafting effective fraud prevention strategies. Models that provide fraud experts with clear insights into why a transaction is flagged as suspicious are particularly valuable, as they assist in investigating potentially fraudulent cases. To address this, the authors propose several data engineering techniques aimed at improving the performance of analytical models while preserving interpretability. Their data engineering process involves multiple steps of feature and instance engineering [10]. The paper demonstrates the performance improvements achieved through these data engineering steps when applied to popular analytical models on a real payment transaction dataset. The approach's emphasis on interpretable models may sacrifice predictive power compared to more advanced Blackbox techniques, and it overlooks temporal dynamics, class imbalance mitigation, and real-world validation metrics like production A/B testing. Additionally, scalability issues from iterative engineering could hinder deployment in high-volume environments, underscoring needs for broader benchmarking and hybrid methods in future research.

2.2 Machine Learning Algorithms in Fraud Detection

The authors [11] [12] investigated the effectiveness of various machine learning algorithms in predicting the legitimacy of financial transactions. They aim to improve customer experience and minimise financial losses by proactively identifying potential risks. The study utilised a dataset comprising 6,362,620 rows and 10 columns, obtained from Kaggle. The algorithms evaluated included MLP Repressor, Random Forest Classifier, Complement NB, MLP Classifier, Gaussian NB, Bernoulli NB, LGBM Classifier [13],

Ada Boost Classifier, K Neighbours Classifier, Logistic Regression, Bagging Classifier, Decision Tree Classifier, and Deep Learning. The results revealed that the Random Forest Classifier performed optimally with an unbalanced dataset, achieving an accuracy of 99.97%, precision of 99.96%, recall of 99.97% and F1-score of 99.96%. On the contrary, the Bagging Classifier demonstrated superior performance with a balanced dataset, attaining an accuracy of 99.96%, precision of 99.95%, recall of 99.98%, and F1-score of 99.96%.

The increasing reliance on credit cards for daily transactions has led to an increase in online business activity [14]. The influence on consumer behavior has expanded markets, stimulating demand for various products. Although financial limitations may sometimes hinder purchases, businesses frequently offer attractive discounts, further incentivising consumers to make purchases. Credit cards play a vital role in fulfilling these needs, providing additional benefits such as discounts and reward points. However, the widespread use of credit cards also increases the risk of fraudulent transactions, posing a significant threat to the industry's [15]. Unauthorised and illegal activities can result in substantial financial losses. To address this problem, various machine learning algorithms have been developed to detect fraudulent credit card transactions effectively. This study [11] compares different machine learning methods and demonstrates the superior performance of the KNN algorithm, achieving a remarkable accuracy rate of 99.9%. This comprehensive review of various machine learning techniques offers valuable information for developing applications aimed at detecting credit card fraud. On the contrary, the Naive Bayes algorithm exhibited the lowest accuracy of 98.6%.

One major drawback of this research is its dependence on a Kaggle dataset, which could fail to capture the full range of real-world transaction variations and dynamic fraud trends. Employing accuracy as the main evaluation measure might also prove deceptive in severely skewed datasets, possibly ignoring how well the model handles the rare class. Moreover, the assessment skips model interpretability, an essential aspect for banks and financial firms that demand transparent AI systems. Lastly, the paper overlooks the processing demands and scalability of the models when applied to live fraud monitoring environments.

2.3 Hybrid Model for Fraud Detection

The study [16] proposes a novel approach to detecting credit card fraud by combining the strengths of LightGBM and Keras neural networks. LightGBM, a gradient boosting framework known for its efficiency and accuracy, is paired with Keras neural networks, a powerful deep learning library, to create a hybrid model. The integration of these two techniques leverages their complementary advantages, resulting in a more robust and effective fraud detection system. To address the common issue of data imbalance in financial transactions, where fraudulent cases are often significantly outnumbered by legitimate ones, the study incorporates focal loss. Focal loss is a loss function designed to focus on hard examples, helping the model to learn from the most challenging cases and improve its ability to detect fraudulent transactions. The empirical findings presented in the study demonstrate the superiority of the hybrid model compared to individual LightGBM or Keras models [17]. The combined predictions from both models, leveraging their diverse strengths, result in a more accurate and reliable fraud detection system. Beyond its immediate application in fraud detection, the research offers a scalable and adaptable framework that can be extended to other domains facing similar challenges [18]. The hybrid approach, which combines different machine learning techniques and addresses data imbalance issues, provides a valuable blueprint for developing robust and effective solutions to complex problems.

