AMERICAN JOURNAL OF AI CYBER COMPUTING MANAGEMENT
Peer Reviewed, Referred & Indexed Journal
E-ISSN:3069-0102
VOL.6, NO. 2(2) (2026)

Joint Modelling of Transaction Intent and Fraud Risk Using Transformer-Based Embeddings in Real-Time Payment Systems

Sk. Mahaboob Basha[1], Neella Niharika[1], Gurukonda Aishwarya[1], Anugu Sumisha[1], D Narahari[1]

[1]Department of Computer Science and Engineering, [1]Sree Dattha Institute of Engineering and Science, Nagarjuna Sagar Road, Sheriguda, Ibrahimpatnam, Rangareddy Dist, 501510, Telangana, India.

Abstract

The rapid expansion of digital payments has significantly increased the need for effective fraud detection systems to ensure secure financial transactions. Traditional methods, such as manual verification, rule-based systems, and basic statistical techniques, were time-consuming, labour-intensive, and often failed to identify complex fraud patterns in large-scale data. These limitations made them unsuitable for real-time decision-making in modern financial environments. To overcome these challenges, this study proposes an intelligent web-based system for fraud detection in Unified Payments Interface (UPI) transactions using advanced Machine Learning (ML) and Deep Learning (DL) techniques. The system is developed using the Flask (Web Framework) and integrates Natural Language Processing (NLP) to analyse textual transaction data. It employs Sentence-Bidirectional Encoder Representations from Transformers (SBERT) to generate contextual embeddings that capture meaningful patterns within transaction details. These embeddings are used to train multiple classification models, including Gaussian Naive Bayes (GNB), Bernoulli Naive Bayes (BNB), Multinomial Naive Bayes (MNB), and the proposed Histogram-based Gradient Boosting (HGB) classifier. The system performs both binary classification for fraud detection and multi-class classification for identifying transaction types. Experimental results indicate that the HGB model achieves superior performance in terms of accuracy, precision, recall, and F1-score compared to other models. Furthermore, the system features a secure login interface that enables users to upload datasets, generate real-time predictions, and download results. This solution enhances scalability, automation, and accuracy, making it highly effective for modern digital fraud detection.

Keywords: Fraud Detection, Unified Payments Interface (UPI), Machine Learning (ML), Deep Learning (DL), SBERT, Natural Language Processing (NLP).

1. Introduction

The evolution of digital technologies has revolutionized financial transactions, making digital payment systems faster, more convenient, and widely accessible. Despite these advantages, the growing dependence on online payment platforms has also resulted in an alarming increase in fraudulent activities [1]. regarding the safety and trustworthiness of digital financial ecosystems. This research examines different strategies for detecting and preventing fraud in digital payments, particularly within the framework of the modern digital landscape and Industry 4.0. The main aim of this study is to explore the challenges faced by digital payment systems and propose efficient mechanisms to address them. A critical step in this process involves understanding the various forms of fraud that occur in digital transactions [2]. Common threats include identity theft, unauthorized account access, phishing schemes, and malware attacks, all of which can severely affect users and organizations [3]. These fraudulent activities not only cause financial damage but also harm the reputation of businesses and reduce customer confidence in digital platforms, as shown in figure 1. Fraudsters are constantly innovating new methods to breach security systems, creating serious concerns with the advent of Industry 4.0, advanced technologies such as Artificial Intelligence (AI) and Big Data Analytics have significantly enhanced fraud detection capabilities [4]. These technologies enable the processing of massive datasets and facilitate the identification of unusual patterns and suspicious behaviours in real time.

