ONLINE FRAUD PAYMENT DETECTIONUSING BALANCED ML ALGORITHMS

Mr.V. SUDHAKAR[1], MORAMPUDI JOSHNITHA[2],PILLI NAGA VIJAYA LAKSHMI[3], MUDDUNENI KOMALI NAGA SRI SAI KEERTIKA[4], MADDALI ROHITH BABU[5]

[1]Associate Professor, Dept. of CSE, V.K.R, V.N.B & A.G.K COLLEGE OF ENGINEERING

[2345]UG Students, Dept. of CSE, V.K.R, V.N.B & A.G.K COLLEGE OF ENGINEERING, GUDIVADA

ABSTRACT

The rapid growth of online payment systems and digital transactions has significantly increased the risk of financial fraud, making fraud detection an essential component of modern financial security. Online fraud payment detection aims to identify suspicious or unauthorized transactions in real time to prevent financial losses for both users and financial institutions. However, one of the major challenges in fraud detection is the imbalance in transaction datasets, where fraudulent transactions represent only a small portion compared to legitimate ones. Traditional machine learning models often fail to accurately detect fraud due to this class imbalance problem.

This study proposes an online fraud payment detection system using balanced machine learning algorithms to improve detection accuracy and reliability. The proposed approach applies data balancing techniques such as oversampling, under sampling, and synthetic data generation to ensure equal representation of fraudulent and non-fraudulent transactions during model training. Balanced algorithms including Random Forest, Support Vector Machine, Logistic Regression, and Gradient Boosting are utilized to enhance classification performance. Feature engineering and preprocessing techniques are also employed to extract meaningful transaction patterns and reduce noise in the dataset.

The system evaluates performance using metrics such as accuracy, precision, recall, F1-score, and ROC-AUC, which are more suitable for imbalanced datasets. Experimental results demonstrate that balanced machine learning models significantly improve fraud detection rates while minimizing false positives. The proposed framework provides a scalable and efficient solution for secure online payment environments, enabling financial organizations to detect fraudulent activities proactively and enhance customer trust.

Keywords: Online Payment Fraud Detection, Balanced Machine Learning, Imbalanced Data Handling, Financial Security, Classification Algorithms, Data Preprocessing, Fraud Analytics.

I INTRODUCTION

The rapid advancement of digital technologies and widespread adoption of online payment platforms have transformed the global financial ecosystem. Online transactions through mobile banking, e-commerce platforms, digital wallets, and internet banking have become an essential part of daily life due to their convenience, speed, and accessibility. However, the increasing dependency on digital payment systems has also led to a significant rise in online payment fraud. Cybercriminals continuously develop sophisticated techniques such as phishing attacks, identity theft, card-

not-present fraud, and unauthorized transactions, causing substantial financial losses to individuals, businesses, and financial institutions.

Traditional fraud detection systems mainly rely on rule-based mechanisms that use predefined conditions to identify suspicious activities. Although these systems are effective for known fraud patterns, they struggle to detect new and evolving fraud strategies. Moreover, online payment datasets are highly imbalanced, where fraudulent transactions represent only a very small percentage compared to legitimate transactions. This imbalance makes it difficult for conventional machine learning models to learn fraud patterns effectively, often resulting in biased predictions toward normal transactions and poor fraud detection performance.

Machine learning techniques have emerged as powerful tools for fraud detection because they can automatically learn complex transaction behaviors and identify hidden patterns from large volumes of data. However, without proper handling of class imbalance, many algorithms fail to accurately detect rare fraudulent events. Balanced machine learning algorithms address this challenge by applying techniques such as oversampling, under sampling, and synthetic data generation to improve model learning and classification accuracy.

II RELATED WORK

Online payment fraud detection has gained significant attention in recent years due to the rapid growth of digital transactions and increasing cyber-financial crimes. Early fraud detection systems mainly relied on rule-based approaches and statistical analysis; however, these methods were limited in detecting new and evolving fraud patterns. With the advancement of machine learning, researchers began applying supervised and ensemble learning algorithms to automatically identify fraudulent transaction behavior from large datasets.

Several studies have focused on applying traditional machine learning algorithms such as Logistic Regression, Decision Trees, Support Vector Machines (SVM), and Random Forest for fraud detection. Research comparing multiple classifiers showed that ensemble models, particularly Random Forest and XGBoost, provide higher detection accuracy and better generalization when handling complex transaction patterns. These models effectively learn nonlinear relationships between transaction features and fraud behavior, improving prediction performance compared to classical techniques .

