SCIENTIFIC CULTURE
www.sci-cult.com

SCIENTIFIC CULTURE, Vol. 12, No. 2.1, (2026), pp. 271-290
Open Access. Online & Print

DOI: 10.5281/zenodo.122.12624

A HYBRID MACHINE-LEARNING FRAMEWORK FOR FRAUD DETECTION IN MOBILE BANKING USING BEHAVIORAL BIOMETRICS AND TRANSACTIONAL PATTERNS

Md Tuhin Rana[1], Shuvashish Roy[2], Ashim Sen Gupta[3], Nadia Mehjabeen Oyshi[4], Rokhshana Parveen[5], Dipankar Das[6] and Abhigyan Bhattacharjee[7]

[1]Student, Department of Statistics, University of Dhaka, Bangladesh, mdtuhin-2016913783@stat.du.ac.bd
[2]Senior Researcher, Research & Innovation Division, Prime Bank PLC, Dhaka, Bangladesh, shuvashishroy@gmail.com
[3]First Assistant Vice President, International Division, Social Islami Bank PLC, Bangladesh, asgcubd@gmail.com
[4]Student, Department of Statistics, University of Dhaka, Bangladesh, nadiamehjabeen-2017413967@stat.du.ac.bd
[5]Research Scholar, Dhaka, Bangladesh, rokhshana2006@gmail.com
[6]Assistant Professor, Department of Commerce, University of Science & Technology, Meghalaya, Ri-Bhoi, Meghalaya, India, dipankardas.dds@gmail.com
[7]Professor, Department of Management, North-Eastern Hill University, Tura Campus, India abhigyan.nehu@gmail.com

Received: 01/12/2025
Accepted: 02/01/2026

Corresponding Author: Md Tuhin Rana (mdtuhin-2016913783@stat.du.ac.bd)

ABSTRACT

The proliferation of mobile banking has been accompanied by a surge in sophisticated financial fraud, necessitating detection systems that go beyond traditional methods. This paper designs and validates a multi-faceted, hybrid machine learning framework that synergizes behavioral biometrics (e.g., typing speed, swipe patterns) with transactional data (e.g., amount, geolocation) for high-accuracy fraud detection. We evaluate the progression of models, demonstrating that while unsupervised autoencoders are effective at profiling normal behavior, they fail to detect over 65% of fraudulent activities. Supervised Long Short-Term Memory (LSTM) networks, capturing temporal sequences, significantly improve performance, achieving fraud recall rates as high as 97%. However, gradient-boosting models (LightGBM and XGBoost) yield the most balanced standalone performance, with 98% recall and 94% precision. Feature importance analysis from these models confirms that the framework's predictive power is derived from a hybrid of both behavioral and transactional features. The framework culminates in a stacked ensemble model that optimizes the precision-recall trade-off, achieving 97% accuracy, 97% fraud recall, and 95% fraud precision. This final model registers the lowest false positive rate, presenting a robust, reliable, and deployable solution that maximizes fraud capture while minimizing unnecessary friction for legitimate users.

KEYWORDS: Fraud Detection, Mobile Banking, behavioral Biometrics, Transactional Pattern.

1. INTRODUCTION

The convergence of wireless communication, smartphones, and banking infrastructure has fostered the digital payment environment that has been instinctively replacing conventional transactions by cash. As technology is rapidly developing, there has been a remarkable shift in the transaction method to cashless payment. Moreover, the Governments all over the world are actively encouraging this shift, yet developing and emerging markets are not exempt from that (Nguyen & Huynh, 2018; Namweli & Magali, 2018; Gupta et al., 2020; Hung et al., 2021; Lonkani et al., 2020; Malaquias et al., 2021; Omigie et al., 2020; Wamba et al., 2021). While, cash remains widely accepted along with familiar due to its longstanding nature, digital payments offer greater convenience by reduced time and effort in the transactions process (Hassan, et al., 2021; Shree, et al, 2021).

E-payment refers to the use of electronic networks throughtransferring money during commercial transactions (Al-Sabaawi, et al., 2023; Nguyen & Huynh, 2018; Ming-Yen Teoh et al., 2013). Masihuddin et al. (2017) depicts, it involves converting cash into digital form between buyers and sellers using electronic technologies.However, the terms e-payment methods, digital payments, and online payments are often used interchangeably, as they all refer to the use of electronic technologies to transfer money in commercial transactions (Arjunwadkar, 2018;).Electronic payment technologies provide more than just ease of use—they help businesses reach more customers, reduce cash-handling expenses, and bring informal economic activities into the formal system, leading to increased tax revenue for governments. Additionally, online payment systems offer innovative features that benefit both customers and banks by eliminating the challenges of traditional banking methods, such as the need to visit a bank for withdrawals or deposits, by reducing delays, minimizing miscommunication and so on (Al-Okaily, et al, 2024; Khando, et al, 2022; Al-Okaily et al., 2020).

The digital commerce landscape is rapidly transforming, with online transactions becoming a dominant mode of exchange due to the unparalleled convenience and global accessibility they provide. This growing dependence on electronic payment systems has simultaneously created opportunities for cybercriminals to exploit users' trust and system vulnerabilities (Vimal et al., 2021). Besides, various benefits associated with online payment systems also face notable limitations and challenges. A key concern among users is the fear of security breaches, which could lead to financial losses. Additionally, risks of fraud and cyberattacks, along with insufficient protection mechanisms, could bring decline in consumer trust and usage of digital payment technologies (Khando, et al, 2022b; Al-Qadi, 2018; Kabir, et al, 2015)

Online payment fraud encompasses a broad spectrum of deceptive tactics aimed at unlawfully acquiring funds or goods through digital platforms (Sun et al., 2021). These fraudulent actions include account hijacking, phishing attacks, synthetic identity creation, and payment card misuse, all of which capitalize on weaknesses within the digital transaction ecosystem (Strelcenia & Prakoonwit, 2023). Among the most prevalent methods is identity theft, wherein criminals obtain sensitive personal information—such as banking credentials or social identification numbers—to impersonate individuals and execute unauthorized transactions, often resulting in severe financial and reputational consequences for the victims (Pan, 2024).

Understanding these evolving fraud methodologies is crucial for developing robust countermeasures and enhancing the security posture of online transactional environments (Mehana & Pireva, 2020). Fraudsters execute online transaction fraud through various sophisticated methods, leveraging vulnerabilities in payment systems and exploiting human factors to gain unauthorized access to financial resources (Sadgali et al., 2019). The rapid expansion of e-commerce has unfortunately provided fertile ground for these illicit activities, making it imperative for financial institutions to implement automated deterrent mechanisms to safeguard against the surging tide of fraudulent credit card transactions (Forough & Momtazi, 2020). The challenge in combating these illicit activities lies in the need for real-time detection and prevention, as transactions are often processed instantaneously (Chalapathy & Chawla, 2019).Given that online transactions do not necessitate the physical presence of the cardholder, cybercriminals can readily impersonate legitimate users to conduct unauthorized transactions, thereby exacerbating the challenge of fraud detection (Manek et al., 2019). The sheer volume and complexity of data generated by digital payment systems also render traditional fraud detection methods inadequate, necessitating advanced computational approaches (Chang et al., 2024). The dynamic nature of fraudulent behavior, which constantly evolves to evade detection, presents a formidable challenge for static fraud detection systems (Carcillo et al., 2019; Rzayeva &

Malekzadeh, 2022). This necessitates the development of robust, real-time fraud detection systems capable of analyzing vast datasets for anomalies and emerging patterns (Tran, 2022). One effective approach to combatting this challenge involves employing machine learning algorithms, which can analyze vast datasets to identify fraudulent transactions by discerning subtle patterns and anomalies that human analysts might overlook (Dornadula & Geetha, 2019). These models can use historical data to distinguish between legitimate and fraudulent transactions, significantly improving detection accuracy (Bello & Olufemi, 2024).Therefore, this research aims to identify the shift underscoring the sophisticated fraud detection systems of mobile banking throughanalyzing transaction patterns as well as behavioral anomalies to identify suspicious activities in real-time.

2. LITERATURE REVIEW

The advancement of global communication and technological infrastructure has, regrettably, been accompanied by a corresponding increase in fraudulent activities, thereby the urgent need for robust fraud detection strategies gets the importance. Fraud may be addressed either through proactive prevention or detection after the occurrence of illicit activity (Alkhateeb & Maolood, 2019). The identification of fraudulent transactions commonly involves the examination of transactional data to recognize irregular or suspicious behavioral patterns (Tran, 2022). Although digital payment systems offer substantial convenience, their widespread adoption has simultaneously escalated the prevalence of online fraud, primarily due to the overlapping behavioral characteristics shared between genuine and deceptive transactions (Keskenler et al., 2021).As Fintech operates on the existing IT infrastructure, it remains vulnerable to exploitation through targeted fraudulent activities. Detecting such threats poses significant technical challenges. To address this, the industry increasingly employs Machine Learning (ML) techniques, including anomaly detection, to automatically identify suspicious patterns. ML methods such as learning algorithms, statistical models, and artificial neural networks (ANN) are used to analyze data and inform effective fraud prevention strategies (Khando, et al, 2022b).

There are many ML methods used in the literature to detect fraud transaction in online platforms-such as Logistic Regression, K-Nearest Neighbors, Decision Tree, Naïve Bayes, Random Forest, Gradient Boosting Machines, Light Gradient Boosting Machine, Extreme Gradient Boosting, and Long Short Term Model (Dileep, et al, 2021; Guezzaz, et al, 2021; Gupta, et al, 2021; Hancock & Khoshgoftaar, 2021; Itoo, et al, 2021; Jemima Jebaseeli, et al, 2021; Owolafe, et al, 2021; Mishra, K. N., & Pandey, 2021; Jan, 2021; Benchaji, et al, 2021; Vassallo, et al, 2021; Zarezadeh, et al, 2021; Alfaiz & Fati, 2022; Aslam, et al, 2022;Aziz, et al, 2022; Chang, et al., 2022; Zhang , 2022; Aburbeian & Ashqar, 2023; Ali, et al, 2023; Douiba, et al, 2023; Du, et al, 2023; Hajek, et al, 2023; Naeem, et al, 2023; Ugale & Midhunchakkaravarthy, 2023; Vishwakarma & Kesswani, 2023, Khalid, et al, 2024; Mehdary, et al, 2024).

In order to determine the best three models for detecting credit card fraud, Alfaiz et al. (2022) used nine Machine Learning Algorithms in the first stage to test their performance. In both phases, each model assessed using the F1-Score, Accuracy, Recall, Precision, and Area under the Receiver Operating Characteristic Curve (AUC), AllKNN-CatBoost was contrasted with earlier research using the same dataset and comparable methodologies. AllKNN-CatBoost did, in fact, performed better than earlier models in terms of F1-Score (87.40%), AUC (97.94%), and Recall (95.91%).Chang et al. (2022) evaluated various Machine Learning Algorithms—such as logistic regression, decision tree, k-nearest neighbours, random forest, and autoencoder—to create an effective and stable model for fraud detection platforms suitable for Industry 0.4. The Results indicated from the random forest and logistic regression surpassed other techniques as all the models that could achieve more than 96% accuracy, 81% sensitivity, and 97% specificityin most of the cases, the Area Under the Receiver Operating Curves (AUROC) values of the used model are higher than 0.9. The research done by Ileberi, et al (2022) employs a genetic algorithm-based feature selection technique to identify the most significant features for credit card fraud detection. It integrates various machine learning classifiers, such as Decision Tree, Random Forest, Logistic Regression, Artificial Neural Network, and Naive Bayes. The findings indicated that this proposed method surpassed existing systems in fraud detection performance. Additionally, the study reveals that using the genetic algorithm for feature selection enhances the accuracy of the machine learning models.

