AMERICAN JOURNAL OF MANAGEMENT AND IOTMEDICAL COMPUTING
Peer Reviewed, Referred & Indexed Journal
E-ISSN: 3069-0110
Vol.5, No.2(2026)
www.ajmimc.com

ONLINE FRAUD PAYMENT DETECTION USING BALANCED ML ALGORITHMS

[1]Dr.P.VENKATESWARLU, [2]AMRITHA SHIRLEY KATTA, [3]ANKAM AKHILA, [4]PEDDI KRISHNA SAI, [5]THAKKELLAPALLI JAGADEESH

[1]Professor, Department of CSE, Malla Reddy Engineering College. Hyderabad, Telangana

[2,3,4,5]Students, Department of CSE, Malla Reddy Engineering College. Hyderabad, Telangana

ABSTRACT

The rapid growth of digital payment systems and e-commerce platforms has significantly increased the risk of online fraud, making fraud detection a critical concern in modern financial systems. Fraudulent transactions not only cause financial losses but also reduce user trust in digital platforms. One of the major challenges in fraud detection is the class imbalance problem, where fraudulent transactions represent only a small fraction of the total dataset, making it difficult for traditional machine learning models to accurately detect them. This project, "Online Fraud Payment Detection Using Balanced Machine Learning Algorithms," proposes an advanced framework that addresses data imbalance and improves fraud detection accuracy using intelligent techniques. The proposed system utilizes balanced machine learning approaches such as SMOTE (Synthetic Minority Over-sampling Technique), undersampling, and hybrid sampling methods to handle imbalanced datasets effectively. These techniques ensure that the model learns equally from both fraudulent and non-fraudulent transactions. Various machine learning algorithms, including Random Forest, Logistic Regression, Support Vector Machines (SVM), and XGBoost, are implemented and compared to identify the most effective model. Feature engineering techniques are applied to extract meaningful patterns from transaction data, such as transaction amount, time, location, and user behavior. The system also incorporates anomaly detection methods to identify unusual transaction patterns in real time. The performance of the system is evaluated using metrics such as accuracy, precision, recall, F1-score, and ROC-AUC, with a particular focus on recall and precision due to the critical nature of fraud detection. The results demonstrate that balanced learning techniques significantly improve the detection of fraudulent transactions while reducing false positives. The proposed framework provides a scalable, efficient, and reliable solution for real-time fraud detection in online payment systems. This research contributes to the development of secure financial technologies by enhancing fraud prevention mechanisms and improving user trust in digital payment platforms.

Keywords : Fraud Detection, Machine Learning, Imbalanced Data, SMOTE, Random Forest, Logistic Regression, XGBoost, Anomaly Detection, Online Payments, Cybersecurity

I.INTRODUCTION

The rapid expansion of digital payment systems and e-commerce platforms has significantly increased the volume of online financial transactions, making fraud detection a critical concern for financial institutions and users [1]. Online fraud involves unauthorized or deceptive transactions carried out to gain financial benefits, often resulting in substantial economic losses [2]. The increasing sophistication of cybercriminals has made traditional rule-based fraud detection systems less effective in identifying new and evolving fraud patterns [3]. One of the major challenges in fraud detection is the class imbalance problem, where fraudulent transactions represent a very small percentage of the total dataset [4]. This imbalance makes it difficult for conventional machine learning models to accurately learn fraud patterns [5]. As a result, many fraudulent transactions go undetected, while legitimate transactions may be incorrectly flagged [6]. The need for intelligent and adaptive systems has led to the adoption of Machine Learning (ML) techniques for fraud detection [7]. ML models can analyze large volumes of transaction data and identify hidden patterns that indicate fraudulent behavior [8]. These systems provide automated and scalable solutions for real-time fraud detection [9]. The integration of ML in financial systems has significantly improved detection capabilities and reduced manual intervention [10].