A major challenge identified in most research is the severe class imbalance in payment datasets, where fraudulent transactions represent less than 1% of total records. To address this issue, researchers introduced balancing techniques such as SMOTE (Synthetic Minority Oversampling Technique), random oversampling, under sampling, and hybrid resampling methods. Studies demonstrate that combining resampling techniques with machine learning models significantly improves recall and F1-score by enabling models to learn minority fraud patterns more effectively. In particular, SMOTE-based preprocessing has been shown to enhance classification capability and reduce bias toward legitimate transactions.

Recent works also explore ensemble learning and hybrid frameworks that integrate balanced datasets with advanced algorithms. Experimental results indicate that balanced Random Forest and gradient boosting approaches achieve strong performance in terms of precision, recall, and ROC-AUC while minimizing false positives, making them suitable for real-time fraud detection systems . Additionally, feature selection and preprocessing techniques have been applied to improve model efficiency and scalability in financial environments

III LITERATURE REVIEW

Online payment fraud detection has been widely studied due to the increasing use of digital financial services and the growing complexity of cyber fraud attacks. Earlier research primarily focused on statistical and rule-based detection methods that relied on predefined transaction thresholds and manual monitoring. Although these approaches were simple to implement, they lacked adaptability and were ineffective in identifying newly emerging fraud patterns. With the development of data mining and machine learning techniques, researchers began exploring automated fraud detection models capable of learning behavioral patterns from historical transaction data.

Several studies investigated the use of supervised learning algorithms such as Logistic Regression, Decision Trees, Naïve Bayes, Support Vector Machines, and Random Forest for fraud classification. Among these, ensemble learning methods demonstrated better performance due to their ability to handle large datasets and capture complex relationships between transaction attributes. Researchers also emphasized the importance of feature engineering, including transaction amount, frequency, location, and time-based behavioral features, which significantly improve model accuracy.

A major focus of recent literature is addressing the class imbalance problem present in fraud datasets, where fraudulent transactions are extremely rare compared to legitimate ones. To overcome this issue, balancing techniques such as oversampling, undersampling, and Synthetic Minority Oversampling Technique (SMOTE) have been widely adopted. Studies show that combining balanced datasets with machine learning algorithms improves recall and fraud detection rate while reducing misclassification.

IV EXISTING SYSTEM

The existing online payment fraud detection systems mainly rely on traditional rule-based and basic machine learning approaches to identify fraudulent transactions. In rule-based systems, predefined conditions such as transaction amount limits, unusual login locations, repeated transaction attempts, or abnormal spending behavior are used to detect suspicious activities. These systems are easy to implement and provide quick responses; however, they depend heavily on manually created rules and cannot effectively adapt to new or evolving fraud techniques.

Later, conventional machine learning models such as Logistic Regression, Decision Trees, and Support Vector Machines were introduced to improve detection accuracy. These models analyze historical transaction data and learn patterns that distinguish fraudulent transactions from legitimate ones. Although these approaches improved automation and reduced manual monitoring, they still suffer from significant limitations when handling highly imbalanced datasets. Since fraudulent transactions represent only a small percentage of total transactions, the models tend to favor the majority class, resulting in high accuracy but poor fraud detection rates.

Another limitation of existing systems is the high number of false positives, where genuine transactions are incorrectly classified as fraud. This creates inconvenience for customers and increases operational workload for financial institutions. Additionally, many traditional systems lack real-time detection capability and struggle to process large volumes of transaction data efficiently.

DISADVANTAGES

The existing online payment fraud detection systems suffer from several limitations that reduce their effectiveness in modern digital transaction environments. One of the major disadvantages is the heavy reliance on rule-based detection mechanisms. These systems depend on predefined rules created by

experts, which makes them unable to adapt quickly to new and evolving fraud techniques. As cybercriminals continuously change their strategies, static rules often fail to detect unknown or sophisticated fraud patterns.

Another significant drawback is the class imbalance problem present in transaction datasets. Fraudulent transactions represent only a very small portion compared to legitimate ones, causing traditional machine learning models to become biased toward normal transactions. As a result, models may achieve high overall accuracy but fail to correctly identify fraudulent activities, leading to low recall and poor detection performance.