Xu and Liu (2018) applied an optimized SVM on commercial bank datasets for online credit card fraud, showing its effectiveness over other models. Mareeswari and Gunasekaran (2016) integrated SVM with spike detection for credit card fraud, outperforming prior approachesCarneiro et al. (2015)

that proposed a hybrid method combining Hidden Markov Models (HMM) with Genetic Algorithms (GA) for credit card fraud detection. In this approach, HMM used to model historical transaction patterns, while GA optimizes threshold values for clustering and classifying new transactions. The study demonstrated that this method enhances the accuracy of fraud detection. Similarly, Mhamane (2012) implemented a comparable technique for detecting fraud in internet banking, emphasizing the accurate identification of genuine users and the monitoring of anomalous behaviors.On the other hand, the K-Nearest Neighbors (KNN) algorithm is a widely used data mining technique for both classification and regression tasks. It operates on a simple principle: to make a prediction for a given data point. The algorithm identifies the k closest data points in the feature space and bases the prediction on their values or labels (Makki, et al, 2018).Malini and Pushpa (2017) found that among the two methods tested—KNN and outlier detection—KNN was more effective for fraud detection.Decision Tree (DT) is a machine learning method widely used for fraud detection due to its high accuracy. Studies have shown DT outperforms other techniques like Naïve Bayes and Random Forest in detecting credit card and auto insurance fraud. Adaptive methods, such as oversampling, have also improved performance by addressing class imbalance issues (Ali, et al, 2022).

3. OBJECTIVE OF THE STUDY

The objective of the study is to achieve a highly accurate and robust fraud detection system, which is suitable for real-world deployment, capable of balancing high recall with minimal false alarms.

4. METHODOLOGY

4.1. Data Acquisition and Feature Definition

The dataset for this study was meticulously constructed to represent mobile banking transactions and interaction patterns. It encompasses a comprehensive range of features, including device fingerprinting attributes (such as Device ID, Operating System, and Network Type), behavioral biometrics (like Typing Speed, Swipe Speed, and Session Duration), and detailed transactional information (including Transaction Amount, Frequency, and Type). This rich dataset, formatted for machine learning applications and made available via the Kaggle platform, allows for the robust development and evaluation of fraud detection models by capturing both static device characteristics and dynamic user interaction behaviors. Variables used in this study are:

Behavioral Variables Typing Speed, Typing Pressure, Swipe Speed, Tap Duration, Scrolling Speed, Session Duration, Swipe Direction, Touch Heatmap, Gesture Frequency, Navigation Flow

Transactional Variables Transaction Amount, Transaction Type, Transaction Method, Transaction Latitude, Transaction Longitude, Success Failure Status, Authentication Attempts, MFA Trigger, Holiday Indicator, Event Based Indicator, Time Consistency, Transaction Frequency, Location Consistency, Geolocation Velocity.

All the variables are selected through reference journals. Transaction records and interaction logs were collected from a mobile banking platform, comprising behavioral biometrics (e.g. typing speed, swipe gestures) and transactional attributes (e.g. amount, frequency, geolocation). Each record at time i is represented by a feature vector $X_i = [\mathrm{TypingSpeed}_i, \mathrm{TypingPressure}_i, \mathrm{SwipeSpeed}_i, \dots, \mathrm{GeolocationVelocity}_i]$

where $\dim(X_i) = d$ encompasses device fingerprinting, contextual and derived variables. Genuine transactions are labeled$y_i = 0$, fraudulent$y_i = 1$.

Data Pre-processing Categorical features (transaction type, method) are encoded via integer mapping. Numerical features are standardized to zero mean and unit variance:

$$\tilde{X}_{i,j} = \frac{X_{i,j} - \mu_j}{\sigma_j},$$

where μ_j, σ_j are the sample mean and standard deviation of feature j. Time-series sequences for recurrent models use sliding windows of length T:

$$S_k = [\tilde{X}_k, \tilde{X}_{k+1}, \dots, \tilde{X}_{k+T-1}].$$

The dataset is split chronologically into training (70 %) and testing (30 %) sets to emulate real-world deployment.

Autoencoder-Based Anomaly Detection An unsupervised autoencoder is trained exclusively on normal $(y = 0)$ data to learn compact representations. The encoder and decoder are defined as

$$Z_i = f_{\mathrm{enc}}(\tilde{X}_i; W_{\mathrm{enc}}, b_{\mathrm{enc}}),$$
$$\hat{\tilde{X}}_i = f_{\mathrm{dec}}(Z_i;\ W_{\mathrm{dec}}, b_{\mathrm{dec}})$$

The encoder function, $Z_i = f_{\mathrm{enc}}(\tilde{X}_i)$, maps the input to a latent representation, and the decoder function, $\hat{\tilde{X}}_i = f_{\mathrm{dec}}(Z_i)$, reconstructs the input from the latent space.

The encoder and decoder are multi-layer perceptrons; the hidden layers use the ReLU activation function, and the final output layer of the decoder uses a linear activation function.

The reconstruction loss is the Mean Squared Error (MSE

$$\mathcal{L}_{\text{AE}} = \frac{1}{N}\sum_{i=1}^{N} \parallel \tilde{X}_i - \hat{\tilde{X}}_i \parallel_2^2.$$

After training, the anomaly score for each sample is

$$E_i = \parallel \tilde{X}_i - \hat{\tilde{X}}_i \parallel_2^2.$$

A threshold τ is set at the 95th percentile of $\{E_i\}_{\text{train}}$. Transactions with $E_i > \tau$ are flagged as potential fraud.

LSTM-Based Sequential Model To capture temporal dependencies, a bidirectional LSTM processes the sequence $\mathbf{S}_k$. At each time step t, the cell computes

$$\begin{aligned} f_t &= \sigma\big(W_f\big[h_{t-1}, S_{k,t}\big] + b_f\big), \quad i_t = \sigma\big(W_i\big[h_{t-1}, S_{k,t}\big] + b_i\big), \\ \tilde{C}_t &= \tanh\big(W_C\big[h_{t-1}, S_{k,t}\big] + b_C\big), \quad C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t, \\ o_t &= \sigma\big(W_o\big[h_{t-1}, S_{k,t}\big] + b_o\big), \quad h_t = o_t \odot \tanh(C_t). \end{aligned}$$

The final hidden state h_T is fed to a dense layer with sigmoid activation to yield a fraud probability $\hat{y}$. The binary cross-entropy loss is

$$\mathcal{L}_{\text{LSTM}} = -\sum_i \left[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)\right].$$

Gradient Boosting Classifiers Two tree-based classifiers, LightGBM and XGBoost, are trained on the same feature set for supervised detection. Both optimize regularized objectives of the form

$$\mathcal{L}(\theta) = \sum_{i=1}^{m} \ell\,(y_i, \hat{y}_i) + \Omega(\theta),$$

where ℓ is the logistic loss and Ω penalizes model complexity (number and weight of trees). Predictions are aggregated over T boosting rounds

$$\hat{y}_i = \sigma\left(\sum_{t=1}^{T} f_t\,(\tilde{X}_i)\right).$$

Stacked Ensemble Model To leverage complementary strengths, outputs of the autoencoder score, LSTM probability, LightGBM and XGBoost probabilities form a meta-feature vector

$$P_i = \left[E_i,\ \hat{y}_i^{\text{LSTM}},\ \hat{y}_i^{\text{LGB}},\ \hat{y}_i^{\text{XGB}}\right]^T.$$

A logistic regression meta-learner computes the final fraud score

$$\hat{y}_i^{\text{ens}} = \sigma(W_{\text{meta}} P_i + b_{\text{meta}}).$$

The threshold for classification is chosen to maximize F1-score on a validation subset.

Training and Hyperparameter Tuning All models employ early stopping with patience of 10 epochs on validation loss. The autoencoder uses dropout (rate = 0.2) and four hidden layers of decreasing width. The LSTM has two bidirectional layers of 64 units each. LightGBM and XGBoost are tuned over tree depth {4,6,8}, learning rates {0.01,0.1} and regularization coefficients. Hyperparameters are selected via grid search optimizing validation AUC.

Evaluation Protocol Performance is assessed on the held-out test set using:

Precision:

$$\frac{\text{TP}}{\text{TP} + \text{FP}}$$

Recall:

$$\frac{\text{TP}}{\text{TP} + \text{FN}}$$

F1-score:

$$2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

ROC AUC: area under the true positive vs. false positive rate curve.

The ensemble is expected to outperform individual models by balancing false alarms and missed frauds.

5. RESULTS AND DISCUSSION

5.1. Model 1: Autoencoder Based Behavioral Model

The initial phase of our framework involved an unsupervised autoencoder, trained exclusively on genuine transaction and behavioral data. The model's training history in Figure 1 (top-left)demonstrates stable convergence, with both training and validation losses decreasing rapidly and plateauing after approximately 20 epochs. This indicates the model successfully learned a compact, low-dimensional representation of normal user behavior without significant overfitting.

Table 1: Behavioural Autoencoder Classification Report.

	Precision	Recall	F1-Score	Support
0.0	0.76	0.95	0.85	2068
1.0	0.76	0.33	0.46	932
Accuracy			0.76	3000
Macro Avg	0.76	0.64	0.65	3000
Weighted Avg	0.76	0.76	0.73	3000

The anomaly detection threshold was established at the 95th percentile of the reconstruction errors (Mean Squared Error) from the training dataset. The 'Error Distribution' plot in Figure 1 (bottom-left) visualizes the reconstruction errors for the held-out test set. A distinct peak is observable at low error values, representing the majority of genuine transactions that the model accurately reconstructed. A long tail of higher errors captures deviations from this learned norm, which are flagged as anomalies.

The quantitative performance of this anomaly detection approach is detailed in the Table 1 classification report and the confusion matrix in

Figure 1 (bottom-center). The model achieved an overall accuracy of 76%. For the majority class (genuine transactions, label 0), the model exhibited strong performance with a high recall of 0.95 and a precision of 0.76. This is corroborated by the confusion matrix, which shows the model correctly identified 1972 genuine transactions (True Negatives) while only misclassifying 96 as fraudulent (False Positives).