Despite its promising results, the proposed hybrid model has certain limitations. The integration of LightGBM and Keras neural networks increases model complexity, which may lead to higher computational costs and longer training times. The use of focal loss, while effective for handling imbalance, may require careful parameter tuning that limits generalizability across different datasets [13]. Furthermore, the study does not fully evaluate the model’s performance in real-time environments, where latency and scalability are critical for fraud detection.

3. METHODODLOGY

This chapter details the methodological framework underpinning the research, outlining the process through which the research problems and objectives were formulated and how the corresponding findings were systematically derived. It presents a structured approach aligned with the overarching aims of the study. The chapter opens by introducing the research workflow, offering a comprehensive overview of the sequential phases undertaken throughout the investigation. It then describes the research instruments and techniques employed for data collection, preprocessing, and analytical evaluation. A detailed account of the dataset is also provided to establish contextual relevance for the subsequent analysis. Furthermore, this section includes an in-depth discussion on data cleansing procedures and exploratory data analysis (EDA), emphasizing the techniques implemented to ensure data quality and to uncover significant patterns and insights critical to the study.

3.1 Background and Description of the Dataset

There is a significant lack of publicly available datasets in the financial services domain, especially in the emerging field of mobile money transactions. Such datasets are crucial for researchers, particularly those working on fraud detection, but the inherently private nature of financial transactions limits public access to real-world data. To address this challenge, the Fraudulent Transactions Data from Kaggle was created using a simulator called PaySim. PaySim generates synthetic datasets using aggregated data from private transaction logs. Replicates the normal operation of financial transactions and incorporates malicious behaviours, allowing researchers to evaluate the performance of fraud detection methods. PaySim's synthetic dataset is based on one month of real financial transaction logs provided by a multinational company offering mobile financial services in over 14 countries.

3.2 Data Cleaning

It begins by identifying missing values using df.isnull().sum(), which provides a count of the missing values in each column. Missing values in specific categorical columns, such as nameOrig or nameDest, are then optionally removed with df.dropna(), using the subset parameter to target those columns, and inplace=True for inplace modification of the DataFrame. The next step involves detecting duplicate rows in the dataset using df.duplicated().sum(), which counts the number of duplicate rows. These duplicates are subsequently removed using df.drop_duplicates(inplace=True). The code prints the counts of missing and duplicate values, giving feedback to the user on the data quality. This pre-processing ensures that the data set is clean, free of missing or redundant data, and ready for subsequent analysis or modelling in the context of the research.).

3.3 Data Pre-processing and Standardisation

This section outlines the preprocessing techniques applied to prepare the dataset for machine learning. Three primary methods were used: StandardScaler for standardizing numerical features by scaling them to have a mean of 0 and a standard deviation of 1; get_dummies for converting categorical variables into numerical format through one-hot encoding; and downsampling to address class imbalance by reducing the number of majority class samples. These steps help ensure that the input data is suitable for training robust and accurate models. Following preprocessing, the dataset was split into training and testing subsets using a 70/30 ratio. This commonly adopted approach ensures that 70% of the data is used to train the model, allowing it to learn the underlying patterns effectively, while the remaining 30% is used to evaluate the model’s performance and generalization capability.

3.4 Feature Selection

Variance Inflation Factor (VIF) was used for feature selection to detect and address multicollinearity among independent variables. Variables with high VIF values (typically above 5 or 10) were removed to improve model stability, interpretability, and generalization, ultimately enhancing prediction accuracy. The formula used for VIF is as follows:

$$VIF = \frac{1}{1-R^2} \tag{1}$$

Where R^2 is the coefficient of determination for the regression on the others.