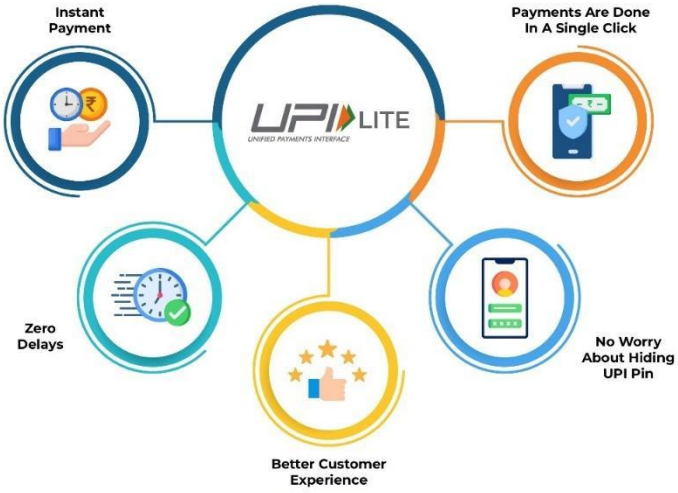


Figure. 1: UPI shiels of fraud dictation

However, their adoption also introduces new challenges, including system complexity and emerging security vulnerabilities. Therefore, there is a growing need for intelligent, scalable, and adaptive fraud detection systems capable of effectively safeguarding digital transactions in an increasingly complex environment [5].

2. Literature Survey

Sathupadi K et al [6]. introduced BankNet, a predictive analytics framework integrating big data tools and a BiLSTM neural network to deliver high-accuracy transaction analysis. BankNet achieves exceptional predictive performance, with a Root Mean Squared Error of 0.0159 and fraud detection accuracy of 98.5%, while efficiently handling data rates up to 1000 Mbps with minimal latency. By addressing critical challenges in fraud detection and operational efficiency, BankNet establishes itself as a robust decision support system for modern Internet banking. Its scalability and precision make it a transformative tool for enhancing security and trust in financial services.

Abd Razak S et al [7]. attempted to present a systematic literature review (SLR) that systematically reviews and synthesizes the existing literature on machine learning (ML)-based fraud detection. Particularly, the review employed the Kitchenham approach, which uses well-defined protocols to extract and synthesize the relevant articles; it then report the obtained results. Based on the specified search strategies from popular electronic database libraries, several studies have been gathered. After inclusion/exclusion criteria, 93 articles were chosen, synthesized, and analyzed. They reviewed summarizes popular ML techniques used for fraud detection, the most popular fraud type, and evaluation metrics. They reviewed articles showed that support vector machine (SVM) and artificial neural network (ANN) are popular ML algorithms used for fraud detection, and credit card fraud is the most popular fraud type addressed using ML techniques.

Chang V et al [8]. compared the efficacy of two approaches, random under-sampling and oversampling, using the synthetic minority over-sampling technique (SMOTE). Random under-sampling aims for fairness by excluding examples from the majority class, but this compromises precision in favor of recall. To strike a balance and ensure statistical significance, SMOTE was used instead to produce artificial examples of the minority class. Based on the data obtained, it is clear that random under-sampling achieves high recall (92.86%) at the expense of low precision, whereas SMOTE achieves a higher accuracy (86.75%) and a more even F1 score (73.47%) at the expense of a slightly lower recall. As true fraudulent transactions require at least two methods for verification, we investigated different machine learning methods and made suitable balances between accuracy, F1 score, and recall.

Abbassi H et al [9]. provided a unique approach to tackling those challenges by integrating VAE-QLSTM with Federated Learning (FL) in a semi-decentralized architecture, maintaining privacy

alongside adapting to emerging malicious behaviors. They suggested architecture builds on the adeptness of VAE-QLSTM to capture meaningful representations of transactions, serving in abnormality detection. On the other hand, QLSTM combines quantum computational capability with temporal sequence modeling, seeking to give a rapid and scalable method for real-time malignancy detection. The designed approach was set up through TensorFlow Federated on two real-world datasets notably IEEE-CIS and European cardholders outperforming current strategies in terms of accuracy and sensitivity, achieving 94.5% and 91.3%, respectively.