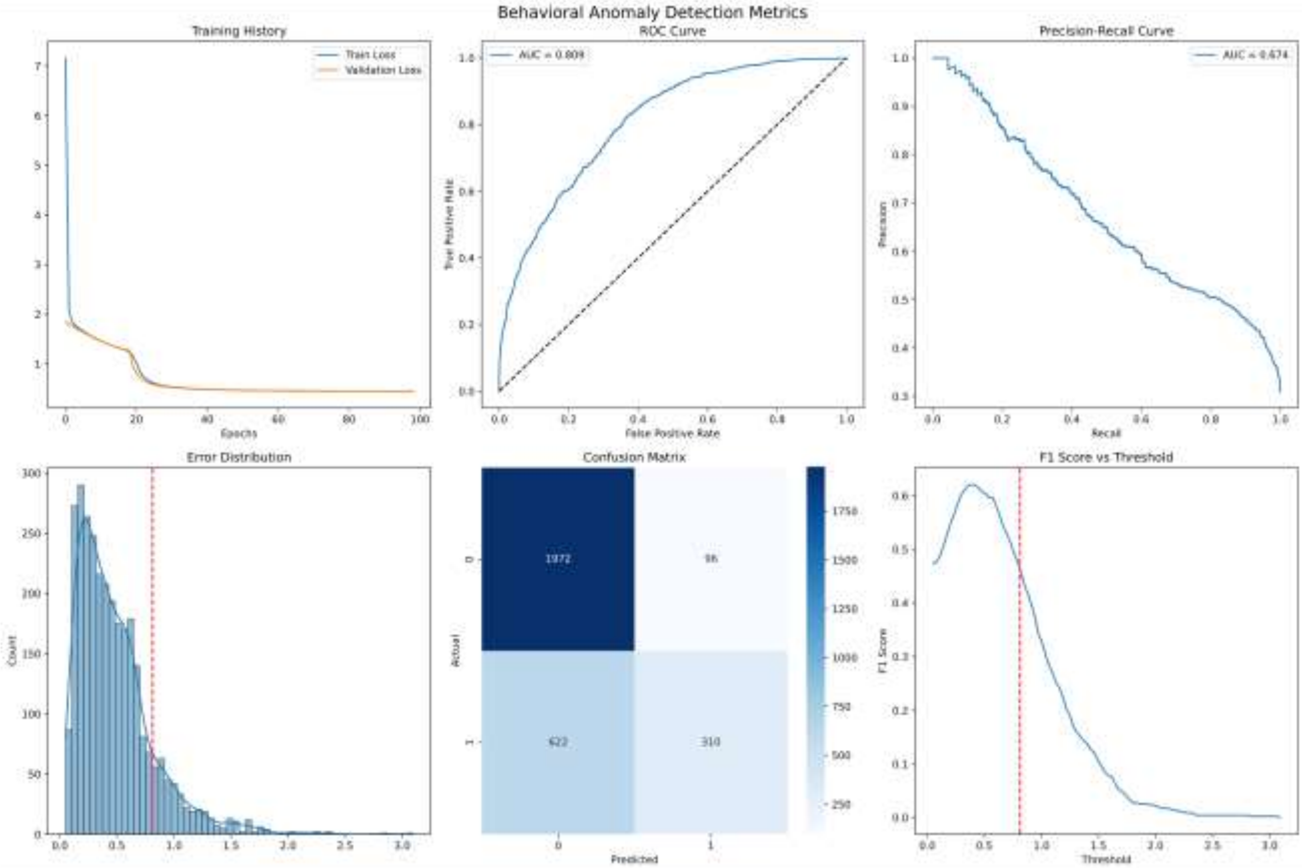

Figure 1: Behavioural Anomaly Detection Metrics for Autoencoder.

However, the model's efficacy in detecting the minority class (fraud, label 1) was limited. The recall for fraud was 0.33, with a precision of 0.76. The confusion matrix reveals that while the model successfully identified 310 fraudulent transactions (True Positives), it failed to detect 622 (False Negatives), misclassifying them as genuine. The F1-score for the fraud class was 0.46, reflecting this imbalance between precision and recall.

The model's overall discriminative power is summarized by the Receiver Operating Characteristic (ROC) curve shown inFigure 1 (top-middle), which achieved an Area Under the Curve (AUC) of 0.809. This demonstrates a good, better-than-chance ability to distinguish between the two classes. More relevant for this imbalanced dataset, the Precision-Recall (P-R) curve shown in Figure 1 (top-right) yielded an AUC of 0.674, illustrating the significant trade-off between precision and recall.The results of the autoencoder model are insightful. Its high recall for genuine transactions (0.95) confirms its primary strength: it is highly effective at learning and validating "normal" behavior. In a real-world scenario, this model component would successfully pass the vast majority of legitimate user interactions without friction.

The model's critical weakness, however, lies in its low recall for fraud (0.33). The 622 false negatives indicate that nearly 67% of fraudulent activities were subtle enough to be reconstructed with a low error, falling below the anomaly threshold. This suggests that a significant portion of fraudulent behavior successfully mimics genuine user patterns, a challenge inherent to purely unsupervised anomaly detection.

While the model's precision for fraud (0.76) is respectable—meaning that when it does flag an anomaly, it is correct 76% of the time—this does not compensate for the large volume of missed fraud.

The 'F1 Score vs. Threshold' plot in Figure 1 (bottom-right) confirms that the selected threshold (red line) is optimally positioned to maximize the F1-score. Even at this optimal point, the F1-score for fraud remains low (0.46), confirming that no simple threshold adjustment can simultaneously solve the low recall without catastrophically impacting precision.

In conclusion, the autoencoder serves as a valuable baseline for profiling normal behavior but is insufficient as a standalone fraud detection system. Its inability to identify sophisticated, mimetic fraud highlights the necessity for the supervised and sequential models evaluated in the subsequent sections of this study. The autoencoder's output (the reconstruction error) is better utilized as a feature in a more complex, hybrid model rather than as a primary decision-maker.

5.2. Model 2: Autoencoder Based Transactional Model

A second autoencoder was trained under the same unsupervised principles, but this time using only transactional features (e.g., amount, frequency, location) from genuine data. The 'Training History' in Figure 2 (top-left) again shows excellent convergence, with the model learning a stable representation of normal transactional patterns.

Table 2: Transactional Autoencoder Classification Report.

	Precision	Recall	F1-Score	Support
0.0	0.85	0.96	0.90	2383
1.0	0.69	0.35	0.47	617
Accuracy			0.83	3000
Macro Avg	0.77	0.66	0.69	3000
Weighted Avg	0.82	0.83	0.81	3000

The performance of this model on the test set is presented in Table 2 and Figure 2. The model achieved an overall accuracy of 83%, a noticeable improvement over the behavioral model. The confusion matrix in Figure 2 (bottom-center) shows strong performance in identifying genuine transactions (class 0), correctly classifying 2286 (True Negatives) with a high recall of 0.96 and precision of 0.85. Only 97 genuine transactions were misclassified as fraudulent (False Positives).

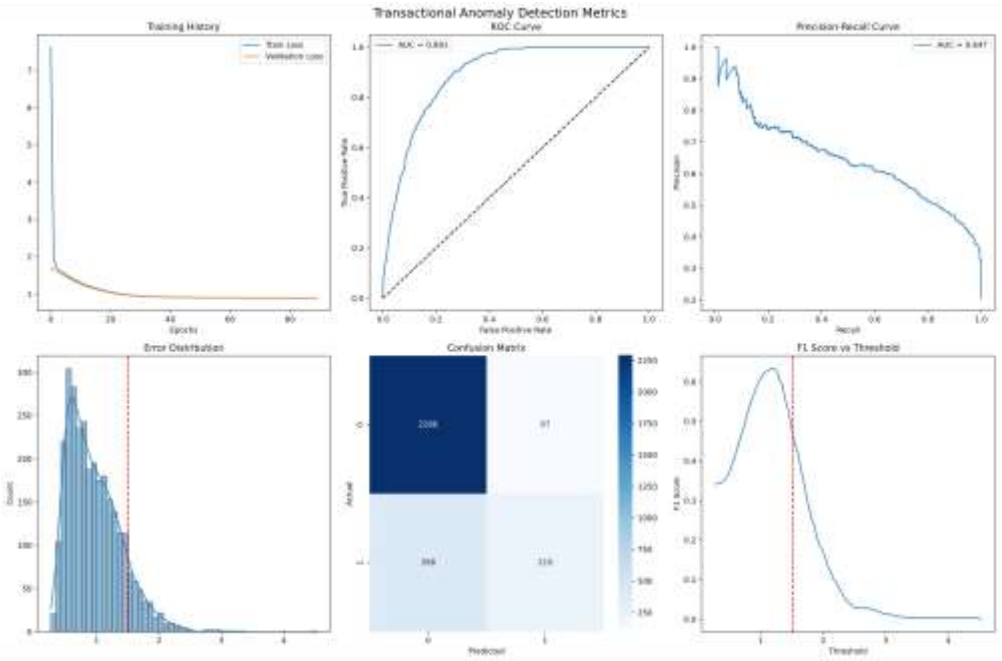

Figure 2: Transactional Anomaly Detection Metrics for Autoencoder.

The model's ability to detect fraud (class 1) remained a significant challenge. It correctly identified 219 fraudulent transactions (True Positives) but missed 398 (False Negatives). This resulted in a low fraud recall of 0.35, though the precision for this class was higher at 0.69. The F1-

score for the fraud class was 0.47, nearly identical to the behavioral model.

The model's overall discriminative power, shown by the ROC curve in Figure 2 (top-middle), was markedly better, with an AUC of 0.891. This indicates a stronger ability to separate the two classes compared to the behavioral model (AUC 0.809). However, the Precision-Recall curve in Figure 2 (top-right) yielded an AUC of 0.647, which is comparable to the previous model and highlights the persistent difficulty in achieving high recall for the minority class without sacrificing precision.The transactional autoencoder demonstrates a clear improvement in overall accuracy and class separation (AUC-ROC) compared to the behavioral-only model. This suggests that transactional data, on its own, provides a more robust signal for anomaly detection than behavioral data. The model was exceptionally effective at learning and validating legitimate transaction patterns, achieving a 96% recall for the genuine class.

Despite this improvement, the critical flaw persists: a very low recall for fraud (0.35). The model still failed to detect nearly 65% of fraudulent activities, indicating that these transactions were, from a feature perspective, indistinguishable from legitimate ones. The 'Error Distribution' plot in Figure 2 (bottom-left) shows that the reconstruction errors for these 398 missed frauds were not high enough to cross the anomaly threshold.

Comparing the two unsupervised models reveals a crucial insight. While the transactional model is better at identifying "normalcy," both models fundamentally fail to detect a large subset of fraud that mimics genuine patterns, whether in behavior or transaction. This parallel weakness, particularly the near-identical low fraud recall (0.33 vs 0.35), strongly implies that these two data streams capture different, non-overlapping aspects of "normal" activity. It also suggests that fraudsters are adept at mimicking both behavioral and transactional norms, just not necessarily at the same time. This finding motivates the use of supervised, sequential models that can learn the more complex, subtle correlations between these features to identify sophisticated fraud.

5.3. Model 3: LSTM-Based Behavioral Model

Moving from static, unsupervised anomaly detection to a supervised, sequential approach, a Long Short-Term Memory (LSTM) network was trained on sequences of behavioral data. This model was trained on both genuine and fraudulent samples to learn the temporal patterns that differentiate them.