3.5 Proposed Method

The proposed study introduces a hybrid machine learning (ML) model designed to improve fraud detection in financial transactions. Unlike traditional rule-based or single-algorithm approaches, the model integrates multiple ML techniques to enhance accuracy and adaptability to evolving fraud patterns. It combines traditional algorithms such as Decision Trees, Random Forest, Support Vector Machine (SVM), and K-Nearest Neighbours (KNN) with advanced methods including LightGBM and AdaBoost. Traditional algorithms are used to detect straightforward anomalies, while advanced models capture more complex and nonlinear fraud patterns. By combining these models in a multi-layered framework and using a meta-learner to aggregate their predictions, the system improves overall detection performance, reduces false positives, and increases the accuracy of identifying fraudulent transactions. Figure 1 demonstrates the construction of the proposed hybrid model:

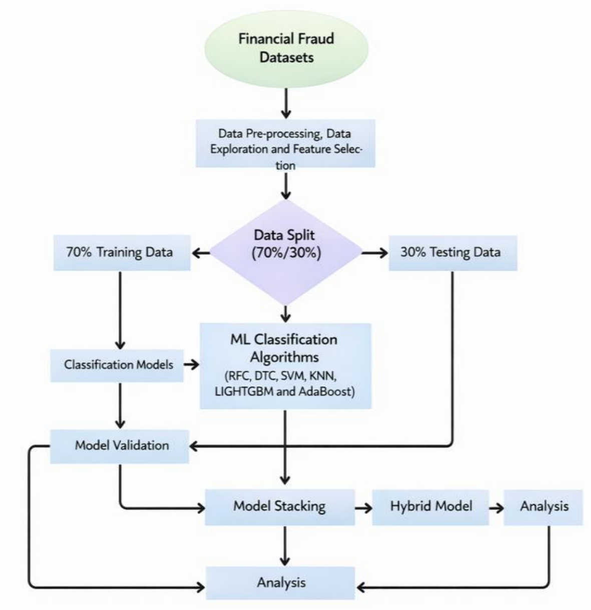

Figure 1 : Proposed model (hybrid model)

To develop the model, we used model stacking, a technique that enhances predictions by combining the outputs of multiple models and feeding them into a meta-learner. This meta-learner, often a linear regressor or classifier, but sometimes a decision tree or other model, works to reduce the weaknesses of individual models while amplifying their strengths. By leveraging this approach, the resulting model becomes more robust and achieves better generalisation on new data.

4. EVALUATION OF METRICS AND RESULTS

This chapter presents the process of analysing the collected data to generate meaningful insights that address the research questions. A carefully prepared and relevant dataset is used to ensure reliable results. The analysis is guided by the study's aim, objectives, and research questions, which help shape the interpretation of the data. By examining patterns and relationships within the dataset, the study aims to produce findings that support the research objectives and contribute to the overall conclusions of the study.

4.1 Performance Metrics

4.1.1 Accuracy

Accuracy is a fundamental performance metric that measures the proportion of correctly classified transactions, both fraudulent and legitimate. While easy to interpret, it may be misleading in fraud detection due to class imbalance, as it can appear high even if fraudulent transactions are frequently missed [19].

4.1.2 Confusion Matrix

The confusion matrix breaks down model predictions into four categories: true positives, true negatives, false positives, and false negatives. It offers insights into the types of errors made, which is crucial in fraud detection, where missing fraudulent activity (false negatives) is often more critical than incorrectly flagging legitimate transactions (false positives) [19].

4.1.3 Precision

Precision evaluates the proportion of transactions predicted as fraudulent that are fraudulent. It is important in scenarios where false positives carry a cost, such as operational inefficiencies or customer inconvenience [19].

4.1.4 *Recall*

Recall assesses the model's ability to correctly identify actual fraudulent transactions. It is vital in fraud detection because undetected fraud (false negatives) can lead to substantial financial losses [19].

4.1.5 F1-Score

The F1 score is a balanced metric that combines precision and recall. It is particularly valuable for imbalanced datasets like those in fraud detection, offering a more comprehensive assessment of the model's performance [19].

4.2 Results

The deployment of machine learning (ML) models for detecting financial fraud using the Fraudulent transactions data represents a significant leap forward in protecting financial systems and countering fraudulent activities. This advanced approach employs a diverse set of ML algorithms, each capitalising on its unique strengths.