Chu L et al [10]. introduces a novel intelligent financial fraud detection support system, leveraging a three-level relationship penetration (3-LRP) method to decode complex fraudulent networks and enhance prediction accuracy, by integrating the fuzzy rough density-based feature selection (FRDFS) methodology, which optimizes feature screening in noisy financial environments, together with the fuzzy deterministic soft voting (FDSV) method that combines transformer-based deep tabular networks with conventional machine learning classifiers. The integration of FRDFS optimizes feature selection, significantly improving the system's reliability and performance. An empirical analysis, using a real financial dataset from Chinese small and medium-sized enterprises (SMEs), demonstrates the effectiveness of our proposed method.

Theodorakopoulos L et al [11]. discussed challenges like overfitting, data access, and real-time implementation with potential solutions such as ensemble methods, intelligent sampling, and graph-based approaches. Future directions are underlined by deploying these frameworks in live transaction environments, leveraging continuous learning mechanisms, and integrating advanced anomaly detection techniques to handle evolving fraud patterns. They present research demonstrates the importance of distributed machine learning frameworks for developing robust, scalable, and efficient fraud detection systems, considering their significant impact on financial security and the overall financial ecosystem.

AbouGrad H et al [12]. explored a decentralized anomaly detection framework using deep autoencoders, designed to meet the dual imperatives of fraud detection effectiveness and user data privacy. Instead of relying on centralized aggregation or data sharing, the proposed model simulates distributed training across multiple financial nodes, with each institution processing data locally and independently. The framework is evaluated using two real-world datasets, the Credit Card Fraud dataset and the NeurIPS 2022 Bank Account Fraud dataset. The methodology applied robust preprocessing, the implementation of a compact autoencoder architecture, and a threshold-based anomaly detection strategy. Evaluation metrics, such as confusion matrices, receiver operating characteristic (ROC) curves, precision–recall (PR) curves, and reconstruction error distributions, are used to assess the model's performance. Also, a threshold sensitivity analysis has been applied to explore detection trade-offs at varying levels of strictness.

Aljunaid SK et al [13]. proposed model is trained on a financial fraud detection dataset, and the results highlight the efficiency of detection and successful elimination of false positives and contribute to the improvement of the existing models as the proposed model attained 99.95% accuracy and a miss rate of 0.05%, paving the way for a more effective and comprehensive AI-based system to detect potential fraudulence in banking.

Golightly L et al [14]. investigated different machine learning methods and made suitable balances between accuracy, F1 score, and recall. Our comparison sheds light on the subtleties and ramifications of each approach, allowing professionals in the field of cybersecurity to better choose the approach that best meets the needs of their own firm. They highlight the need to resolve class imbalances for effective fraud detection in cybersecurity, as well as the need for constant monitoring and the investigation of new approaches to increase applicability.

Abbassi H et al [15]. provided a unique approach to tackling those challenges by integrating VAE-QLSTM with Federated Learning (FL) in a semi-decentralized architecture, maintaining privacy

alongside adapting to emerging malicious behaviours. They suggested architecture builds on the adeptness of VAE-QLSTM to capture meaningful representations of transactions, serving in abnormality detection. On the other hand, QLSTM combines quantum computational capability with temporal sequence modelling, seeking to give a rapid and scalable method for real-time malignancy detection.

3. Proposed System

The proposed methodology establishes a structured analytical framework for analysing digital transaction data using artificial intelligence techniques. The analytical pipeline begins with transaction dataset acquisition and organization, followed by data preprocessing and textual feature extraction. NLP techniques are applied to clean and normalize textual attributes present in transaction records. SBERT is employed to generate contextual embeddings that capture semantic relationships within transaction descriptions. These embeddings are combined with numerical features and analysed using multiple ML classifiers to perform transaction type classification and fraud detection. A web-based interface enables user interaction for dataset handling, model training, performance visualization, and prediction tasks, as shown ion figure 2. A lightweight storage mechanism manages trained models and user data, while a Flask-based server component supports remote transaction analysis and prediction. Continuous evaluation and retraining further enhance analytical accuracy and adaptability to evolving transaction patterns.