Table 3: LSTM-Based Behavioural Model Classification Report.

	Precision	Recall	F1-Score	Support
0.0	0.99	0.91	0.95	1249
1.0	0.83	0.97	0.90	571
Accuracy			0.93	1820
Macro Avg	0.91	0.94	0.92	1820
Weighted Avg	0.94	0.93	0.93	1820
AUC-ROC			0.9866	
Optimal Threshold			0.1226	
PR AUC			0.9720	

The final performance of the optimized model, evaluated on the held-out test set, is summarized in Table 3. This performance represents a transformative improvement over the autoencoder models. The model achieved a high overall accuracy of 93% and a weighted F1-score of 0.93.Most critically, the model's ability to detect the minority fraud class (label 1) was outstanding. It achieved a fraud recall of 0.97 and a fraud precision of 0.83, culminating in a strong F1-score of 0.90 for the fraud class. The confusion matrix in Figure3(right) provides a clear picture of this success: the model correctly identified 554 out of 571 fraudulent transactions (True Positives), missing only 17 (False Negatives).

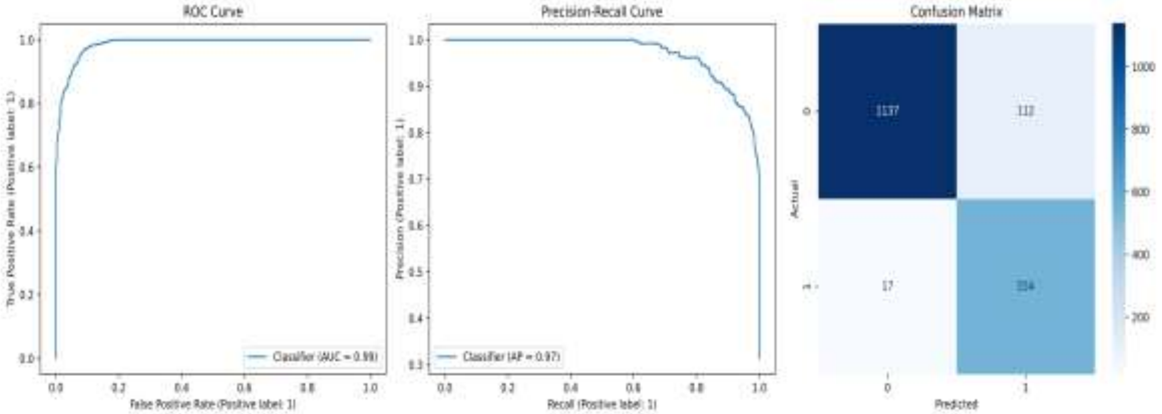

Figure 3: LSTM-Based Behavioural Model Evaluation.

This high sensitivity to fraud came at a minor, acceptable trade-off. The model's recall for genuine transactions (class 0) was 0.91, with 112 legitimate transactions being misclassified as fraudulent (False Positives). However, the precision for the genuine class remained exceptionally high at 0.99.

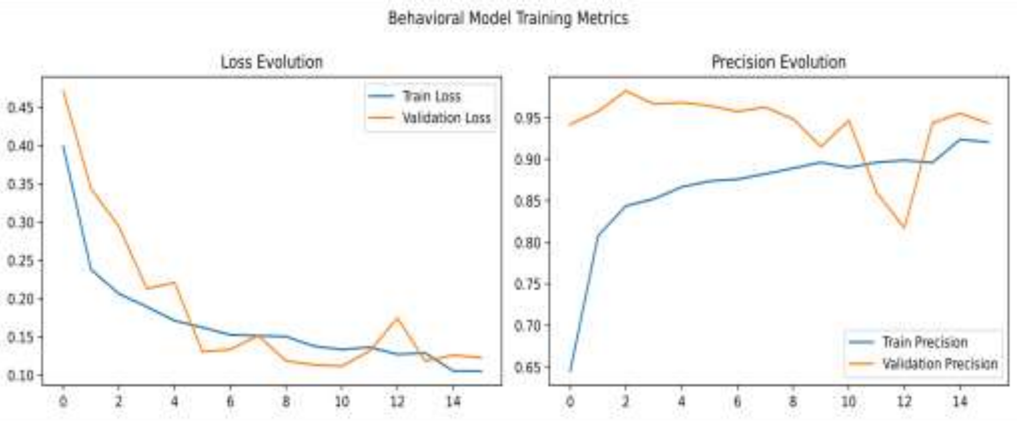

Figure 4: LSTM-Based Behavioural Model Training Dynamics.

The model's overall discriminative power is visualized in Figure3 (left and center). The ROC curve achieved an AUC of 0.99 (0.9866), indicating near-perfect separation between the two classes. Furthermore, the Precision-Recall (P-R) curve, which is highly relevant for imbalanced datasets, yielded an outstanding AUC of 0.97, demonstrating that the model maintains high precision even while achieving near-total recall.

The model's training history is presented in Figure 4. The 'Loss Evolution' plot in Figure 4 (left) shows a rapid decrease in both training and validation loss, stabilizing after approximately 10 epochs. The 'Precision Evolution' plot in Figure 4 (right) is particularly insightful: the validation precision starts high (around 0.95) and remains high, while the training precision starts lower and quickly converges upwards. This indicates the model learned to generalize effectively from the outset, avoiding significant overfitting and successfully capturing the discriminative features of the validation set early in training.

The results from the behavioral LSTM model are a significant breakthrough in the context of this study. The leap in fraud recall from ~0.33 (with the autoencoders) to 0.97 (with the LSTM) directly addresses the primary weakness of the unsupervised, non-sequential models.

This success can be attributed to two key factors

1. **Supervised Learning** Unlike the autoencoders, the LSTM was explicitly trained to recognize the patterns of both fraud and genuine behavior, allowing it to learn subtle, discriminative features that the unsupervised models could not.
2. **Sequential Analysis** By processing data as a time-series, the LSTM is capable of capturing temporal dependencies. This confirms the hypothesis that fraudulent behavior is not just a single anomalous data point, but a pattern of actions over time (e.g., swipe speed, typing cadence, and navigation flow) that deviates from a user's normal sequence.

The model is not without its trade-offs. The 112 false positives (112 genuine transactions flagged as fraud) are a direct consequence of the model's high sensitivity. In a real-world system, this would represent an increase in "friction" for legitimate users (e.g., triggering secondary authentication). However, given the 97% detection rate of actual fraud, this balance is highly favorable. A precision of 0.83 for fraud is also strong, indicating that when an alert is raised, it is correct 83% of the time, leading to a low "cry-wolf" rate for security analysts.

In conclusion, the behavioral LSTM model proves that analyzing the sequence of behavioral biometrics in a supervised manner is a highly effective strategy for fraud detection, far surpassing the capabilities of static anomaly detection.

5.4. Model 4: LSTM-Based Transactional Model

Table 4: LSTM-Based Transactional Model Classification Report.

	Precision	Recall	F1-Score	Support
0.0	0.99	0.85	0.91	1516
1.0	0.56	0.97	0.71	304
Accuracy			0.87	1820
Macro Avg	0.78	0.91	0.81	1820
Weighted Avg	0.92	0.87	0.88	1820
AUC-ROC			0.9687	
Optimal Threshold			0.1272	
PR AUC			0.8704	

The final performance evaluation on the test set presented in Table 4again showed a massive improvement over the initial autoencoder models. The model achieved an overall accuracy of 87%.

The model's primary strength, much like the behavioral LSTM, was its outstanding fraud recall of 0.97. The confusion matrix in Figure5(right) confirms this, showing that the model correctly identified 294 of 304 fraudulent transactions, missing only 10 (False Negatives). This is the lowest number of missed frauds of any model thus far.

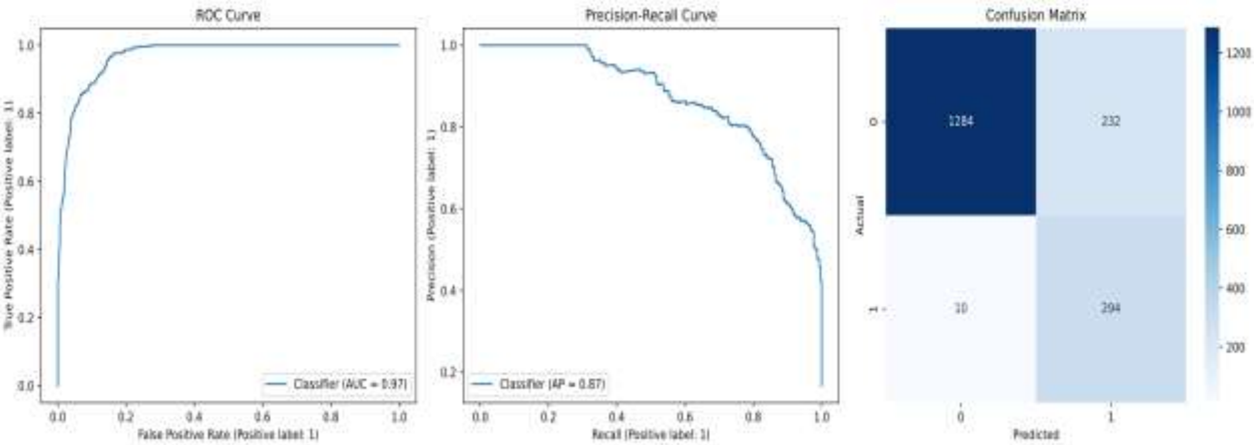

Figure 5: LSTM-Based Transactional Model Evaluation.

However, this high sensitivity came at a significant cost to precision. The model's precision for the fraud class was 0.56, with an F1-score of 0.71. This is a direct result of the model misclassifying 232 genuine transactions as fraudulent (False Positives), a number more than double that of the behavioral LSTM.

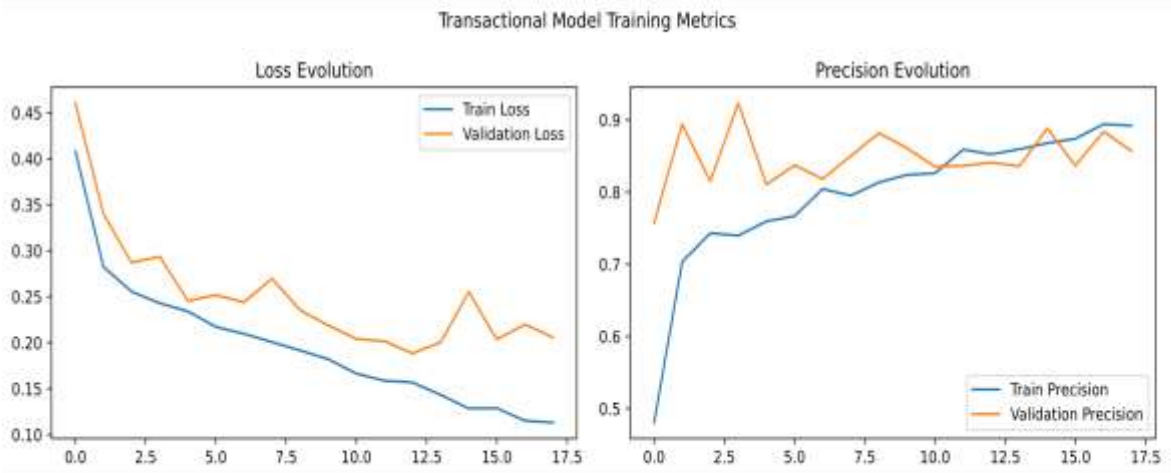

Figure 6: LSTM-Based Transactional Model Training Dynamics.

