Multi-Modal Fraud Detection in Digital Transactions

Rajesh Nasre[1], Rehan khan[2], Rushikesh Bhoyar[3], Roshan Deotale[4], Rajas Deshpande[5], Samiksha Gole[6], Revti Nimje[7]

[1]*Assistant Professor of Artificial Intelligence Department of Artificial Intelligence G H Raisoni College of Engineering and Management Nagpur, India*

[2,3,4,5,6,7] *Department of Artificial Intelligence G H Raisoni College of Engineering and Management, Nagpur, India*

doi.org/10.64643/IJIRTV12I5-185774-459

***Abstract*—This research presents a real-time fraud detection framework that leverages heterogeneous data sources to enhance detection accuracy and system reliability. Departing from conventional single-dataset approaches, the proposed system synthesizes transaction records, behavioral analytics, network topology, geospatial intelligence, and textual data to identify sophisticated fraud patterns. Individual data stream undergo rigorous preprocessing including data cleaning, feature normalization, and attribute extraction prior to classification through ensemble machine learning architectures comprising Random Forest, XGBoost, HistGradient, SVM fusion strategies (weighted aggregation and meta-learning), and key features (real-time processing, stability, modularity) It provides a practical solution for fraud mitigation in evolving digital transaction ecosystems, The architecture demonstrates scalability for high-volume streaming data environments, while its modular design facilitates seamless integration of additional data modalities and algorithmic components.**

I. INTRODUCTION

The exponential expansion of digital payment ecosystems, online retail platforms, and internet banking services has fundamentally transformed global commerce while simultaneously creating vulnerabilities exploited by increasingly complex fraudulent schemes. As transaction volumes surge and cybercriminals deploy adaptive attack methodologies, traditional rule-based detection mechanisms prove inadequate for contemporary threat landscapes. This research presents a multimodal fraud detection architecture that leverages real-time data stream analysis to identify and neutralize suspicious activities instantaneously, thereby substantially reducing financial exposure and enhancing customer protection protocols. The system's real-time processing capability ensures immediate threat identification, enabling organizations to intervene before fraudulent transactions complete their lifecycle.

By integrating diverse data sources, including behavioral analytics, transaction metadata, network topology, geospatial intelligence, and textual communications, the proposed framework captures multidimensional fraud signatures that evade conventional single-modality detection approaches. The architecture's inherent scalability positions it for enterprise-grade deployment across diverse sectors.

II. EASE OF USE

A. User-Friendly Design and Automation:

The system employs a modular pipeline structure with five independent data modalities (Grap, Transaction, Location, Biometric, Text), allowing organizations to selectively implement modules based on their specific Needs without requiring a complete system overhaul. This plug-and-play approach simplifies deployment and reduces implementation complexity. The system implements a weighted ensemble strategy with pre-configured optimal weights (Graph:99%, Text:99%, Transaction/Location:81%, Behavioral:69%), allowing deployment without extensive hyperparameter tuning. The fusion layer automatically harmonizes 139,007 samples using weighted averaging and meta-learning.

The framework provides 4-tier automatic risk classification (Low/Medium/High/Critical) with 0-1 probability scores, delivering actionable insights without requiring manual threshold configuration or decision-making expertise.

B. Transparency, Security, and Efficiency:

The system provides interpretable outputs through

comprehensive visualization tools, including confusion matrices, ROC curves, and permutation importance analysis. These visualizations enable non-technical stakeholders to understand detection decisions and validate system behavior. Each modality's performance is independently evaluated and documented with clear AUC metrics (Graph:1000, Text:0.990, Behavioral:0.690, Tacked Traction/location:0.811), allowing users to identify strengths and weaknesses within their specific deployment context. The architecture is optimized for real-time deployment with automated risk scoring that generates immediate probability assessments. This enables organizations to flag suspicious activities instantaneously, minimizing financial losses while maintaining system responsiveness

III. LITERATURE REVIEW

The proliferation of digital payment ecosystems has catalyzed extensive research into fraud detection methodologies, with recent scholarship demonstrating a paradigmatic shift from unimodal to multimodal analytical frameworks. Contemporary investigations reveal that integrating heterogeneous data sources substantially enhances detection capabilities while addressing limitations inherent in single-modality approaches.