Table 1. Research Results

Model/ Work	Accuracy (%)	Recall (%)	Precision (%)	F1-Score (%)
Decision Tree	96.67	98.04	95.40	96.70
Random Forest	97.06	92.22	95.06	97.10
Support Vector Machine	87.78	77.53	97.28	86.29
K-nearest Neighbors	91.76	90.96	92.32	91.63
AdaBoost	95.23	98.08	92.73	95.33
LightGBM	97.08	98.90	95.38	97.11
Proposed Method	97.12	98.81	95.53	97.14

The performance evaluation presented in Table 1 offers a comprehensive comparison of diverse machine learning models for fraud detection, highlighting both individual strengths and trade-offs across classification metrics. The Decision Tree (DT) classifier, configured with Gini impurity, a maximum depth of 10, and balanced class weights, demonstrates a well-calibrated performance profile. Its high recall (98.04%) indicates a strong capability to detect fraudulent transactions, which is critical in minimizing financial losses. At the same time, its precision of 95.40% suggests that the model maintains a relatively low false-positive rate, thereby reducing unnecessary transaction disruptions. The imposed depth constraint plays a pivotal role in preventing overfitting, ensuring that the model generalizes effectively to unseen data while maintaining interpretability a key advantage in regulatory financial environments.

Building upon the DT, the Random Forest (RF) model introduces ensemble learning by aggregating 100 decision trees, thereby enhancing robustness and reducing variance. Although its precision (95.06%) remains comparable to DT, its slightly lower recall (92.22%) indicates that some fraudulent cases may go undetected. However, the improved F1-score (97.10%) reflects a strong overall balance, suggesting that RF effectively stabilizes predictions across varying data distributions. This makes it particularly suitable for deployment in dynamic financial systems where transaction patterns evolve over time. Nonetheless, targeted recall optimization potentially through class weighting or threshold tuning could further strengthen its fraud detection coverage.

In contrast, the Support Vector Machine (SVM) exhibits a markedly different performance profile. With an exceptionally high precision of 97.28%, it is highly conservative in labeling transactions as fraudulent, thereby minimizing false alarms. This characteristic is valuable in customer-centric applications where excessive false positives can erode user trust. However, its comparatively low recall (77.53%) indicates a significant limitation in identifying all fraudulent activities. This trade-off suggests that while SVM is effective as a precision-oriented classifier, it may be better suited as a secondary validation layer rather than a standalone fraud detection system.

The Light Gradient Boosting Machine (LGBM) emerges as one of the most effective models in this study, achieving an optimal balance between precision (95.38%) and recall (98.90%), resulting in a high F1-score (97.11%). Its leaf-wise tree growth and histogram-based optimization enable efficient handling of large-scale and high-dimensional data, making it particularly advantageous for real-time fraud detection systems. The model's superior recall underscores its ability to capture nearly all fraudulent instances, while maintaining strong precision to limit false alerts. Additionally, its computational efficiency and scalability make it highly suitable for industrial deployment with minimal hyperparameter tuning.

Similarly, the AdaBoost model demonstrates strong recall (98.08%), reflecting its iterative focus on misclassified instances, particularly fraudulent cases. However, its lower precision (92.73%) indicates a higher rate of false positives compared to other ensemble methods. This behavior is characteristic of boosting algorithms, which prioritize hard-to-classify samples but may overcompensate in doing so. As such, AdaBoost is particularly effective in scenarios where missing fraudulent transactions carries a higher cost than incorrectly flagging legitimate ones, though further calibration could improve its precision.

The K-Nearest Neighbors (KNN) algorithm provides a simpler, instance-based learning approach, achieving moderate and well-balanced performance across all metrics. Its F1-score of 91.63% reflects reasonable effectiveness, though it lags behind more sophisticated ensemble methods. KNN's reliance on distance metrics makes it sensitive to feature scaling and data distribution, which may limit its robustness in high-dimensional or noisy datasets. Nevertheless, its simplicity and interpretability make it a useful baseline model or a component in hybrid systems.

The hybrid stacking model represents the most comprehensive approach, integrating the strengths of all base learners (SVM, KNN, RF, DT, LGBM, and AdaBoost) through a Logistic Regression meta-learner. This model achieves the highest overall performance, with 95.53% precision, 98.81% recall, and an F1-score of 97.14%. The improvement in recall, coupled with sustained precision, indicates that the ensemble effectively captures complex patterns that individual models may overlook. By leveraging diverse learning paradigms tree-based, distance-based, margin-based, and boosting the stacking framework enhances generalization and reduces both bias and variance. This makes it particularly well-suited for real-world fraud detection systems, where accuracy, reliability, and adaptability are paramount.