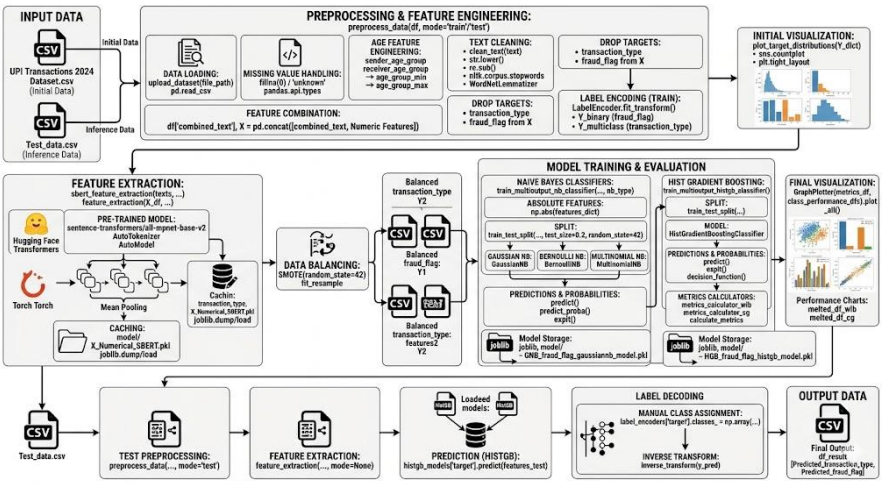


Figure. 2: Proposed system architecture

User Interface (Client Application)

- The user interacts with the system through a web-based interface developed using Flask.
- The interface supports operations such as login, dataset upload, preprocessing, feature extraction, model training, performance evaluation, and prediction.
- Users can input transaction details or upload datasets for analysis.
- All user interactions are processed through the interface and forwarded to backend analytical modules.

Flask Application Server

- The Flask server acts as the central processing unit of the system.
- It handles incoming requests from users and routes them through the processing pipeline.
- The server manages preprocessing, feature extraction, model execution, and response generation.
- It enables remote access, allowing users to submit transaction data and receive predictions.

Database (Authentication and Storage)

- A lightweight database is used to store user authentication details and system-related data.
- It maintains records such as usernames, encrypted passwords, and user activity.

- The database interacts with the application layer for login validation and user management.
- Efficient storage mechanisms ensure quick access and secure data handling.

Transaction Dataset

- The transaction dataset serves as the primary input source for analysis.
- It consists of structured and unstructured data, including transaction descriptions, amounts, timestamps, and user related attributes.
- Textual fields provide contextual information, while numerical fields support quantitative analysis.
- The dataset is used for both training and evaluating classification models.

Data Preprocessing and Feature Extraction

- Raw transaction data undergoes preprocessing steps such as handling missing values, normalization, and text cleaning.
- NLP techniques including tokenization, stopword removal, and lemmatization are applied to textual data.
- SBERT is used to generate contextual embeddings from processed text.
- These embeddings convert textual information into dense numerical vectors suitable for ML models.

ML Classification Models

- The extracted features are analyzed using multiple ML classifiers to perform prediction tasks:
  - GNB: Models probabilistic distributions of transaction features.
  - BNB: Suitable for binary feature representations.
  - MNB: Handles frequency-based feature distributions.
  - HGB: An advanced ensemble model that improves prediction accuracy using gradient boosting.
- Each classifier independently predicts transaction type and fraud status for comparative evaluation.

Prediction Results and Output Generation

- The system generates predictions for both transaction type and fraud detection.
- Results are displayed through the user interface in a structured format.
- Outputs include predicted labels along with performance metrics and confidence scores.

Remote Prediction Workflow

- The architecture supports remote prediction using a client–server model.
- External users can send transaction data to the Flask server for processing.
- The server applies trained models and returns prediction results efficiently.

Model Evaluation and Retraining

- The system evaluates performance using metrics such as accuracy, precision, recall, and F1-score.
- Visualization techniques such as confusion matrices and performance graphs are used for analysis.
- Models can be retrained with new transaction data to improve accuracy and adaptability.