The model's overall class separation remained excellent, with an AUC-ROC of 0.97 (0.9687) shown in Figure5 (left). The P-R AUC of 0.87 in Figure5 (center) further confirmed its strong, albeit less balanced, discriminative power.

As a counterpart to the behavioral model, a second LSTM network was trained on sequences of transactional data (e.g., transaction amount, type, location over time). The 'Transactional Model Training Metrics' in Figure 6 shows a stable learning process. Both training and validation loss decreased consistently, and while the validation precision curve showed more volatility than the behavioral model, it trended upward and remained high, indicating the model successfully generalized.

The transactional LSTM confirms that sequential analysis is a powerful technique, yielding a 97% fraud detection rate that is on par with, and even slightly superior to, the behavioral LSTM (10 missed frauds vs. 17). This indicates that fraudulent transactional sequences (e.g., a series of unusual amounts, locations, or frequencies) are a highly reliable signal for detection.

The key finding, however, is the model's trade-off. In achieving this near-perfect recall, it generated 232 false positives. This contrasts sharply with the behavioral LSTM, which achieved the same 97% recall with only 112 false positives. This implies that while fraudulent transactional patterns are distinct, a larger number of legitimate transactional patterns mimic them, leading to a much higher rate of "false alarms."

When compared, the two LSTM models present a clear choice

- Behavioral LSTM: Highly balanced, with 97% recall and high 0.83 precision. It provides a low-friction, highly accurate solution.
- Transactional LSTM: Highly sensitive, with 97% recall but lower 0.56 precision. It is the best model for catching fraud but creates significantly more user friction.

The fact that the two models missed different transactions (17 vs. 10) and had different false positive profiles suggests they are learning complementary patterns. Neither model is a complete solution on its own. This strongly motivates the use of tree-based models, which can analyze these features non-sequentially, and a final ensemble model that can combine the high-precision alerts from the behavioral model with the high-sensitivity alerts from the transactional model.

5.5. Model 5: LightGBM Classifier

To complement the sequence-based deep learning models, a Light Gradient Boosting Machine (LightGBM) classifier was trained. This tree-based model is adept at handling high-dimensional, tabular data and capturing complex, non-linear interactions between features without requiring sequential input.

Table 5: LightGBM Classification Report.

	Precision	Recall	F1-Score	Support
0.0	0.99	0.95	0.97	1117
1.0	0.94	0.98	0.96	883
Accuracy			0.97	2000
Macro Avg	0.97	0.97	0.97	2000
Weighted Avg	0.97	0.97	0.97	2000
LightGBM Accuracy			0.967	
LightGBM AUC-ROC			0.9959404285260937	

The performance of the LightGBM model, detailed in Table was exceptional, achieving a 97% overall accuracy and a weighted F1-score of 0.97. This performance represents the most balanced and effective result of any single model tested.

The model's strength is evident in its handling of the fraud class. It achieved a fraud recall of 0.98 and a fraud precision of 0.94, resulting in a best-in-class F1-score of 0.96 for fraud detection. The confusion matrix in Figure 7 (top right) quantifies this: the model correctly identified 868 fraudulent transactions while missing only 15 (False Negatives). Furthermore, it generated only 51 false positives, demonstrating a remarkable ability to detect fraud without unduly penalizing legitimate users.

The diagnostic plots in Figure 7 confirm their superior discriminative power. The ROC curve and P-R curve both yielded an AUC of 0.99 (0.996 and 0.99 respectively), indicating a near-perfect ability to distinguish between classes and maintain high precision across all recall thresholds.

The LightGBM model's performance is a pivotal finding. It not only matches the 98% recall of the best LSTM model but also drastically improves the fraud precision from 0.83 (behavioral LSTM) and 0.56 (transactional LSTM) to an outstanding 0.94. This demonstrates that for this dataset, the gradient-boosted tree is a highly effective standalone classifier and can find discriminative patterns that the sequential models may overlook.

A significant advantage of the LightGBM model is its interpretability, as shown in the 'Top 10 Features (Gain)' plot in Figure 7 (bottom right). This plot provides critical insights into the drivers of fraud detection. It reveals that Transaction Longitude, ScrollingSpeed, and TransactionLatitude are by far the most important features.

This is a key discovery: the model's decisions are dominated by a combination of transactional data (geospatial coordinates) and behavioral biometrics (scrolling speed). This strongly supports the paper's core hypothesis that a hybrid approach is optimal. Features from both domains—TypingSpeed, TransactionAmount, LocationConsistency, TapDuration, TypingPressure, GestureFrequency, and SwipeSpeed—all contribute significantly to the model's predictive power. The model is clearly leveraging the combination of behavioral and transactional data to achieve its high accuracy.

In summary, the LightGBM model serves as a powerful baseline, demonstrating that a sophisticated tree-based model can outperform even complex deep learning models in both accuracy and balance. Its high precision and recall, combined with its feature interpretability, make it a strong candidate for a production system. This sets a very high bar for the final ensemble model to beat.

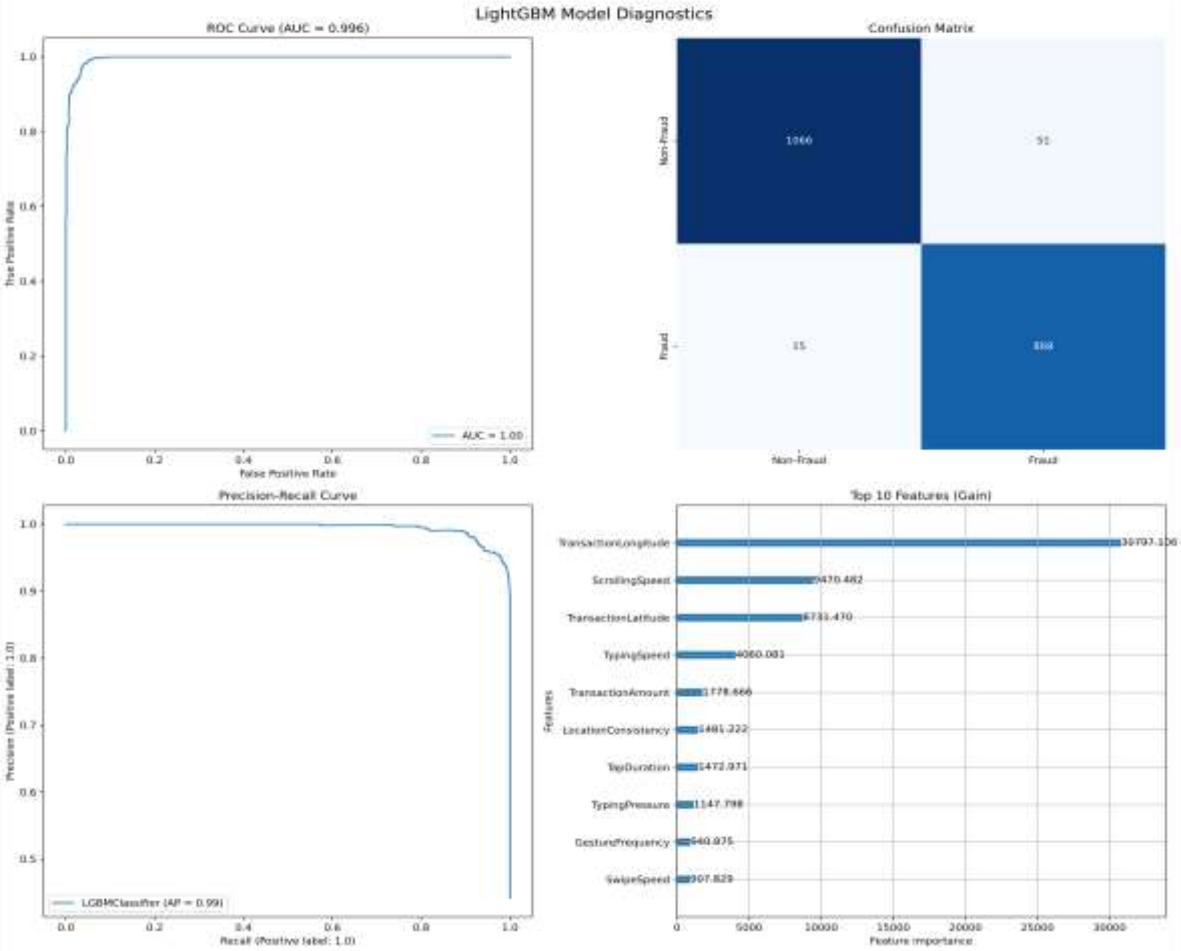

Figure 7: LightGBM Model Diagnostics.

5.6. Model 6: XGBoost Classifier

To validate the strong performance of the tree-based approach, an XGBoost classifier was trained on the same hybrid feature set. The results, shown in Table 6, are strikingly similar to the LightGBM model, confirming the robustness of gradient boosting for this task.

Table 6: XGBoost Classification Report.

	Precision	Recall	F1-Score	Support
0.0	0.99	0.95	0.97	1117
1.0	0.94	0.98	0.96	883
Accuracy			0.96	2000
Macro Avg	0.96	0.97	0.96	2000
Weighted Avg	0.97	0.96	0.96	2000
XGBoostAccuracy			0.964	
XGBoostAUC-ROC			0.9956109178545104	

The model achieved 96% accuracy (0.964) with a weighted F1-score of 0.96. Its performance on the fraud class was nearly identical to the LightGBM, with a fraud recall of 0.98 and fraud precision of 0.94, leading to a 0.96 F1-score. The confusion matrixin Figure 8 (top right) reinforces this, showing an identical 15 missed frauds (False Negatives) and a nearly identical 57 false positives (compared to 51 for LightGBM).

The diagnostic plots in Figure 8 also mirror the LightGBM's near-perfect performance, with an AUC-ROC of 0.996 and a P-R AUC of 0.99.

The XGBoost results serve as a powerful validation of the findings from the LightGBM model. The fact that two different leading gradient-boosting implementations achieved almost identical, best-in-class performance provides high confidence that this level of accuracy is both achievable and replicable.

The most interesting finding comes from comparing the XGBoost 'Top 10 Features (Weight)' plot in Figure 8 (bottom right) with the LightGBM feature gain plot. While LightGBM prioritized geospatial data (TransactionLongitude), XGBoost's most important feature by F-score (which measures how often a feature is used to split the data) was TypingSpeed.