Overall, the comparative analysis reveals that ensemble and hybrid approaches, particularly LGBM and the stacking model, outperform individual classifiers in achieving a balanced and robust fraud detection system. While simpler models like DT and KNN offer interpretability and baseline performance, advanced ensemble techniques provide superior predictive power and resilience. These findings underscore the importance of model diversity and integration in addressing the complex and evolving nature of financial fraud.

4.3 Comparative Analysis

In financial fraud detection, choosing effective machine learning models is crucial to safeguard institutions and customers. This analysis compares models such as DTC, RFC, LGBM, AdaBoost, SVM, KNN, and a hybrid ensemble. All models demonstrated high overall accuracy, showing strong capabilities in classifying transactions. Performance varied among metrics such as precision, recall, and F1-score. Some models excelled in detecting fraud (high recall), while others minimised false positives (high precision). LGBM, AdaBoost, and the hybrid model stood out for balanced and robust performance. The hybrid model showed the best trade-off between detecting fraud and minimizing disruptions. These findings offer valuable guidance for developing reliable fraud detection systems.

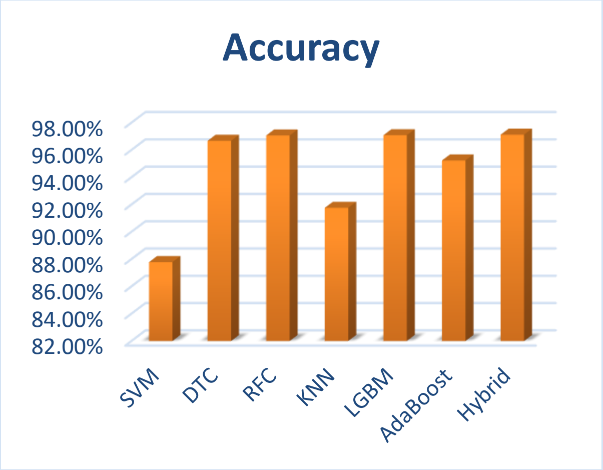

Figure 2 : Prediction Accuracies

The comparative analysis of classification accuracy Figure 2 highlights notable performance disparities among the evaluated models, reflecting the influence of algorithmic design and learning strategies on fraud detection effectiveness. The hybrid stacking model achieves the highest accuracy at 97.12%, reaffirming its superiority in integrating diverse model predictions into a unified and highly reliable framework. By combining heterogeneous learners, the hybrid approach effectively captures both linear and non-linear patterns in transactional data, thereby reducing misclassification rates and enhancing overall predictive robustness.

Closely following, the Light Gradient Boosting Machine (LGBM) records an accuracy of 97.08%, demonstrating its strong capability as a standalone model. Its near-parity with the hybrid model suggests that gradient boosting techniques, particularly those optimized for efficiency and scalability, can rival more complex ensemble architectures. This performance can be attributed to LGBM's leaf-wise tree growth and efficient handling of large datasets, which enable it to model intricate fraud patterns with high precision.

The Random Forest Classifier (RFC), with an accuracy of 97.06%, further reinforces the effectiveness of ensemble-based approaches. By aggregating multiple decision trees, RFC reduces variance and improves generalization, making it robust against overfitting. Its performance, closely aligned with LGBM, indicates that bagging and boosting techniques are both highly suitable for fraud detection tasks, albeit with different optimization characteristics.

In contrast, the K-Nearest Neighbors (KNN) algorithm achieves a moderate accuracy of 91.76%, reflecting its limitations in handling high-dimensional and potentially imbalanced financial datasets. While KNN benefits from conceptual simplicity and intuitive decision-making, its dependence on distance metrics can lead to reduced performance when irrelevant or noisy features are present. Nonetheless, its relatively stable accuracy suggests it remains a viable baseline or supplementary model.