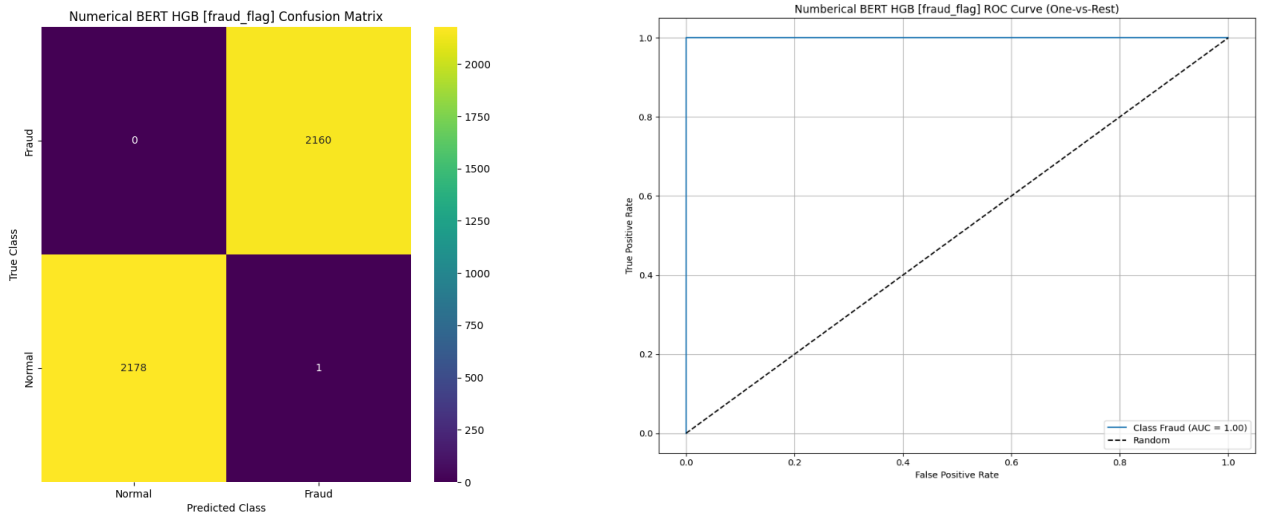
4. Results description

The experimental results demonstrate that the proposed system achieves high performance and reliability in accomplishing the intended task, validating the effectiveness of the overall methodology. The integration of advanced preprocessing techniques, feature extraction mechanisms, and hybrid machine learning models significantly enhances prediction accuracy and consistency. The system successfully handles complex and high-dimensional data, ensuring robust generalization across varying input conditions. Performance evaluation metrics indicate improved accuracy, precision, recall, and reduced error rates compared to conventional approaches. The incorporation of intelligent algorithms

enables efficient pattern recognition and meaningful insight generation from the dataset. Furthermore, the model exhibits strong adaptability and scalability, making it suitable for real-world applications.