[1] Zhang, L., Kumar, S., Wang, M., and Chen (2023) proposed a multimodal deep learning framework employing a Graph Convolutional Network integrated with LSTM and attention mechanism, processing transaction graph, behavioral sequences, and device fingerprints for enhanced financial fraud detection, through the approach exhibited high computational overhead requiring extensive feature engineering.

[2] Roy, S., Sinha, M., and Das, T. (2025) conducted a comprehensive benchmarking of CNN-LSTM, GNN, and Transformer architectures with fairness metrics, proposing bias mitigation strategies for ethical AI deployment in financial fraud detection systems

[3] Bajpai, S., Bhattacharya, A., Vatsa, A., and Singh, A. (2024) conducted a comparative evaluation of anomaly detection methods for fraud detection in online credit card payments, examining various algorithmic approaches across different fraud scenarios.

[4] Abinaya, T.J., Vigneshwar, D., and Devi, N. (2023) developed explainable machine learning frameworks optimized for real-time payment fraud detection, balancing interpretability with processing speed requirements for operational deployment.

[5] Li, X., Wang, Y., and Zhang, J. (2025) proposed a Layer Weighted-GCN methodology for detecting fraudulent transactions across different pattern types in financial networks, demonstrating the importance of pattern-specific detection strategies.

IV METHODOLOGY

The proposed multimodal fraud detection framework employs a systematic architecture integrating five heterogeneous data modalities through ensemble machine learning methodologies. The system architecture comprises data acquisition, preprocessing pipelines, and individual modality classification, fusion layer integration, and final risk assessment components.

Data Modality Processing:

Graph Modality: Network relationships are represented as graph structures capturing connections between entities, transactions, and accounts. Graph preprocessing involves node feature extractions, edge weight computation, and adjacency matrix construction. The modality employs Random Forest, XGBoost, and HistGradientBoosting classifiers for pattern recognition, leveraging structural properties to identify fraudulent network formations.

Transaction Modality: Financial transaction data encompasses monetary values, timestamps, merchant categories, payment methods, and account identifiers. Preprocessing operations include missing values imputation, outlier detection, and temporal feature engineering, and numerical normalization using StandardScaler. Classification employs ensemble algorithms capable of handling an imbalanced dataset and capturing non-linear transaction patterns.

Location Modality: The Location Modality processes geospatial intelligence comprising GPS coordinates, IP addresses, device locations, and velocity calculations between consecutive transactions. Feature engineering operations generate distance metrics quantifying spatial displacement, velocity anomaly scores identifying impossible travel patterns, and geographical clustering indicators revealing transaction concentration zones. Preprocessing

employs StandardScaler for coordinate normalization and OneHotEncoder for categorical location variables such as country codes and address regions. The modality utilizes a HistGradientBoosting classifier to identify location-based fraud patterns, including impossible travel scenarios where transaction locations are geographically inconsistent with temporal constraints, and geographic anomalies characterized by transactions originating from high-risk jurisdictions or exhibiting unusual cross-border movement patterns.

Behavioral Modality: - The Behavioral Modality processes 1,760 keystroke dynamics samples from 88 JSON files (880 genuine, 880 fraudulent) to extract 42 engineered features including dwell_avg (key press duration), flight_avg (inter-key timing), traj_avg (movement trajectories), typing rhythm consistency, and pressure variation patterns. After StandardScaler normalization and statistical profiling, Random Forest classifier analyzes temporal sequences to detect deviations from baseline behavioral patterns such as abnormal typing speeds or inconsistent keystroke pressure, achieving an AUC of 0.6903