Existing systems also generate a high number of false positives, where genuine transactions are incorrectly flagged as fraudulent. This negatively impacts customer experience by blocking valid payments and creates additional verification workload for banks and financial institutions. Furthermore, many traditional approaches require frequent manual updates and monitoring, increasing operational complexity and maintenance costs.

Scalability and real-time processing are additional challenges. Conventional systems often struggle to handle large volumes of transaction data generated by modern online platforms, resulting in delayed detection and response. Moreover, limited feature analysis and lack of advanced learning capabilities reduce the system's ability to identify hidden behavioral patterns.

V PROPOSED SYSTEM

The proposed system introduces an online fraud payment detection framework using balanced machine learning algorithms to improve the accuracy and reliability of fraud identification. Unlike traditional systems, the proposed approach focuses on handling the class imbalance problem present in online transaction datasets, where fraudulent transactions are significantly fewer than legitimate ones. The system applies data balancing techniques such as oversampling, undersampling, and Synthetic Minority Oversampling Technique (SMOTE) to create a balanced training dataset, enabling machine learning models to effectively learn fraud patterns.

The proposed model includes multiple stages such as data collection, preprocessing, feature extraction, data balancing, model training, and evaluation. During preprocessing, missing values, noise, and duplicate records are removed to enhance data quality. Important transaction features such as transaction amount, time, frequency, location, and user behavior patterns are extracted to improve prediction capability. Balanced machine learning algorithms including Random Forest, Logistic Regression, Support Vector Machine, and Gradient Boosting are then trained using the balanced dataset to classify transactions as fraudulent or legitimate.

The system is designed to support real-time fraud detection by analyzing incoming transactions instantly and generating alerts when suspicious activities are identified. Advanced evaluation metrics such as precision, recall, F1-score, and ROC-AUC are used to measure performance more effectively than simple accuracy measures. By reducing false positives and improving fraud detection rates, the proposed system enhances customer trust and financial security.

ADVANTAGES

The proposed online fraud payment detection system using balanced machine learning algorithms offers several significant advantages over traditional fraud detection approaches. One of the primary benefits is improved fraud detection accuracy. By applying data balancing techniques such as oversampling and undersampling, the system effectively handles the class

imbalance problem, enabling the model to learn fraudulent transaction patterns more efficiently and increasing the detection rate.

Another important advantage is the reduction of false positives. The balanced learning process helps the system distinguish clearly between legitimate and fraudulent transactions, minimizing incorrect fraud alerts and improving customer experience. This reduces unnecessary transaction blocking and lowers the workload for financial institutions involved in manual verification processes.

The proposed system also provides adaptability and scalability. Machine learning algorithms continuously learn from transaction data and can adapt to new and evolving fraud strategies without requiring frequent manual rule updates. Additionally, the system can process large volumes of online transactions, making it suitable for real-time payment platforms and modern digital banking environments.

Improved performance evaluation is another benefit, as the system uses advanced metrics such as precision, recall, F1-score, and ROC-AUC instead of relying only on accuracy. This ensures more reliable assessment when dealing with imbalanced datasets. Furthermore, automated detection reduces human intervention, improves operational efficiency, and enables faster response to suspicious activities.

VI METHODOLOGY

The methodology for the proposed online fraud payment detection system using balanced machine learning algorithms consists of several structured stages designed to ensure accurate and efficient fraud identification. The process begins with data collection, where historical online transaction datasets are gathered from financial records or publicly available sources. These datasets typically include features such as transaction amount, time, location, transaction type, and user behavior patterns.

In the preprocessing stage, the collected data is cleaned by removing missing values, duplicate entries, and irrelevant attributes to improve data quality. Data normalization and transformation techniques are applied to ensure consistency and improve model performance. After preprocessing, feature engineering is performed to extract meaningful attributes that help distinguish fraudulent transactions from legitimate ones.

Since fraud datasets are highly imbalanced, data balancing techniques play a crucial role in the methodology. Methods such as oversampling, undersampling, and Synthetic Minority Oversampling Technique (SMOTE) are applied to create a balanced dataset, allowing machine learning algorithms to learn minority fraud patterns effectively. Following this, balanced machine learning models including Random Forest, Logistic Regression, Support Vector Machine, and Gradient Boosting are trained using the prepared dataset.

The trained models are then evaluated using performance metrics such as precision, recall, F1-score, accuracy, and ROC-AUC to measure detection effectiveness. Finally, the best-performing model is deployed for real-time transaction monitoring, where incoming payments are analyzed instantly and flagged if suspicious behavior is detected. This systematic methodology ensures improved fraud detection accuracy, reduced false alarms, and enhanced security in online payment systems.