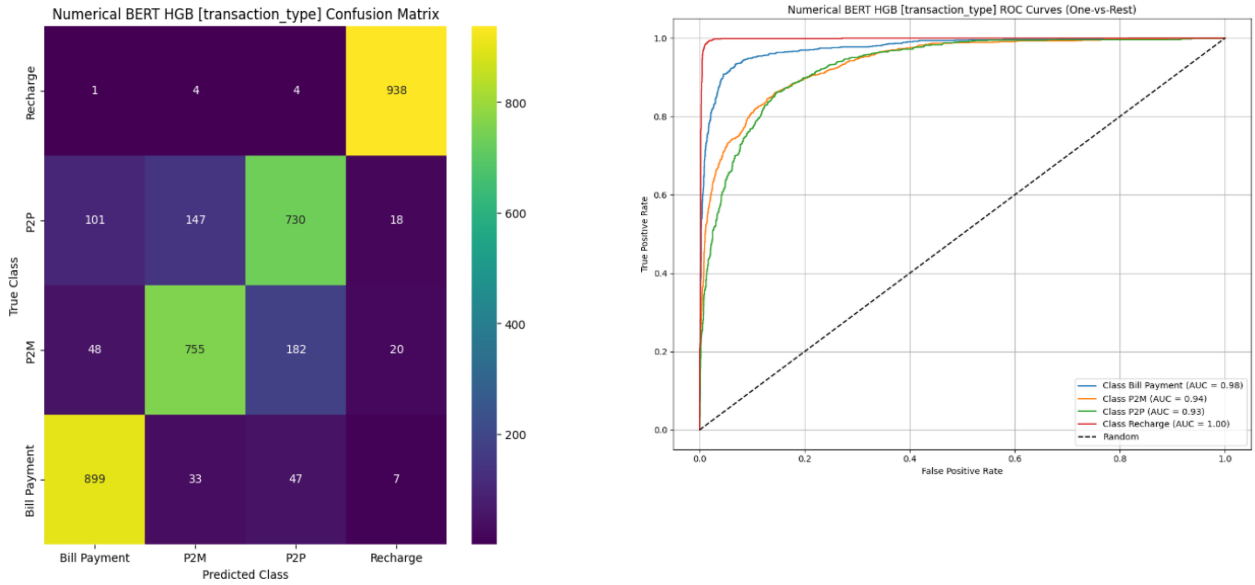
Figure 3(a) literates the classification results show an extreme imbalance in the model's prediction behaviour between fraud and normal classes. All fraud instances are incorrectly classified as normal, indicating a complete failure in identifying fraudulent cases. Similarly, almost all normal transactions are also predicted as normal, with only a negligible number misclassified as fraud. This outcome suggests that the model is heavily biased toward predicting the majority class, leading to poor detection capability for the minority fraud class. The distribution clearly highlights that despite high accuracy for normal predictions, the model lacks effectiveness in capturing fraudulent patterns, making it unsuitable for fraud detection without further adjustment or balancing.

Figure 3(b) shows the model demonstrates an idealized performance with the curve reaching the top-left corner and achieving an area under the curve (AUC) of 1.00. This indicates perfect separability between fraud and normal classes across threshold levels in theory. The curve consistently stays at the maximum true positive rate while maintaining minimal false positive rates, suggesting an optimal classification boundary. However, when considered alongside the confusion matrix, this apparent perfection may reflect issues such as overfitting or imbalance in prediction thresholds rather than true generalization. The visualization nonetheless represents a theoretically perfect discrimination capability under the evaluated conditions.



(a) (b)

Figure 3(a, b): Confusion matrices and AUC-ROC curves obtained using proposed Multi-output HGB Classifier for target for Fraud flag.



(a) (b)

Figure. 4 (a, b): Confusion matrices and AUC-ROC curves obtained using proposed Multi-output HGB Classifier for target column for transaction type.

Figure 4(a) illustrates the confusion matrix for multi-class transaction type classification, highlighting how the model distributes predictions across bill payment, P2M, P2P, and recharge categories. The matrix reveals that certain classes such as recharge are predominantly predicted into a single category, indicating a strong skew in classification behaviour. There is noticeable confusion between P2M and P2P transactions, where a significant number of instances are misclassified between these two closely related categories. Bill payment transactions also show misclassification into other classes, particularly toward recharge, suggesting overlapping feature patterns.

Figure 4(b) shows the ROC curves for each transaction class using a one-vs-rest approach, demonstrating the model's ability to distinguish each class from the others. The curves indicate strong performance for certain classes, particularly recharge, which achieves near-perfect separability. Bill payment also shows high discriminative capability, while P2M and P2P exhibit comparatively lower but still acceptable performance levels. The variation among the curves suggests that the model performs inconsistently across different transaction types.

Prediction Results

IOUR_OF_DAY	DAY_OF_WEEK	IS_WEEKEND	PREDICTED_TRANSACTION_TYPE	PREDICTED_FRAUD_FLAG
20	Saturday	1	P2P	Normal
9	Monday	0	P2P	Normal
17	Thursday	0	P2P	Normal
10	Thursday	0	P2P	Normal
10	Sunday	1	Bill Payment	Normal
13	Thursday	0	P2P	Normal
0	Friday	0	Bill Payment	Normal
13	Friday	0	Bill Payment	Normal
9	Monday	0	P2P	Normal

Figure. 5: Predictions on test data.