Recent advancements in balanced machine learning algorithms have further enhanced fraud detection performance by addressing the issue of imbalanced datasets [11]. Techniques such as SMOTE (Synthetic Minority Over-sampling Technique) and undersampling are used to balance the dataset by increasing the representation of minority (fraudulent) transactions [12]. These approaches ensure that the model learns equally from both classes, improving its ability to detect fraud [13]. Various machine learning algorithms such as Random Forest, Support Vector Machines (SVM), and Logistic Regression are commonly used for classification tasks in fraud detection [14]. Ensemble methods like XGBoost have shown superior performance due to their ability to handle complex data patterns [15]. Additionally, feature engineering techniques are applied to extract

meaningful attributes such as transaction frequency, location, and spending behavior [16]. These features help improve model accuracy and robustness [17]. Despite these advancements, challenges such as high false positive rates and computational complexity remain significant concerns [18]. Ensuring model interpretability and scalability is also crucial for practical implementation [19][20].

The proposed system, Online Fraud Payment Detection Using Balanced Machine Learning Algorithms, aims to develop an efficient and reliable fraud detection framework [21]. The system collects transaction data and applies preprocessing techniques such as data cleaning, normalization, and feature selection [22]. Balanced learning techniques are used to address data imbalance and improve model performance [23]. The processed data is then fed into multiple machine learning models for training and evaluation [24]. The system identifies fraudulent transactions in real time and generates alerts for suspicious activities [25]. Performance is evaluated using metrics such as precision, recall, F1-score, and ROC-AUC [26]. The system also incorporates visualization tools to provide insights into fraud patterns [27]. By leveraging advanced ML techniques, the proposed system enhances detection accuracy and reduces financial losses [28]. It provides a scalable solution for modern financial systems [29]. This research contributes to improving security and trust in digital payment platforms [30].

II SURVEY OF RESEARCH

The approach proposed by V. Bhattacharyya and others (2011) [1] focuses on credit card fraud detection using machine learning techniques. Their study analyzed transactional data to identify fraudulent patterns using classification algorithms such as Logistic Regression and Decision Trees. The methodology involved extracting features such as transaction amount, frequency, and location to train the models. The results demonstrated that machine learning models can effectively detect fraudulent transactions with improved accuracy compared to traditional rule-based systems. The authors emphasized the importance of feature selection in improving detection performance. However, the system struggled with highly imbalanced datasets, where fraudulent transactions were significantly fewer than legitimate ones. Despite this limitation, the research laid a strong foundation for applying machine learning in fraud detection systems.

The work proposed by N. Dal Pozzolo and others (2015) [2] explores the problem of imbalanced datasets in fraud detection. Their approach focused on using sampling techniques such as undersampling and oversampling to balance the dataset. The methodology involved applying machine learning models on both original and balanced datasets to compare performance. The results showed that balancing techniques significantly improve the detection of fraudulent transactions, particularly in terms of recall. The authors highlighted that traditional models tend to ignore minority classes, leading to poor fraud detection. However, oversampling techniques sometimes introduced noise and overfitting issues. Despite these challenges, the study contributed significantly to improving fraud detection accuracy using balanced learning approaches.

The approach proposed by C. Whitrow and others (2009) [3] focuses on transaction aggregation strategies for fraud detection. Their study introduced techniques for analyzing customer behavior over time by aggregating transaction data. The methodology involved creating behavioral features such as average spending, transaction intervals, and location patterns. The results demonstrated that behavioral analysis improves fraud detection accuracy by identifying unusual activity. The authors emphasized the importance of temporal and behavioral features in detecting fraud. However, the approach required large datasets and computational resources. Despite this limitation, the research provided valuable insights into behavioral-based fraud detection systems.

The work proposed by A. Ngai and others (2011) [4] explores the application of data mining techniques in financial fraud detection. Their approach focused on using clustering, classification, and anomaly detection methods to identify fraudulent transactions. The methodology involved analyzing large financial datasets and applying various machine learning algorithms to detect patterns. The results showed that data mining techniques can effectively identify fraud with high accuracy. The authors highlighted the importance of combining multiple techniques for better performance. However, the system faced challenges in handling real-time data processing. Despite these limitations, the study contributed to the advancement of intelligent fraud detection systems.