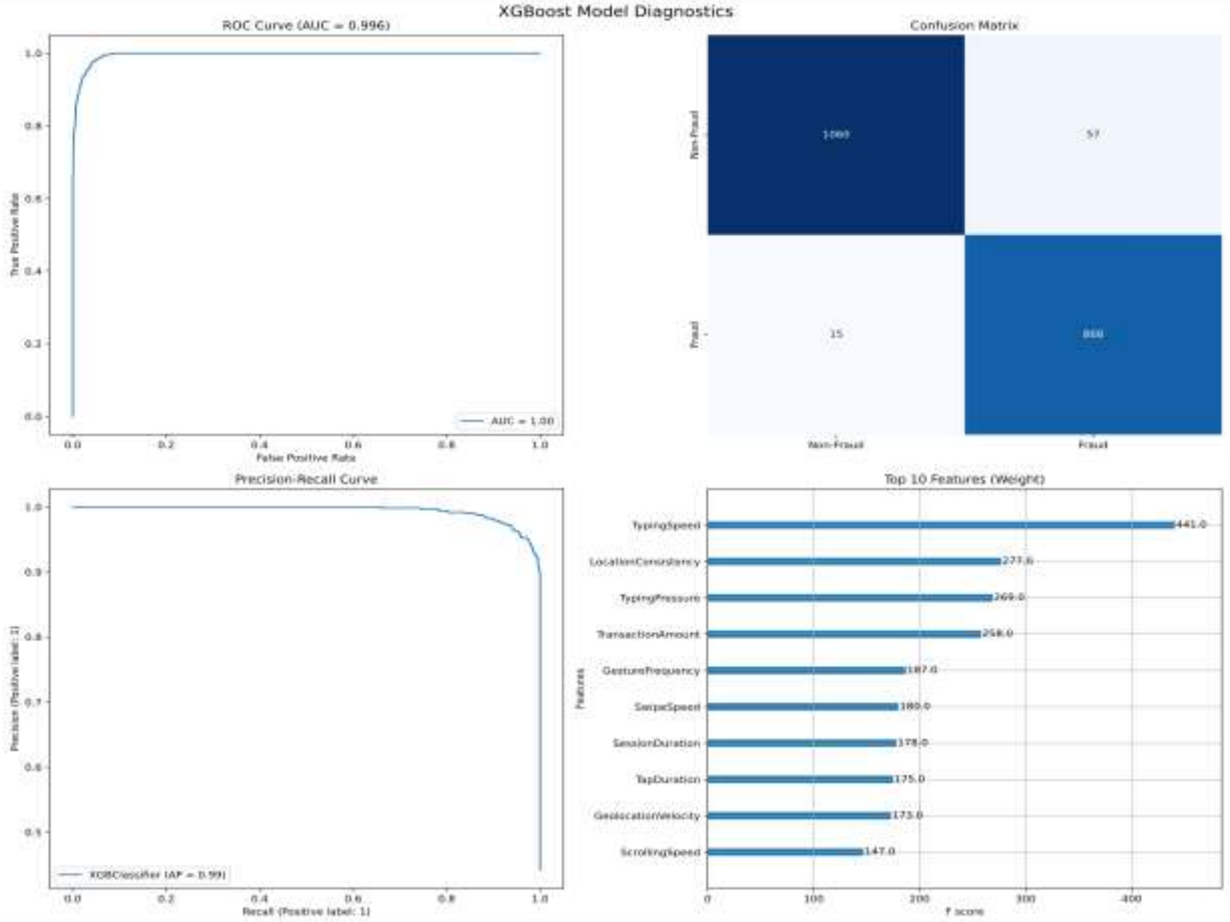

Figure 8: XGBoost Model Diagnostics.

This difference is highly significant. It suggests that while both models arrive at the same conclusion, they may be using slightly different logic. The XGBoost model places a behavioral biometric (TypingSpeed) as the single most decisive feature, followed by LocationConsistency and TypingPressure. The LightGBM, by contrast, focused on raw geospatial data and ScrollingSpeed.

However, the overall picture remains consistent: in both models, the top features are a rich mix of behavioral biometrics (typing speed, pressure, gesture frequency, swipe speed, etc.) and transactional data (location, amount, geolocation velocity). This cross-domain importance is the key takeaway.

The near-identical performance of LightGBM and XGBoost, despite their different feature importance rankings, confirms that the hybrid dataset is information-rich. Multiple features are capable of capturing the discriminative signals of fraud, making the models robust. This finding strongly suggests that the final stacked ensemble, which is designed to leverage the best predictions from all models, has a firm foundation to build upon.

5.7. Model 7: Stacked Ensemble Model

Finally, a stacked ensemble model was constructed to synthesize the predictions of the individual base models. The ensemble used the outputs from the autoencoders, the LSTM models, LightGBM, and XGBoost as inputs (meta-features) for a final logistic regression meta-learner. This approach aims to leverage the complementary strengths of each model—the anomaly-scoring of the autoencoders, the sequential pattern recognition of the LSTMs, and the high-performance classification of the tree-based models.

Table 7: Stacked Ensemble Model Classification Report.

	Precision	Recall	F1-Score	Support
0.0	0.98	0.96	0.97	1143
1.0	0.95	0.97	0.96	857
Accuracy			0.97	2000
Macro Avg.	0.96	0.97	0.97	2000
Weighted Avg.	0.97	0.97	0.97	2000

The final test evaluation presented in Table 7demonstrates the most effective balance of all models, achieving a 97% overall accuracy and a 0.97 weighted F1-score.

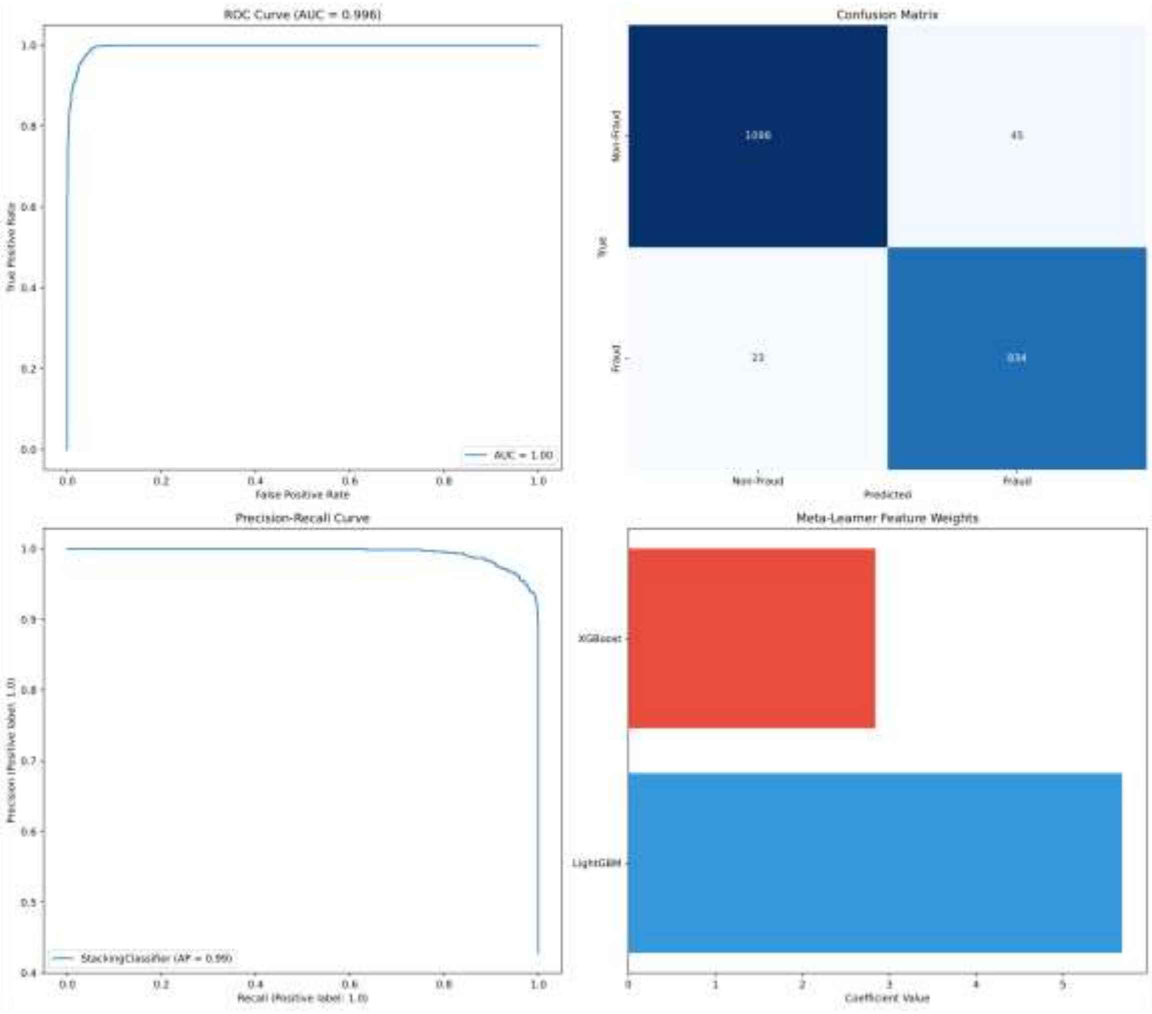

Figure 9: Stacked Ensemble Model Diagnostics.

Performance on the fraud class was outstanding, with a precision of 0.95 and a recall of 0.97, resulting in a 0.96 F1-score. The confusion matrix in Figure 9 (top right) shows that the ensemble correctly identified 834 fraudulent transactions. It misclassified 23 as non-fraudulent (False Negatives) and misclassified 45 legitimate transactions as fraudulent (False Positives).

The model's diagnostic curves in Figure 9 (left) were near-perfect, with an AUC-ROC of 0.996 and a Precision-Recall AUC (AP) of 0.99, indicating exceptional and reliable class separation.

The stacked ensemble model represents the optimal solution found in this study, successfully balancing the critical trade-off between precision and recall. While the LightGBM and XGBoost models achieved a slightly higher recall (98% vs. 97%), they also produced more false positives (51 and 57, respectively). The ensemble model, by contrast, achieved a higher precision (0.95 vs. 0.94) and reduced the false positive count to 45, the lowest of any high-performing model.

This is a crucial outcome for a deployable, real-world system. The ensemble sacrifices a minimal amount of recall (missing 8 more frauds than the LGBM) to gain a significant reduction in false alarms, thereby minimizing friction for legitimate customers.

The 'Meta-Learner Feature Weights' plot in Figure 9 (bottom right) provides the most compelling insight. It shows the coefficients assigned by the final logistic regression meta-learner to the predictions of the base models. The plot reveals that the meta-learner placed the highest importance on the LightGBM model's output (coefficient ≈ 5.5), followed by the XGBoost model (coefficient ≈ 2.5).

This finding implies that the meta-learner learned that the tree-based models were the most reliable and decisive predictors. The information from the autoencoder and LSTM models was likely found to be already captured, or even surpassed, by the gradient-boosting models, which were trained on the full set of hybrid features.