VII SYSTEM MODEL

SYSTEM ARCHITECTURE

VIII RESULTS AND DISCUSSIONS

IX CONCLUSION

The online fraud payment detection system using balanced machine learning algorithms provides an effective and intelligent solution for identifying fraudulent transactions in modern digital payment environments. The study addresses one of the major challenges in fraud detection, namely the class imbalance problem, by applying data balancing techniques such as oversampling, undersampling, and SMOTE. These techniques enable machine learning models to learn fraud patterns more accurately and improve overall detection performance.

By integrating preprocessing, feature engineering, balanced learning models, and performance evaluation metrics, the proposed system achieves better precision,

recall, and F1-score compared to traditional rule-based and unbalanced machine learning approaches. The system also reduces false positives, thereby improving customer experience and minimizing unnecessary transaction interruptions. Additionally, the ability to analyze transactions in real time enhances financial security and allows organizations to respond quickly to suspicious activities.

The proposed framework is scalable, adaptive, and capable of handling large volumes of online transaction data, making it suitable for banking, e-commerce, and digital payment platforms. Overall, the implementation of balanced machine learning algorithms significantly strengthens fraud prevention mechanisms and contributes to building a safer and more reliable online payment ecosystem.

REFERENCES

[1]. V. Chandola, A. Banerjee, and V. Kumar, "Anomaly Detection: A Survey," *ACM Computing Surveys*, vol. 41, no. 3, pp. 1–58, 2009.

[2]. A. Dal Pozzolo, O. Caelen, Y. Le Borgne, S. Waterschoot, and G. Bontempi, "Learned Lessons in Credit Card Fraud Detection from a Practitioner Perspective," *Expert Systems with Applications*, vol. 41, no. 10, pp. 4915–4928, 2014.

[3]. N. Japkowicz and S. Stephen, "The Class Imbalance Problem: A Systematic Study," *Intelligent Data Analysis*, vol. 6, no. 5, pp. 429–449, 2002.

[4]. H. He and E. A. Garcia, "Learning from Imbalanced Data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, 2009.

[5]. W. Chawla, K. Bowyer, L. Hall, and W. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[6]. L. Breiman, "Random Forests," *Machine Learning Journal*, vol. 45, no. 1, pp. 5–32, 2001.

[7]. C. Cortes and V. Vapnik, "Support-Vector Networks," *Machine Learning*, vol. 20, pp. 273–297, 1995.

[8]. T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016.

[9]. S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. Westland, "Data Mining for Credit Card Fraud: A Comparative Study," *Decision Support Systems*, vol. 50, no. 3, pp. 602–613, 2011.

[10]. Sharma, P., & Gupta, R. (2023). AI and Geo-Fencing Based Smart Tourist Safety Framework. International Journal of Advanced Computer Science.

[11]. Sharma, S., & Kaur, R. (2019). Automated recruitment using natural language processing: Techniques and challenges. International Journal of Advanced Computer Science and Applications, 10(6), 1–8.

[12]. Dayal, P. S., Chandra, B. R., Keerthi, M., Sruthi, M., Venkatesh, K., Appalaraju, G., & Eswari, G. (2013). Design of Pyramidal Horn Antenna at 10GHz Using WIPL-D Optimizer. International Journal of Electronics Communication and Computer Engineering, 4(2).

[13]. Viswanathan, V., Polagani, S. S., Agarwal, R., Akula, S., Dey, S., & Kashyap, R. (2025, September). AI-Augmented Threat Intelligence for Proactive Intrusion Detection in Multi-Cloud Ecosystem. In 2025 IEEE International Conference on Advanced

Computing Technologies (ICACT) (pp. 567-572). IEEE.

[14]. Sruthi, M. V., Sree, V. U., & Soundararajan, K. (2012). Specific removal of motion artifacts in medical image processing. IJECCE, 3(3), 227-229.

[15]. Viswanathan, V., Shah, A. K., Kubam, C. S., Dontu, S., Gandhi, A., & Singla, P. (2025, August). Deep Learning-Driven Stock Market Forecasting Using Cloud-Based Financial Time Series Analytics. In 2025 International Conference on Emerging Trends in Networks and Computer Communications (ETNCC) (pp. 1-6). IEEE.

[16]. Viswanathan, V. (2025). Agentic AI for Employment: Reducing Unemployment through Intelligent Job-Seeker Support. LEX LOCALIS–Journal of Local Self-Government.