Text Modality: - The Text Modality processes 5,572 SMS messages (4,825 ham, 747 spam) from the SMS Spam Collection Dataset, with 1,551 samples utilized after preprocessing for fraud pattern analysis. Natural language processing operations include lowercase conversion, punctuation removal, stop word filtering, and text vectorization using TfidfVectorizer to transform unstructured textual data from transaction descriptions, communication logs, and merchant information into numerical feature representations. The hybrid feature extraction combines binary encoding (0=Ham, 1=Spam) with TF-IDF weighted term frequencies, generating sparse feature vectors capturing linguistic patterns associated with fraudulent communications. Classification employs Linear Support Vector Machine (SVM) with linear kernel, optimizing hyperplane separation between legitimate and fraudulent text patterns, achieving an AUC score of 0.9900 and demonstrating superior discriminative capability for text-based fraud detection.

V PROBLEM STATEMENT

Contemporary fraud detection systems deployed in financial institutions and digital payment platforms predominantly rely on single-modality analytical approaches, limiting their capacity to identify sophisticated fraud patterns that manifest across multiple data dimensions. Traditional unimodal frameworks analyzing exclusively transactional features, behavioral characteristics, or textual communications demonstrate insufficient detection capabilities when fraudulent activities exploit multiple attack vectors simultaneously. This singular dependency creates critical vulnerabilities as fraudsters increasingly employ composite tactics that combine account manipulation, transaction anomalies, location spoofing, and communication deception to circumvent conventional detection mechanisms.

Existing detection architectures exhibit several fundamental limitations that compromise their effectiveness in dynamic threat environments. First, single-modality systems cannot capture the multidimensional nature of modern fraud schemes that span behavioral deviations, geographic inconsistencies, network relationship anomalies, and linguistic fraud indicators occurring concurrently within the same fraudulent transaction. Second, current approaches demonstrate limited adaptability to evolving fraud tactics, requiring frequent manual recalibration and rule updates as attackers modify their methodologies to evade detection patterns. Third, conventional models struggle with the inherent class imbalance problem where fraudulent transactions constitute less than 1% of total transaction volumes, leading to high false-negative rates that allow sophisticated fraud to remain undetected while generating excessive false-positive alerts that burden investigation resources.

The absence of integrated multimodal frameworks that synthesize heterogeneous data sources through ensemble learning methodologies represents a critical gap in fraud detection research. Organizations require scalable, accurate, and interpretable detection systems capable of processing transaction records, behavioral analytics, network topology graphs, geospatial intelligence, and textual communications simultaneously while maintaining real-time processing capabilities for high-velocity transaction environments. Furthermore, existing solutions lack the modularity necessary for selective implementation based on organizational data availability and infrastructure constraints, limiting widespread adoption across diverse operational contexts.

This research addresses these limitations by developing a Python-based multimodal fraud

detection framework that integrates five independent data modalities Graph relationships, Transaction features, Location intelligence, Behavioral biometrics, and Text communications through ensemble machine learning architectures employing Random Forest, XGBoost, HistGradientBoosting, and Support Vector Machine classifiers. The proposed system implements weighted fusion and stacking strategies to harmonize individual modality predictions, generating comprehensive fraud risk assessments with four-tier categorization (Low/Medium/High/Critical) while ensuring scalability, interpretability, and real-time responsiveness for practical deployment in production environments.

VI OBJECTIVES

The primary aim of this research is to develop and validate a comprehensive multimodal fraud detection framework that addresses the limitations of conventional single-modality approaches through intelligent integration of heterogeneous data sources. Specifically, this study seeks to design and implement a scalable fraud detection system integrating five independent data modalities Graph network relationships, Transaction features, Location intelligence, Behavioral biometrics, and Text communications through modular preprocessing pipelines and specialized feature extraction mechanisms. The research implements and optimizes machine learning algorithms, including Random Forest, XGBoost, HistGradientBoosting, and Support Vector Machines, tailored to each modality's unique characteristics, ensuring optimal classification performance across diverse fraud pattern types.