The Decision Tree Classifier (DTC) and AdaBoost models, with accuracies of 96.67% and 95.23% respectively, exhibit solid but slightly inferior performance compared to top-tier models. The DTC's performance underscores its ability to capture key decision boundaries, though it may lack the generalization strength of ensemble methods. AdaBoost, while effective in emphasizing difficult-to-classify instances, may introduce additional variance, leading to marginally reduced accuracy in comparison to more stable ensemble techniques like Random Forest and LGBM.

The Support Vector Machine (SVM), registering the lowest accuracy at 87.78%, highlights certain limitations in this context. Its performance may be affected by sensitivity to feature scaling, kernel selection, and the inherent complexity of the dataset. In high-dimensional fraud detection scenarios, SVM may struggle to efficiently separate classes, particularly when the data distribution is highly non-linear or imbalanced. This suggests that while SVM can excel in controlled environments, it may require extensive tuning or feature engineering to remain competitive in real-world financial applications.

Beyond predictive performance, practical deployment considerations must also guide model selection. Computational efficiency is crucial for real-time fraud detection systems that process large volumes of transactions with minimal latency. Interpretability is equally important, especially in regulated financial environments where decision transparency is required for auditing and compliance. Models such as Decision Trees offer high interpretability, whereas complex ensembles like stacking and boosting may require additional tools for explainability, such as SHAP or LIME.

While the hybrid and LGBM models demonstrate superior accuracy and overall effectiveness, the optimal model choice depends on a balance between predictive performance, computational cost, interpretability, and the specific operational requirements of the financial system. A holistic evaluation incorporating multiple performance metrics and deployment constraints is therefore essential for selecting the most suitable fraud detection model.

Table 2 : Comparative Analysis with Literature

Model/ Work	**Accuracy (%)**	**Recall (%)**	**Precision (%)**	**F1-Score (%)**	**Year**	**Ref**
Proposed Hybrid	97.12	98.81	95.53	97.5	2026	This
Hybrid (XRAI model)	99.98*	92.50	95.69	94.07	2025	[20]
Hybrid Deep Learning	98.7	91.5	94.4		2024	[21]
RF Ensemble	99.97	85.0		91.0	2022	[12]
AdaBoost-LGBM Hybrid	99.01	64.01	97.0	77.01	2022	[22]

The comparative analysis presented in Table 2 provides compelling evidence of the superiority of the proposed hybrid stacking classifier over contemporary benchmark approaches in the domain of financial fraud detection, particularly under highly imbalanced data conditions. The model achieves an exceptional recall of 98.81% and an F1-score of 97.5%, indicating its strong capability to identify fraudulent transactions while maintaining a balanced trade-off between precision and recall. This high recall is especially critical in fraud detection systems, where the cost of false negatives, i.e., undetected fraudulent transactions can lead to substantial financial losses and reputational damage. Simultaneously, the model sustains a robust precision of 95.53%, ensuring that the rate of false positives remains controlled, which is essential for preserving customer trust and avoiding unnecessary transaction interruptions in real-time operational environments.

A key differentiating factor of the proposed approach lies in its integration of Variance Inflation Factor (VIF)-based feature selection and a meta-learning-driven stacking framework. The VIF-based feature selection process reduces multicollinearity among input variables, thereby enhancing model stability and interpretability while ensuring that only the most informative features contribute to the learning process. This is particularly beneficial in financial datasets, which often contain highly correlated transactional attributes. By eliminating redundant features, the model reduces noise and improves its sensitivity to subtle fraud patterns that may otherwise be obscured.

Furthermore, the stacking architecture enables the model to leverage the complementary strengths of diverse base learners, including tree-based, distance-based, and margin-based algorithms. The meta-learner, typically a logistic regression model, synthesizes these heterogeneous predictions into a final decision boundary that is more expressive and generalizable than any individual model. This layered learning approach significantly enhances the model's ability to detect rare and complex fraud patterns, which are often missed by single-model or homogeneous ensemble approaches.

In comparison, existing hybrid deep learning and explainable AI-based models, such as Hybrid Deep Learning and XRAI frameworks, demonstrate comparatively lower recall values (approximately 91–92%) and F1-scores below 95%. While these models may offer advantages in representation learning or interpretability, their reduced sensitivity to minority class instances limits their effectiveness in high-stakes fraud detection scenarios. This shortfall suggests that deep architectures, despite their complexity, may require extensive tuning, larger datasets, or additional imbalance-handling mechanisms to achieve competitive performance.