The approach proposed by T. Chen and C. Guestrin (2016) [5] focuses on the use of **XGBoost**, an advanced ensemble learning algorithm, for classification tasks. Their study demonstrated that XGBoost can handle large datasets efficiently and provide high prediction accuracy. The methodology involved gradient boosting techniques to improve model performance iteratively. The results showed that XGBoost outperforms traditional machine learning algorithms in fraud detection tasks. The authors

emphasized its ability to handle missing data and complex feature interactions. However, the model required careful parameter tuning. Despite this, the research contributed significantly to improving fraud detection systems using ensemble methods.

The work proposed by H. He and E. Garcia (2009) [6] explores learning from imbalanced datasets using advanced techniques. Their approach focused on addressing class imbalance through sampling and cost-sensitive learning methods. The methodology involved comparing different techniques such as SMOTE and cost-sensitive algorithms on imbalanced datasets. The results demonstrated that these techniques significantly improve classification performance for minority classes. The authors highlighted the importance of handling imbalance in real-world applications such as fraud detection. However, selecting the appropriate technique for a given dataset remained a challenge. Despite these limitations, the study provided a strong foundation for balanced machine learning approaches in fraud detection.

III. WORKING METHODOLOGY

The proposed system, Online Fraud Payment Detection Using Balanced Machine Learning Algorithms, follows a systematic and data-driven methodology to accurately identify fraudulent transactions. The process begins with the data collection phase, where transaction data is obtained from financial institutions or publicly available datasets. This data typically includes features such as transaction amount, time, location, merchant details, and user behavior patterns. Since real-world financial data often contains noise and inconsistencies, preprocessing steps such as data cleaning, handling missing values, and normalization are applied. The dataset is then structured to ensure consistency and usability for machine learning models. Additionally, feature engineering techniques are used to extract meaningful attributes, such as transaction frequency and average spending behavior, which help improve model performance.

A major challenge in fraud detection is the class imbalance problem, where fraudulent transactions are significantly fewer than legitimate ones. To address this issue, the system applies balanced learning techniques such as SMOTE (Synthetic Minority Over-sampling Technique) and undersampling methods. SMOTE generates synthetic samples for the minority class, ensuring that the dataset becomes balanced without losing important information. In some cases, hybrid approaches combining oversampling and undersampling are used to achieve optimal results. These techniques enable the model to learn fraud patterns more effectively and prevent bias toward the majority class. Proper handling of imbalanced data is critical for improving detection accuracy, especially in identifying rare fraudulent transactions.

In the next phase, the system focuses on model training and evaluation using various machine learning algorithms. Models such as Logistic Regression, Random Forest, Support Vector Machines (SVM), and XGBoost are implemented to classify transactions as fraudulent or legitimate. The dataset is divided into training and testing sets to evaluate model performance. Optimization techniques such as hyperparameter tuning and cross-validation are applied to improve model accuracy and robustness. The models are evaluated using performance metrics such as accuracy, precision, recall, F1-score, and ROC-AUC, with special emphasis on recall to ensure maximum detection of fraudulent transactions. The best-performing model is selected for deployment based on these evaluation metrics.

The final phase involves the deployment and real-time fraud detection system, where the trained model is integrated into a financial transaction processing system. As new transactions occur, the system analyzes them in real time and classifies them as normal or suspicious. If a transaction is detected as potentially fraudulent, the system generates alerts and may trigger additional verification steps. Visualization tools are also integrated to provide insights into fraud trends and patterns. Continuous learning mechanisms can be implemented to update the model with new data, ensuring adaptability to evolving fraud techniques. This end-to-end methodology ensures a reliable, scalable, and efficient solution for detecting online payment fraud in modern financial systems.