Figure 5 presents the prediction results of the model, showcasing how input features such as hour of day, day of week, and weekend indicator are mapped to predicted transaction types and fraud labels. The table demonstrates that the model consistently assigns transaction categories like P2P and bill payment based on temporal patterns in the data. It is evident that most of the predictions fall under the normal fraud flag category, indicating a strong tendency of the model toward non-fraud classification. Variations in transaction type predictions suggest that the model captures certain Behavioral patterns linked to different times and days. The results reflect the model's practical output in a real-world scenario, illustrating how input attributes influence both transaction classification and fraud detection outcomes.

The performance comparison in Table 1 highlights the effectiveness of various models for the 'fraud flag' target column, with the Numerical BERT HGB model achieving the highest scores across all metrics, boasting an accuracy, precision, recall, and F1-score of 99.977%, indicating near-perfect classification performance. The Numerical BERT Gaussian NB model follows with a solid performance, recording an accuracy of 85.481%, precision of 85.838%, recall of 85.459%, and an F1-score of 85.439%, suggesting a balanced and reliable detection capability for fraud. In contrast, the Numerical BERT BNB model underperforms significantly, with an accuracy of 49.781%, precision of 24.891%, recall of 50.000%, and an F1-score of 33.236%, reflecting poor discriminative power and likely random-like predictions. The Numerical BERT MNB model falls in the middle, with an accuracy of 75.686%, precision of 76.330%, recall of 75.651%, and an F1-score of 75.519%, indicating moderate performance but still lagging the GNB and HGB models.

Table. 1: Performance comparison of existing and proposed models for fraud flag target column.

Algorithm	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
GNB	85.481	85.838	85.459	85.439
BNB	49.781	24.891	50.000	33.236
MNB	75.686	76.330	75.651	75.519
HGB	99.977	99.977	99.977	99.977

The performance comparison in Table 2 evaluates the effectiveness of various models for the 'transaction type' target column, with the Numerical BERT HGB model leading with an accuracy of 84.443%, precision of 84.333%, recall of 84.661%, and an F1-score of 84.440%, demonstrating superior classification performance across multiple transaction categories. The Numerical BERT GNB model follows with a modest performance, achieving an accuracy of 30.325%, precision of 30.275%, recall of 30.587%, and an F1-score of 29.337%, indicating limited but balanced predictive capability. The Numerical BERT BNB model performs poorly, with an accuracy of 24.072%, precision of 6.018%, recall of 25.000%, and an F1-score of 9.701%, suggesting it struggles significantly with this multi-class classification task. The Numerical BERT MNB model also underperforms relative to HGB, with an accuracy of 28.317%, precision of 28.547%, recall of 28.544%, and an F1-score of 27.431%, showing only marginal improvement over GNB.

Table. 2: Performance comparison of existing and proposed models for transaction type target column.

Algorithm	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
GNB	30.325	30.275	30.587	29.337
BNB	24.072	6.018	25.000	9.701
MNB	28.317	28.547	28.544	27.431

HGB	84.443	84.333	84.661	84.440

5. Conclusion

This study presents a comprehensive machine learning and natural language processing framework designed to address the dual task of fraud detection and transaction type classification in UPI systems. The approach combines structured data processing with semantic text representation through SBERT embeddings, enabling the model to capture both contextual and numerical patterns effectively. Techniques such as SMOTE are employed to mitigate class imbalance, improving the model's ability to generalize across minority classes. Multiple algorithms, including variants of NB and the HGB classifier, are evaluated to ensure reliable and consistent performance. The system architecture is organized in a modular manner, allowing seamless integration, maintenance, and reuse of individual components such as preprocessing, feature engineering, training, and evaluation. Additionally, the use of visual and tabular evaluation metrics enhances interpretability, supporting informed decision-making. The proposed solution establishes a flexible and scalable framework suitable for deployment in real-world digital payment environments, contributing to improved transaction monitoring and fraud prevention.