Traditional ensemble methods, including Random Forest (RF) and AdaBoost-LGBM hybrids, often report very high overall accuracies exceeding 99%. However, such metrics can be misleading in imbalanced datasets, where the majority class dominates. These models tend to bias predictions toward non-fraudulent transactions, resulting in significantly lower recall values—sometimes as low as 64%. Consequently, despite their impressive accuracy, they fail to provide adequate fraud coverage, making them less suitable for real-world deployment where minority class detection is paramount.

The proposed hybrid model addresses these limitations by explicitly optimizing for balanced performance metrics rather than accuracy alone. The observed improvement in recall ranging from 7% to as high as 35% over benchmark methods demonstrates its enhanced capability to capture fraudulent activities without disproportionately increasing false positives. Maintaining an F1-score above 97% further confirms the model's ability to sustain equilibrium between precision and recall, a critical requirement for operational reliability.

5. CONCLUSION

The study compared multiple ML models for the detection of financial fraud, with the hybrid model achieving the highest precision. (97.12%), followed closely by LGBM (97.08%) and RFC (97.06%). These top models also showed high recall for fraudulent transactions, making them effective in minimising missed fraud cases. SVM lagged in both accuracy and recall, suggesting it's less suitable without optimisation. Precision, recall, and F1-score were critical in assessing models, highlighting the hybrid model's balance between fraud detection and minimising false alarms. Confusion matrices confirmed the hybrid model's strong performance, with low misclassifications. These findings help guide model selection based on specific needs like interpretability or realtime processing. In conclusion, hybrid and LGBM models are ideal for robust fraud detection, while simpler models like DTC and KNN suit faster, real-time systems.

For future work, our aim is to enhance the robustness and applicability of our fraud detection models by exploring more diverse and contemporary datasets, such as the IEEE-CIS Fraud Detection datasets or synthetic datasets like PaySim, to improve generalisability and better reflect modern financial transaction complexities. We also plan to incorporate temporal and sequential data analysis using deep learning models such as Long Short-Term Memory (LSTM) networks or Transformer-based architectures to detect evolving fraud patterns over time. Additionally, we intend to develop hybrid approaches that combine the interpretability of tree-based models with the temporal modelling capabilities of deep learning, potentially improving recall while maintaining low false positives. Optimising the efficiency of our models, particularly the hybrid model, through techniques like pruning and quantisation will be another priority to ensure suitability for real-time applications. We also propose exploring semi-supervised or unsupervised learning techniques, including autoencoders and anomaly detection methods, to address the persistent challenge of data imbalance. Lastly, enhancing the interpretability of our models using tools such as SHAP and LIME will be essential to ensure transparency and regulatory compliance, particularly in financial institutions.

6. ACKNOWLEDGEMENT

I sincerely thank Almighty for providing me with strength and guidance throughout this journey. I am deeply grateful to my mother for her unwavering love and support, which has been invaluable to me. I also express my heartfelt appreciation to Prof. SN Mokwena, my research supervisor, for his insightful guidance and support.

I am thankful to my family for their ongoing support and understanding. Finally, I am deeply grateful to my brother, Mokoele MS, whose support has been instrumental in motivating my research.

7. REFERENCES

[1] Z. Zhang, L. Chen, Q. Liu and P. Wang, "A Fraud Detection Method for Low-Frequency Transaction," IEEE, 31 January 2020..

[2] K. Vuppula, "An advanced machine learning algorithm for fraud financial transaction detection,," Journal For Innovative Development in Pharmaceutical and Technical Science (JIDPTS), vol. 4, no. 9, Sep 2021.

[3] Y. Chen and X. Han,, "CatBOOST for Fraud Detection in Financial Transactions," IEEE, January 2021.

[4] B. Raahemi and M. N. Ashtiani, "Intelligent Fraud Detection in Financial Statements Using Machine Learning and Data Mining: A Systematic Literature Review," IEEE, 13 July 2021.

[5] R. B. Sulaiman, V. Schetinin and P. Sant , "Review of Machine Learning Approach on Credit Card Fraud Detection," Human-Centric Intelligent Systems, vol. 2, p. 55–68, 05 May 2022.