[17]. Viswanathan, V. (2024). Pioneering Ethical AI Integration in Enterprise Workflows: A Framework for Scalable Team Governance. Available at SSRN 5375619.

[18]. Sruthi, M. V., Soundararajan, K., & Sree, V. U. (2012). Accurate Multimodality Registration of medical images. International Journal of Engineering Research and Development, 1(3), 33-36.

[19]. Ranjbareslamloo, S., Dzukeya, G. A., Muhit, M. M. I., & Qattawi, A. (2025). Numerical and experimental study of residual stress in additively manufactured IN718. Manufacturing Letters, 44, 915–927. https://doi.org/10.1016/j.mfglet.2025.915927

[20]. Mahtabi, M., Roshan, M., Muhit, M. M. I., Behvar, A., & Haghshenas, M. (2026). Cryogenic ultrasonic fatigue: Mechanisms, advancements, and insights. Cryogenics, 153, 104257. https://doi.org/10.1016/j.cryogenics.2025.104257

[21]. Kotte, G. (2025). Enhancing Cloud Infrastructure Security on AWS with HIPAA Compliance Standards. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.5283660

[22]. GIRISH KOTTE. (2025). ETHICAL ISSUES SURROUNDING THE INTEGRATION OF AI-POWERED DIAGNOSTIC TOOLS IN THE HEALTHCARE SECTOR. American Journal of AI Cyber Computing Management, 5(4), 329–334. https://doi.org/10.64751/ajaccm.2025.v5.n4.pp329-334

[23]. Kumara, S. (2025). Identity-Driven IoT Security in Telecom Ecosystems: Implications for Scalable and Trustworthy Digital Infrastructure. Int. J. Appl. Math, 38(12s), 2797-2816.

[24]. Poojari, R. INTELLIGENT SYSTEMS+B108 AND APPLICATIONS IN ENGINEERING.

[25]. Cyril, H. P., & Kumara, S. (2026, February). DevSecOps-Driven Security Integration in the Software Development Lifecycle Using CI/CD Pipelines. In 2026 IEEE 5th International Conference on AI in Cybersecurity (ICAIC) (pp. 1-6). IEEE.

[26]. Prodduturi, S. M. K. To Secure Your Paper as Per UGC Guidelines We Are Providing A ElectronicBar code.

[27]. Santthosh Saai Reddy Purmani. (2026). Artificial Intelligence First Enterprise Architecture: The Design of Scalable, Secure, and Intelligent IT Ecosystems. American Journal of AI Cyber Computing Management, 6(1(2)), 1–8. https://doi.org/10.64751/ajaccm.2026.v6.n1(2).pp1-8

[28]. Purmani, S. S. R. (2025). Optimizing IT project management through advanced ROI

analysis techniques. International Journal for Innovative Engineering and Management Research, 14(3), 301–312.

[29]. Patyrykin, K. (2025). CANCEL CULTURE PROBLEM. Lex Localis: Journal of Local Self-Government, 23.

[30]. Kalae, U. K. (2021). Creating tailored Power Apps to optimize data collection and reporting across multiple platforms. International Journal for Innovative Engineering and Management Research, 10(10), 49–56.

[31]. Patel, S., & Patyrykin, K. (2025). Strategic Impacts of Salesforce Automation on Organisational Competitive Advantage in Emerging Markets. Journal of Posthumanism, 5(12), 357–372. https://doi.org/10.63332/joph.v5i12.3782

[32]. Vasagam, M., Kumar, A., & Garg, A. (2026). Learning Execution Plan Embeddings for Multi-Dimensional Query Resource Prediction. IEEE Access.

[33]. Kalae, U. K. (2023). Enhancing deployment efficiency through CI/CD pipelines and containerization with Docker and Kubernetes. International Journal of Communication Networks and Information Security, 15(4), 728–736.

[34]. Poojari, R. Enhancing Healthcare Decision-Making through Machine Learning and the Analysis of Large-Scale Medical Data.

[35]. Akhilaiswarya, B., Sree, B. T., Lilly, K., Chowdary, K. H., & Sruthi, M. (2023). Elderly fall detection and location tracking system using heterogeneous networks. Journal of Engineering Sciences, 14(05).

[36]. Reddy, S. K. R. Developing a Modular AI Framework to Enhance Scalability and Personalization in Next-Generation Reward Platforms.