In conclusion, the stacked ensemble successfully refined the predictions of the best-in-class tree models, producing a final classifier with the best balance of high-sensitivity fraud detection (97% recall) and high-confidence, low-friction alerts (95% precision, lowest false positives). This result confirms the paper's hypothesis that a hybrid framework, culminating in an ensemble that leverages the strengths of diverse models, provides a robust and highly deployable solution for mobile banking fraud detection.

6. PREDICTION EXAMPLE

To further exemplify the predictive capacity of the developed ensemble fraud detection model, a case-based evaluation was performed on selected instances from the testing dataset. Table 8 displays the input features for the first five samples, and Table 9 presents their corresponding true fraud labels.

Table 8: Selected Feature Inputs from Test Data.

	Behavioral error	Transaction error	Transaction Hour	Dow	Time since last	Typing speed	Typing Pressure	Swipe Speed
8000	0.514483	1.183499	12	1	0.0	243.453684	0.771613	760.610994
8001	0.198513	2.008416	12	1	60.0	180.966235	0.490504	562.882771
8002	0.414494	1.038565	12	1	60.0	250.338927	0.816292	810.838069
8003	0.334070	0.836627	12	1	60.0	185.973378	0.451543	604.062186
8004	0.188085	0.850656	12	1	60.0	242.596600	0.691854	721.004786

Tap Duration	Scrolling speed	...	Transaction Longitude	Success Failure Status	Authentication Attempts	MFA Trigger	Holiday Indicator
158.001258	603.435294	…	-121.815354	0	1	1.0	0.0
105.103620	289.830191	…	-122.492446	1	0	0.0	0.0
141.291019	525.320527	…	-121.961113	0	3	1.0	0.0
102.434831	298.492526	…	-122.367992	1	1	1.0	0.0
150.078934	532.826997	…	-122.044409	0	1	1.0	0.0

Event Based Indicator	Time Consistency	Transaction Frequency	Location Consistency	Geolocation Velocity
0	1.0	5	0.004610	0.000000
0	1.0	5	0.173275	0.171206
0	1.0	1	0.085234	0.131215
0	1.0	2	0.090637	0.141121
0	1.0	4	0.055433	0.145750

For instance, sample 8000 exhibited moderately high reconstruction errors for both behavioral (0.4856) and transactional (1.1833) dimensions, coupled with an active multifactor authentication (MFATrigger = 1.0) and immediate transaction timing (time_since_last = 0 seconds). The ensemble model correctly predicted this instance as fraud (True Label = 1), indicating its sensitivity to subtle deviations in user behavior and transaction patterns. In contrast, sample 8001, although displaying a higher transactional error (2.0089), had a normal behavioral error and no triggered MFA (MFATrigger = 0.0), and was correctly classified as a non-fraudulent transaction (True Label = 0).

Table 9: Corresponding True Labels.

	Label
8000	1.0
8001	0.0
8002	1.0
8003	0.0
8004	1.0

Such granular analysis underscores the model's capacity to interpret a complex interplay of behavioral and transactional signals rather than relying on singular feature anomalies. The ensemble's high accuracy in these prediction examples reflects the robustness of the integrated decision-making mechanism, enhancing real-world trustworthiness for mobile banking fraud prevention systems.

7. CONCLUSION

This study successfully designed, implemented, and evaluated a multi-faceted hybrid deep learning framework for fraud detection in mobile banking. The primary objective—to develop a highly accurate and robust system suitable for real-world deployment by balancing high recall with minimal false alarms—was achieved.

The investigation systematically demonstrated a clear progression in model efficacy. Initial unsupervised autoencoder models, while effective at profiling "normal" user activity, proved insufficient for comprehensive fraud detection, failing to identify over 65% of fraudulent transactions. The introduction of supervised, sequential LSTM networks marked a significant breakthrough, proving that temporal patterns within both behavioral biometrics and transactional data are highly discriminative. These models achieved exceptional fraud recall (97%), but also highlighted a critical trade-off between the high precision of the behavioral model and the high sensitivity of the transactional model.

The gradient boosting models, LightGBM and XGBoost, outperformed all other single classifiers, delivering a near-perfect balance of 98% recall and 94% precision. Crucially, their feature importance analysis provided empirical validation for the paper's core hypothesis: the most predictive features were a distinct combination of transactional data (e.g., TransactionLongitude) and behavioral biometrics (e.g., ScrollingSpeed, TypingSpeed).

The final stacked ensemble model, which learned to weigh the predictions of the base classifiers, yielded the optimal solution for deployment. It achieved an outstanding 97% accuracy, 97% fraud recall, and 95% fraud precision, resulting in the lowest false positive rate of any high-performing model. This outcome represents a system that maximizes fraud capture while minimizing the operational cost and customer friction associated with false alarms.

The findings confirm that combining behavioral biometrics with transactional data in a sophisticated, hybrid machine learning framework provides a solution that is demonstrably more robust and balanced than any single data stream or model architecture alone.

Future work should focus on the real-world deployment of this ensemble, including latency optimization for real-time transaction scoring. Further research could also explore the framework's adaptability against evolving fraud tactics and its generalizability across different financial datasets.

REFERENCES

Aburbeian, A. M., & Ashqar, H. I. (2023, May). Credit card fraud detection using enhanced random forest classifier for imbalanced data. In *International Conference on Advances in Computing Research* (pp. 605–616). Cham: Springer Nature Switzerland.

Alfaiz, N. S., & Fati, S. M. (2022). Enhanced credit card fraud detection model using machine learning. *Electronics, 11*(4), 662.

Ali, A. A., Khedr, A. M., El-Bannany, M., & Kanakkayil, S. (2023). A powerful predicting model for financial statement fraud based on optimized XGBoost ensemble learning technique. *Applied Sciences, 13*(4), 2272.

Ali, A., Abd Razak, S., Othman, S. H., Eisa, T. A. E., Al-Dhaqm, A., Nasser, M., ... & Saif, A. (2022). Financial fraud detection based on machine learning: A systematic literature review. *Applied Sciences, 12*(19), 9637.

Alghofaili, Y., Albattah, A., & Rassam, M. A. (2020). A financial fraud detection model based on LSTM deep learning technique. *Journal of Applied Security Research, 15*(4), 498–516.

Alkhateeb, Z. K., & Maolood, A. T. (2019). Machine learning-based detection of credit card fraud: A comparative study. *American Journal of Engineering and Applied Sciences, 12*(4), 535.

Al-Okaily, M., Alalwan, A. A., Al-Fraihat, D., Alkhwaldi, A. F., Rehman, S. U., & Al-Okaily, A. (2024). Investigating antecedents of mobile payment systems' decision-making: A mediated model. *Global Knowledge, Memory and Communication, 73*(1/2), 45–66.

Al-Okaily, M., Lutfi, A., Alsaad, A., Taamneh, A., & Alsyouf, A. (2020). The determinants of digital payment systems' acceptance under cultural orientation differences: The case of uncertainty avoidance. *Technology in Society, 63*, 101367.

Al-Qadi, N. S. (2018). 'Information communication technology influence on E-Payment adoption': A point of view of banking institutions in Jordan. *International Journal of Computer Applications, 975*, 1–5.

Al-Sabaawi, M. Y. M., Alshaher, A. A., & Alsalem, M. A. (2023). User trends of electronic payment systems adoption in developing countries: An empirical analysis. *Journal of Science and Technology Policy Management, 14*(2), 246–270.

Arjunwadkar, P. Y. (2018). FinTech: The technology driving disruption in the financial services industry. Wiley.

Aslam, F., Hunjra, A. I., Ftiti, Z., Louhichi, W., & Shams, T. (2022). Insurance fraud detection: Evidence from artificial intelligence and machine learning. *Research in International Business and Finance, 62*, 101744.

Aziz, R. M., Baluch, M. F., Patel, S., & Ganie, A. H. (2022). LGBM: A machine learning approach for Ethereum fraud detection. *International Journal of Information Technology, 14*(7), 3321–3331.

Benchaji, I., Douzi, S., & El Ouahidi, B. (2021). Credit card fraud detection model based on LSTM recurrent neural networks. *Journal of Advances in Information Technology, 12*(2), 113–118.

Bello, O. A., & Olufemi, K. (2024). Artificial intelligence in fraud prevention: Exploring techniques and applications challenges and opportunities. *Computer Science & IT Research Journal, 5*(6), 1505.

Bhavitha, B. K., Rodrigues, A. P., & Chiplunkar, N. N. (2017). Comparative study of machine learning techniques in sentimental analysis. In *Proceedings of the 2017 International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, India, 10–11 March 2017*, 216–221.

Carcillo, F., Borgne, Y. L., Caelen, O., Kessaci, Y., Oblé, F., & Bontempi, G. (2019). Combining unsupervised and supervised learning in credit card fraud detection. *Information Sciences, 557*, 317.

Carneiro, E. M., Dias, L. A. V., Da Cunha, A. M., & Mialaret, L. F. S. (2015). Cluster analysis and artificial neural networks: A case study in credit card fraud detection. In *2015 12th International Conference on Information Technology-New Generations* (pp. 122–126). IEEE.

Chalapathy, R., & Chawla, S. (2019). Deep learning for anomaly detection: A survey. arXiv (Cornell University).

Chang, V., Ali, B. A. A., Golightly, L., Ganatra, M. A., & Mohamed, M. (2024). Investigating credit card payment fraud with detection methods using advanced machine learning. *Information, 15*(8), 478.

Chang, V., Di Stefano, A., Sun, Z., & Fortino, G. (2022). Digital payment fraud detection methods in digital ages and Industry 4.0. *Computers and Electrical Engineering, 100*, 107734.

Dileep, M. R., Navaneeth, A. V., & Abhishek, M. (2021). A novel approach for credit card fraud detection using decision tree and random forest algorithms. In *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*, 1025–1028. IEEE.

Dornadula, V. N., & Geetha, S. (2019). Credit card fraud detection using machine learning algorithms. *Procedia Computer Science, 165*, 631.

Douiba, M., Benkirane, S., Guezzaz, A., & Azrour, M. (2023). An improved anomaly detection model for IoT security using decision tree and gradient boosting. *The Journal of Supercomputing, 79*(3), 3392–3411.

Du, H., Lv, L., Guo, A., & Wang, H. (2023). AutoEncoder and LightGBM for credit card fraud detection problems. *Symmetry, 15*(4), 870.

Faraji, Z., & States, U. (2022). A review of machine learning applications for credit card fraud detection with a case study. *J. Manag, 5*, 49–59.

Forough, J., & Momtazi, S. (2020). Ensemble of deep sequential models for credit card fraud detection. *Applied Soft Computing, 99*, 106883.

Guezzaz, A., Benkirane, S., Azrour, M., & Khurram, S. (2021). A reliable network intrusion detection approach using decision tree with enhanced data quality. *Security and Communication Networks, 2021*(1), 1230593.

Gupta, A., Lohani, M. C., & Manchanda, M. (2021). Financial fraud detection using naive bayes algorithm in highly imbalanced data set. *Journal of Discrete Mathematical Sciences and Cryptography, 24*(5), 1559–1572.

Gupta, A., Yousaf, A., & Mishra, A. (2020). How pre-adoption expectancies shape post-adoption continuance intentions: An extended expectation-confirmation model. *Int. J. Inf. Manag., 52*, 102094.