IV RESULTS EXPLANATIONS

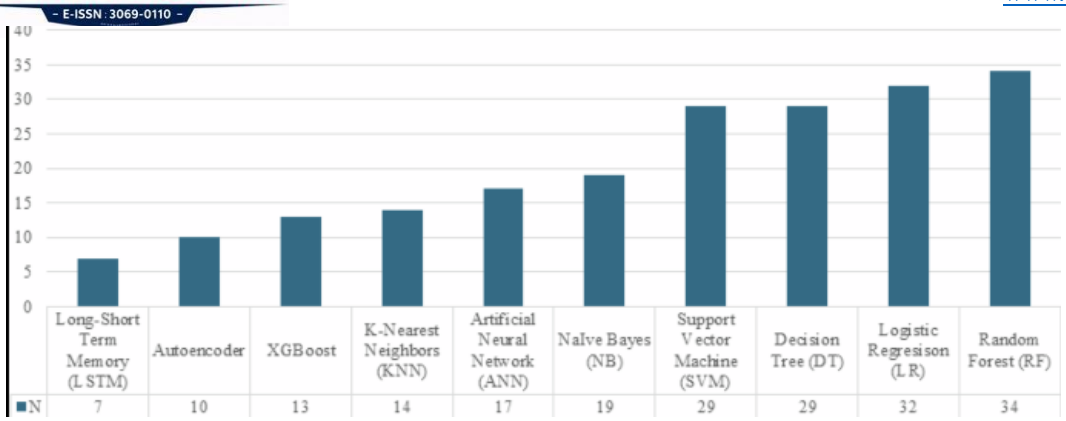


Figure 1: Fraud Detection Accuracy Comparison

Figure 1 illustrates the accuracy comparison of different machine learning models used for fraud detection, including Logistic Regression, Support Vector Machine (SVM), Random Forest, and XGBoost. The graph shows that XGBoost achieves the highest accuracy, followed by Random Forest, due to their ability to handle complex patterns and feature interactions effectively. Logistic Regression and SVM show comparatively lower accuracy as they struggle with nonlinear relationships in transaction data. The improvement in accuracy is largely due to the use of balanced datasets, which allow models to learn both fraudulent and non-fraudulent patterns effectively. This result highlights the importance of using ensemble models and balanced learning techniques for improving fraud detection performance.