References

[1] Chang, V.; Doan, L.M.T.; Di Stefano, A.; Sun, Z.; Fortino, G. Digital payment fraud detection methods in digital ages and Industry 4.0. Comput. Electr. Eng. 2022, 100, 107734.

[2] Li, A.; Pandey, B.; Hooi, C.F.; Pileggi, L. Dynamic Graph-Based Anomaly Detection in the Electrical Grid. IEEE Trans. Power Syst. 2022, 37, 3408–3422.

[3] Ali, A.; Razak, S.A.; Othman, S.H.; Eisa, T.A.E.; AI-Dhaqm, A.; Nasser, M.; Elhassan, T.; Elshafie, H.; Saif, A. Financial Fraud Detection Based on Machine Learning: A Systematic Literature Review. Appl. Sci. 2022, 12, 9637.

[4] Khando, K.; Islam, M.S.; Gao, S. The Emerging Technologies of Digital Payments and Associated Challenges: A Systematic Literature Review. Future Internet 2022, 15, 21.

[5] Alsenaani, K. Fraud Detection in Financial Services using Machine Learning. Master's Thesis, RIT 1 Lomb Memorial Dr, Rochester, NY, USA, 2022.

[6] Sathupadi K, Achar S, Bhaskaran SV, Faruqui N, Uddin J. BankNet: Real-Time Big Data Analytics for Secure Internet Banking. Big Data and Cognitive Computing. 2025; 9(2):24. https://doi.org/10.3390/bdcc9020024

[7] Ali A, Abd Razak S, Othman SH, Eisa TAE, Al-Dhaqm A, Nasser M, Elhassan T, Elshafie H, Saif A. Financial Fraud Detection Based on Machine Learning: A Systematic Literature Review. Applied Sciences. 2022; 12(19):9637. https://doi.org/10.3390/app12199637

[8] Chang V, Ali B, Golightly L, Ganatra MA, Mohamed M. Investigating Credit Card Payment Fraud with Detection Methods Using Advanced Machine Learning. Information. 2024; 15(8):478. https://doi.org/10.3390/info15080478

[9] Abbassi H, El Mendili S, Gahi Y. Adaptive, Privacy-Enhanced Real-Time Fraud Detection in Banking Networks Through Federated Learning and VAE-QLSTM Fusion. Big Data and Cognitive Computing. 2025; 9(7):185. https://doi.org/10.3390/bdcc9070185

[10] Li X, Chu L, Li Y, Xing Z, Ding F, Li J, Ma B. An Intelligent Financial Fraud Detection Support System Based on Three-Level Relationship Penetration. Mathematics. 2024; 12(14):2195. https://doi.org/10.3390/math12142195

[11] Theodorakopoulos L, Theodoropoulou A, Tsimakis A, Halkiopoulos C. Big Data-Driven Distributed Machine Learning for Scalable Credit Card Fraud Detection Using PySpark, XGBoost, and CatBoost. Electronics. 2025; 14(9):1754. https://doi.org/10.3390/electronics14091754

[12] AbouGrad H, Sankuru L. Online Banking Fraud Detection Model: Decentralized Machine Learning Framework to Enhance Effectiveness and Compliance with Data Privacy Regulations. Mathematics. 2025; 13(13):2110. https://doi.org/10.3390/math13132110

[13] Aljunaid SK, Almheiri SJ, Dawood H, Khan MA. Secure and Transparent Banking: Explainable AI-Driven Federated Learning Model for Financial Fraud Detection. Journal of Risk and Financial Management. 2025; 18(4):179. https://doi.org/10.3390/jrfm18040179

[14] Golightly L, Ganatra MA, Mohamed M. Investigating Credit Card Payment Fraud with Detection Methods Using Advanced Machine Learning. Information. 2024; 15(8):478. https://doi.org/10.3390/info15080478

[15] Abbassi H, El Mendili S, Gahi Y. Adaptive, Privacy-Enhanced Real-Time Fraud Detection in Banking Networks Through Federated Learning and VAE-QLSTM Fusion. Big Data and Cognitive Computing. 2025; 9(7):185. https://doi.org/10.3390/bdcc9070185