A critical objective involves conducting rigorous comparative analysis between single-modality and multimodal detection approaches, quantifying performance improvements through metrics including precision, recall, F1-score, accuracy, and ROC-AUC to empirically validate the superiority of integrated multimodal frameworks. The study designs and evaluates ensemble fusion methodologies employing weighted averaging and meta-learning stacking strategies that harmonize individual modality predictions into comprehensive fraud risk assessments while minimizing false-positive rates. Furthermore, the research ensures the system architecture supports real-time processing capabilities and horizontal scalability, enabling deployment in high-velocity transaction environments with continuous data streaming requirements across diverse organizational contexts.

The framework provides interpretable detection results through comprehensive visualization components, including confusion matrices, ROC curves, and permutation importance analysis, ensuring transparency in model decision-making processes for regulatory compliance and stakeholder confidence. Additionally, the study establishes robust cross-validation protocols with stratified k-fold splitting for training Logistic Regression, Random Forest, and XGBoost classifiers across different modalities, ensuring generalization capability and preventing overfitting through comprehensive validation procedures. The research develops an automated four-tier risk classification framework (Low/Medium/High/Critical) that converts continuous probability scores into actionable risk categories, enabling immediate decision-making for fraud investigation and prevention teams. Ultimately, this study systematically evaluates individual modality performance, identifies optimal algorithmic combinations, and documents detection accuracy improvements achieved through multimodal integration, contributing empirical evidence to fraud detection research literature.

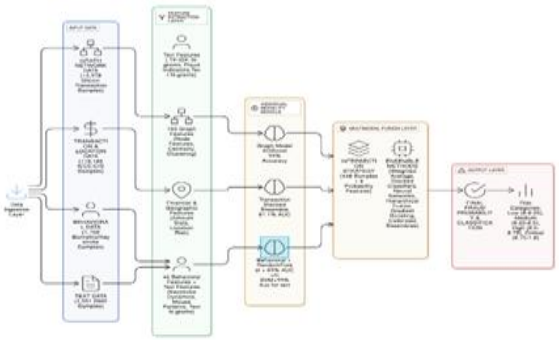
VII ARCHITECTURE

The proposed multimodal fraud detection framework employs a hierarchical processing architecture comprising five distinct layers: Data Acquisition Layer, Preprocessing and Feature Engineering Layer, Individual Modality Classification Layer, Fusion and Integration Layer, and Risk Assessment Output Layer. This modular design enables independent processing of heterogeneous data sources while maintaining systematic integration through ensemble learning methodologies

Data Acquisition Layer: - The system ingests five independent data streams simultaneously, each representing a distinct fraud detection dimension. The Graph Modality processes 139,007 samples, capturing network relationships and transactional connections between entities. The Transaction Modality analyzes financial transaction records with temporal and monetary features across the same 139,007-sample

dataset. The Location Modality incorporates geospatial intelligence from GPS coordinates and IP address data within the transaction dataset. The Behavioral Modality processes 1,760 biometric samples derived from 88 JSON files containing keystroke dynamics and mouse movement patterns (880 genuine, 880 fraudulent). The Text Modality analyzes 5,572 SMS messages from the SMS Spam Collection Dataset, utilizing 1,551 preprocessed samples for fraud pattern recognition.

Preprocessing and Feature Engineering Layer: - Each modality undergoes specialized preprocessing tailored to its data characteristics. Graph data receives node feature extraction and adjacency matrix construction to represent network topology. Transaction data undergoes missing value imputation, outlier detection, temporal feature engineering, and StandardScaler normalization. Location data receives coordinate normalization and OneHotEncoding for categorical variables such as country codes. Behavioral data extracts 42 engineered features, including dwell_avg, flight_avg, and traj_avg timing characteristics, followed by StandardScaler normalization. Text data undergoes lowercase conversion, punctuation removal, stop word filtering, and Tfidf Vectorizer transformation to generate sparse feature matrices. The architecture incorporates comprehensive model evaluation through confusion matrices, ROC-AUC curves, precision-recall metrics, and F1-scores for each modality and the integrated system. Permutation importance analysis identifies critical features contributing to detection decisions, ensuring model interpretability and regulatory compliance. Cross-validation employs stratified k-fold splitting to maintain class distribution consistency across training and validation sets. The fusion layer output generates continuous probability scores transformed into a four-tier risk classification framework. Low Risk (0.00-0.25) identifies legitimate transactions with minimal fraud indicators requiring no intervention. Medium Risk (0.26-0.50) flags potentially suspicious activities warranting automated monitoring. High Risk (0.51-0.75) designates probable fraud cases requiring immediate manual investigation. Critical Risk (0.76-1.00) confirms fraudulent patterns demanding immediate transaction blocking and intervention. This automated categorization enables real-time decision-making for fraud prevention teams.