Hancock, J. T., & Khoshgoftaar, T. M. (2021). Gradient boosted decision tree algorithms for medicare fraud detection. *SN Computer Science, 2*(4), 268.

Hajek, P., Abedin, M. Z., & Sivarajah, U. (2023). Fraud detection in mobile payment systems using an XGBoost-based framework. *Information Systems Frontiers, 25*(5), 1985–2003.

Hassan, M. A., Shukur, Z., & Hasan, M. K. (2021). Electronic wallet payment system in Malaysia. *Data Analytics and Management*, Springer, 711–736.

Hung, W., Tseng, C., Chang, F., & Ho, C. (2021). Effects of utilitarian and hedonic emotion on the use of online banking services. *J. Glob. Inf. Manag., 29*(6), 1–20.

Ileberi, E., Sun, Y., & Wang, Z. (2022). A machine learning based credit card fraud detection using the GA algorithm for feature selection. *Journal of Big Data, 9*(1), 24.

Itoo, F., Meenakshi, & Singh, S. (2021). Comparison and analysis of logistic regression, Naïve Bayes and KNN machine learning algorithms for credit card fraud detection. *International Journal of Information Technology, 13*(4), 1503–1511.

Jan, C. L. (2021). Detection of financial statement fraud using deep learning for sustainable development of capital markets under information asymmetry. *Sustainability, 13*(17), 9879.

Jemima Jebaseeli, T., Venkatesan, R., & Ramalakshmi, K. (2021). Fraud detection for credit card transactions using random forest algorithm. In *Intelligence in Big Data Technologies – Beyond the Hype: Proceedings of ICBDCC 2019*, 189–197. Springer Singapore.

Jeragh, M., & Alsulaimi, M. (2018). Combining auto encoders and one class support vector machine for fraudulent credit card transactions detection. In *Proceedings of the 2018 Second World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4), London, UK, 30–31 October 2018*, 178–184.

Kabir, M. A., Saidin, S. Z., & Ahmi, A. (2015). Adoption of e-payment systems: A review of literature. In *International Conference on E-Commerce, Kuching, Sarawak*, 112–120.

Keskenler, M. F., Dal, D., & Aydın, T. (2021). Yapay zeka destekli ÇOKS yöntemi ile kredi kartı sahtekarlığının tespiti. *El-Cezeri Fen ve Mühendislik Dergisi*.

Khalid, A. R., Owoh, N., Uthmani, O., Ashawa, M., Osamor, J., & Adejoh, J. (2024). Enhancing credit card fraud detection: An ensemble machine learning approach. *Big Data and Cognitive Computing, 8*(1), 6.

Khando, K., Islam, M. S., & Gao, S. (2022a). Factors shaping the cashless payment ecosystem: Understanding the role of participating actors. In *35th Bled eConference-Digital Restructuring and Human (Re) action, Bled, Slovenia, June 26–29, 2022*, 161–186. University of Maribor University Press.

Khando, K., Islam, M. S., & Gao, S. (2022b). The emerging technologies of digital payments and associated challenges: A systematic literature review. *Future Internet, 15*(1), 21.

Lonkani, R., Changchit, C., Klaus, T., & Sampet, J. (2020). A comparative study of trust in mobile banking: An analysis of US and Thai customers. *J. Glob. Inf. Manag., 28*(4), 95–119.

Madhurya, M. J., Gururaj, H. L., Soundarya, B. C., Vidyashree, K. P., & Rajendra, A. B. (2022). Exploratory analysis of credit card fraud detection using machine learning techniques. *Global Transitions Proceedings, 3*(1), 31–37.

Makki, S., Haque, R., Taher, Y., Assaghir, Z., Hacid, M. S., & Zeineddine, H. (2018, December). A cost-sensitive cosine similarity K-nearest neighbor for credit card fraud detection. In *Big Data and Cyber-security Intelligence*.

Malaquias, R. F., Malaquias, F. F., Ha, Y. M., & Hwang, Y. (2021). A cross-country study on intention to use mobile banking: Does computer self-efficacy matter? *J. Glob. Inf. Manag., 29*(2), 102–117.

Malini, N., & Pushpa, M. (2017, February). Analysis on credit card fraud identification techniques based on KNN and outlier detection. In *2017 Third International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-informatics (AEEICB)* (pp. 255–258). IEEE.

Manek, H., Kataria, N., Jain, S., & Bhole, C. (2019). Various methods for fraud transaction detection in credit cards. *Journal of Ubiquitous Systems and Pervasive Networks, 12*(1), 25.

Mareeswari, V., & Gunasekaran, G. (2016). Prevention of credit card fraud detection based on HSVM. In *Proceedings of the 2016 International Conference on Information Communication and Embedded Systems (ICICES), Chennai, India, 25–26 February 2016*, 1–4.

Mehdary, A., Chehri, A., Jakimi, A., & Saadane, R. (2024). Hyperparameter optimization with genetic algorithms and XGBoost: A step forward in smart grid fraud detection. *Sensors, 24*(4), 1230.

Mehana, A., & Pireva, K. (2020). Fraud detection using data-driven approach. arXiv (Cornell University).

Meng, W., Wang, Y., Wong, D. S., Wen, S., & Xiang, Y. (2018). TouchWB: Touch behavioral user authentication based on web browsing on smartphones. *Journal of Network and Computer Applications, 117*, 1–9.

Ming-Yen Teoh, W., Choy Chong, S., Lin, B., & Wei Chua, J. (2013). Factors affecting consumers' perception of electronic payment: An empirical analysis. *Internet Research, 23*(4), 465–485.

Mishra, K. N., & Pandey, S. C. (2021). Fraud prediction in smart societies using logistic regression and k-fold machine learning techniques. *Wireless Personal Communications, 119*(2), 1341–1367.

Mhamane, S. S., & Lobo, L. M. R. J. (2012). Internet banking fraud detection using HMM. In *Proceedings of the 2012 Third International Conference on Computing, Communication and Networking Technologies (ICCCNT'12), Karur, India, 26–28 July 2012*, 1–4.

Modi, K., & Dayma, R. (2017). Computing and control (I2C2), review on fraud detection methods in credit card transactions. *2017 International Conference on Intelligent, IEEE*.

Namweli, H., & Magali, J. (2018). Factors affecting adoption of prepaid electronic payment cards in Tanzania: The case study of Kilimanjaro Christian Medical Centre (KCMC). *African Journal of Business Management, 2*(1), 1–14.

Naeem, A., Javaid, N., Aslam, Z., Nadeem, M. I., Ahmed, K., Ghadi, Y. Y., & Eldin, S. M. (2023). A novel data balancing approach and a deep fractal network with light gradient boosting approach for theft detection in smart grids. *Heliyon, 9*(9).

Nguyen, T. D., & Huynh, P. A. (2018). The roles of perceived risk and trust on E–payment adoption. In *International Econometric Conference of Vietnam, Springer, Cham*, 926–940.

Omigie, N. O., Zo, H., Ciganek, A. P., & Jarupathirun, S. (2020). Understanding the continuance of mobile financial services in Kenya: The roles of utilitarian, hedonic, and personal values. *J. Glob. Inf. Manag., 28*(3), 36–57.

Owolafe, O., Ogunrinde, O. B., & Thompson, A. F. B. (2021). A long short term memory model for credit card fraud detection. In *Artificial Intelligence for Cyber Security: Methods, Issues and Possible Horizons or Opportunities*, 369–391. Cham: Springer International Publishing.

Pan, E. (2024). Machine learning in financial transaction fraud detection and prevention. *Transactions on Economics Business and Management Research, 5*, 243.

Rajak, I., & Mathai, K. J. (2015). Intelligent fraudulent detection system based SVM and optimized by danger theory. In *Proceedings of the 2015 International Conference on Computer, Communication and Control (IC4), Indore, India, 10–12 September 2015*, 1–4.

Rzayeva, D., & Malekzadeh, S. (2022). A combination of deep neural networks and K-nearest neighbors for credit card fraud detection. arXiv (Cornell University).

Sadgali, I., Sael, N., & Benabbou, F. (2019). Performance of machine learning techniques in the detection of financial frauds. *Procedia Computer Science, 148*, 45–54.

Sadgali, I., Sael, N., & Benabbou, F. (2019). Fraud detection in credit card transaction using neural networks. *Proceedings of the 4th International Conference on Smart City Applications, 1*.

Sun, Q., Tang, T., Chai, H., Wu, J., & Chen, Y. (2021). Boosting fraud detection in mobile payment with prior knowledge. *Applied Sciences, 11*(10), 4347.

Shree, S., Pratap, B., Saroy, R., & Dhal, S. (2021). Digital payments and consumer experience in India: A survey based empirical study. *Journal of Banking and Financial Technology, 5*, 1–20.

Strelcenia, E., & Prakoonwit, S. (2023). A survey on GAN techniques for data augmentation to address the imbalanced data issues in credit card fraud detection. *Machine Learning and Knowledge Extraction, 5*(1), 304.

Tran, T. A. (2022). On some studies of fraud detection pipeline and related issues from the scope of ensemble learning and graph-based learning. arXiv (Cornell University).

Tripathi, K. K., & Pavaskar, M. A. (2012). Survey on credit card fraud detection methods. *International Journal of Emerging Technology and Advanced Engineering, 2*(11), 721–726.

Ugale, M., & Midhunchakkaravarthy, J. (2023). Machine learning-based image forgery detection using light gradient-boosting machine. In *Congress on Intelligent Systems*, 463–476. Singapore: Springer Nature.

Vassallo, D., Vella, V., & Ellul, J. (2021). Application of gradient boosting algorithms for anti-money laundering in cryptocurrencies. *SN Computer Science, 2*(3), 143.

Vanini, P., Rossi, S., Zvizdic, E., & Domenig, T. (2023). Online payment fraud: From anomaly detection to risk management. *Financial Innovation, 9*(1), 66.

Vimal, S., Kayathwal, K., Wadhwa, H., & Dhama, G. (2021). Application of deep reinforcement learning to payment fraud. arXiv (Cornell University).

Vishwakarma, M., & Kesswani, N. (2023). A new two-phase intrusion detection system with Naïve Bayes machine learning for data classification and elliptic envelope method for anomaly detection. *Decision Analytics Journal, 7*, 100233.

Wamba, S. F., Queiroz, M. M., Blome, C., & Sivarajah, U. (2021). Fostering financial inclusion in a developing country: Predicting user acceptance of mobile wallets in Cameroon. *J. Glob. Inf. Manag., 29*(4), 195–220.

Zarezadeh, M. R., Aboonajmi, M., & Ghasemi Varnamkhasti, M. (2021). Fraud detection and quality assessment of olive oil using ultrasound. *Food Science & Nutrition, 9*(1), 180–189.

Zhang, Y., & Trubey, P. (2019). Machine learning and sampling scheme: An empirical study of money laundering detection. *Computational Economics, 54*(3), 1043–1063.

Zhang, Q. (2022). Financial data anomaly detection method based on decision tree and random forest algorithm. *Journal of Mathematics, 2022*(1), 9135117.