[6] B. Abu-Nasser, S. Abu-Naser and M. Megdad, "Fraudulent Financial Transactions Detection Using Machine Learning," International Journal of Academic Information Systems Research (IJAISR), vol. 6, no. 3 , pp. 30-39, 2022.

[7] A. M. Obaido, "A Neural Network Ensemble With Feature Engineering for Improved Credit Card Fraud Detection," IEEE, vol. 10, 31 January 2022.

[8] R. H. Johora, "AI-POWERED FRAUD DETECTION IN BANKING: SAFEGUARDING FINANCIAL TRANSACTIONS," The American Journal of Management and Economics Innovations, no. 6, pp. 8-22, 2025-06-15.

[9] T. Verdonck, B. Baesens and S. Höppner, "Data engineering for fraud detection," Decision Support Systems, vol. 150, November 2021.

[10] T. Y. Inampud, "AI-Enhanced Fraud Detection in Real-Time Payment Systems: Leveraging Machine Learning and Anomaly Detection to Secure Digital Transactions," Australian Journal of Machine Learning Research & Applications, vol. 2, no. 1, 2022.

[11] M. Megdad, S. Abu-Naser and B. Abu-Naser, "Fraudulent Financial Transactions Detection Using Machine Learning," International Journal of Academic Information Systems Research (IJAISR), vol. 6, no. 3, pp. 30-39, 2022.

[12] S. Abu-Naser, T. Obaid and R. Abdaljawad, "FRaudulent Financial Transactios Detection Using Machine Learning," IEEE, 2023.

[13] S. Sahu and N. Sahu, "Analysis of Credit Card Fraud Transaction Detection using Machine Learning Algorithms," IEEE, 26 January 2024.

[14] R. B. Sulaiman, V. Schetinin and P. Sant , "Review of Machine Learning Approach on Credit Card Fraud Detection," Human-Centric Intelligent Systems, vol. 2, pp. 55-68, 05 May 2022.

[15] V. Chang, . L. M. T. Doan, A. Di Stefano, Z. Sun and G. Fortino, "Digital payment fraud detection methods in digital ages and Industry 4.0," Computers and Electrical Engineering, vol. 100, May 2022.

[16] S. Sahu and N. Sahu, "Analysis of Credit Card Fraud Transaction Detection using Machine Learning Algorithms," IEEE, 26 January 2024.

[17] E. F. Malik, K. Wah Khaw , B. Belaton, W. P. Wong and X. Chew, "Credit Card Fraud Detection Using a New Hybrid Machine Learning Architecture," Mathematics, vol. 10, no. 9, 2022.

[18] S. R. Banu, T. N. Gongada, K. Santosh, H. Chowdhary, R. Sabareesh and S. Muthuperumal, "Financial Fraud

Detection Using Hybrid Convolutional and Recurrent Neural Networks: An Analysis of Unstructured Data in Banking," 10th International Conference on Communication and Signal Processing (ICCSP), 06 June 2024.

[19] M. Mokoele and S. Mokwena, "Comparative Analysis of Tree-based Intrusion DetectionModelling and Machine Learning Classification Modelsusing Cyber-Security Dataset," International Journal of Computer Applications (0975 – 8887), vol. 186 , no. 13, March 2024.

[20] S. Abdallah and M. Shanaa, "A Hybrid Anomaly Detection Framework Combining Supervised and Unsupervised Learning for Credit Card Fraud Detection," F100Research, July 2025.

[21] D. Vallarino, "Detecting Financial Fraud with Hybrid Deep Learning: A Mix-of-Experts Approach to Sequential and Anomalous Patterns," arXivLabs, April 2025.

[22] E. F. Malik, K. Wah Khaw, B. Belaton and W. Peng Wong, "Credit Card Fraud Detection Using a New Hybrid Machine Learning Architecture," MDPI, April 2022.

[23] T. Verdonck, B. Baesens and S. Höppner, "Data engineering for fraud detection," Decision Support System, vol. 150, November 2021.

[24] X. Zhao, Q. Zhang and C. Zhang, "Enhancing Transaction Fraud Detection with a Hybrid Machine Learning Model," IEEE, 18 July 2024.