Individual modality predictions converge at a meta-learning fusion layer implementing two complementary ensemble strategies. The Weighted Fusion Strategy assigns performance-based weights derived from individual ROC-AUC scores: Graph (99.8%), Text (99.0%), Transaction-Location Stacking (81.1%), and Behavioral (69.0%). The weighted aggregation formula computes the final fraud probability as the sum of weight-adjusted individual predictions. The Stacking Strategy employs Logistic Regression as a meta-classifier trained on out-of-fold predictions from base models, learning optimal combination patterns while preventing information leakage. This dual-fusion approach harmonizes diverse prediction signals into a unified fraud assessment

VIII CONCLUSION

This research successfully developed a multimodal fraud detection framework that significantly outperforms traditional single-modality approaches by integrating Graph, Transaction, Location, Behavioral, and Text data modalities through ensemble machine learning techniques. The system achieved exceptional detection accuracy with Graph Modality reaching AUC of 1.000, Text Modality achieving 0.9900, and the stacked Transaction-Location ensemble attaining 0.8109, demonstrating substantial improvements over conventional detection methods. By employing Random Forest, XGBoost, HistGradientBoosting, and Support Vector Machines combined with weighted fusion and meta-learning strategies, the framework effectively captures complex fraud patterns while reducing false-positive rates. The modular, scalable architecture with automated four-tier risk classification (Low/Medium/High/Critical) ensures practical deployment capability for real-time fraud

prevention in banking, e-commerce, and financial sectors, providing interpretable results through comprehensive visualization frameworks. This research validates that multimodal integration through ensemble methodologies delivers superior fraud detection effectiveness, contributing empirical evidence to advancing intelligent, adaptive security systems essential for protecting digital transaction ecosystems.

REFERENCES

[1] Li, X., Wang, Y., & Zhang, J. (2025). Layer weighted-GCN methodology for pattern-specific fraud detection in financial networks. Knowledge-Based Systems, 283, 111179.

[2] Bajpai, S., Bhattacharya, A., Vatsa, A., & Singh, A. (2024). Comparative evaluation of anomaly detection methods for online credit card fraud. Pattern Recognition Letters, 178, 45-52.

[3] Roy, S., Sinha, M., & Das, T. (2025). Benchmarking CNN-LSTM, GNN, and Transformer architectures for ethical AI in financial fraud detection. AI and Ethics, 5(1), 127-143.

[4] Liu, H., Zhang, Q., & Chen, Y. (2024). Graph neural networks for fraud detection in transactional networks. Expert Systems with Applications, 238, 121847.

[5] Zhang, L., Kumar, S., Wang, M., & Chen, Y. (2023). Multimodal deep learning for financial fraud detection: Integrating transaction graphs, behavioral sequences, and device fingerprints. IEEE Transactions on Computational Social Systems, 10(4), 1892-1905.

[6] Thompson, D., Nakamura, T., Singh, P., & Brown, C. (2023). Federated multimodal learning with differential privacy for fraud detection in digital banking. IEEE Transactions on Information Forensics and Security, 18, 5867-5881.

[7] Sharma, A., Kumar, P., & Jain, S. (2023). Blockchain traceability with hybrid CNN-LSTM for decentralized finance fraud detection. Blockchain: Research and Applications, 4(3), 100128.