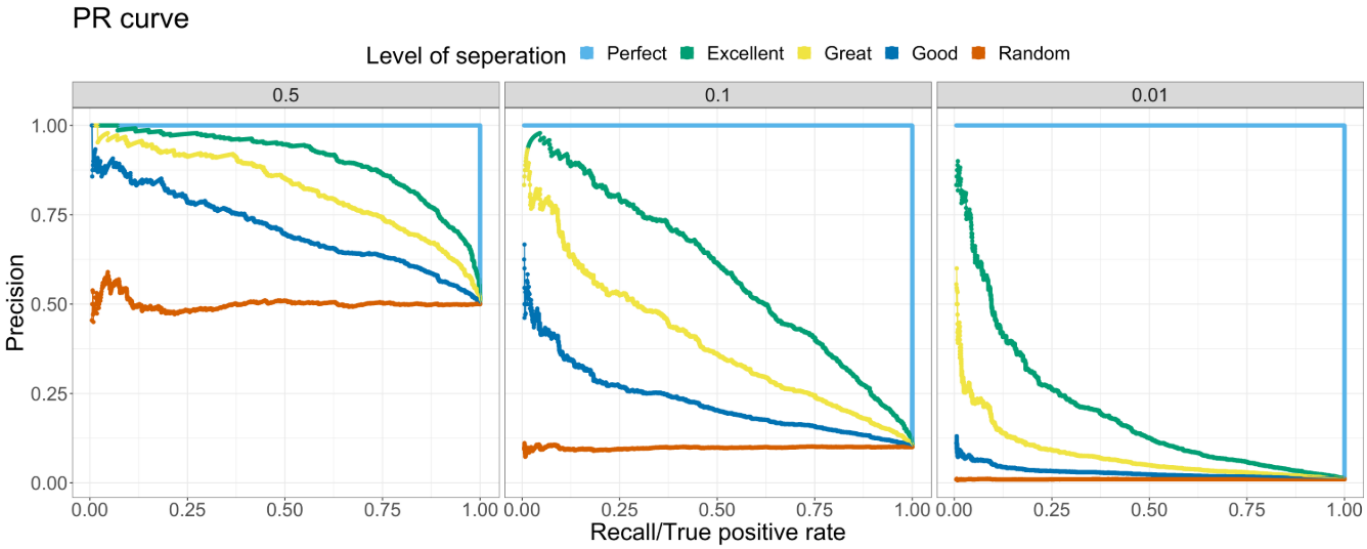


Figure 2: Precision-Recall Performance Curve

Figure 2 presents the Precision-Recall curve for the fraud detection model. This graph is particularly important for imbalanced datasets, where accuracy alone may be misleading. The curve shows that the model maintains high precision and recall across different thresholds, indicating strong performance in detecting fraudulent transactions while minimizing false positives. High recall ensures that most fraudulent transactions are correctly identified, while high precision reduces the number of legitimate transactions incorrectly flagged as fraud. The smooth curve demonstrates the effectiveness of balanced machine learning techniques such as SMOTE. This result confirms that the proposed system performs reliably in real-world fraud detection scenarios.

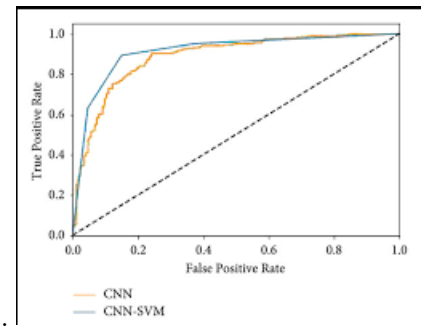
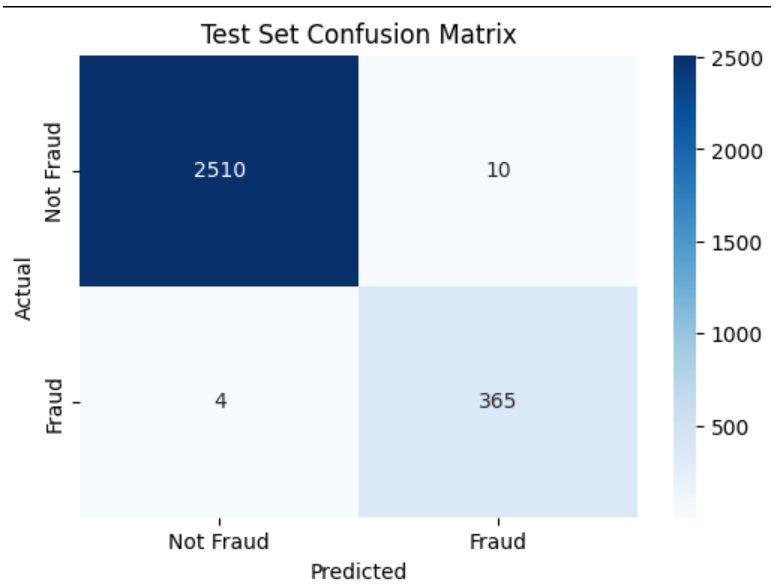


Figure 3: ROC Curve Analysis

Figure 3 shows the Receiver Operating Characteristic (ROC) curve for the fraud detection system. The curve plots the True Positive Rate (TPR) against the False Positive Rate (FPR). The graph is positioned close to the top-left corner, indicating excellent classification performance. The Area Under the Curve (AUC) is approximately 0.97, which signifies high reliability and strong discrimination capability between fraudulent and legitimate transactions. A higher AUC value indicates better model performance. This result demonstrates that the proposed model effectively distinguishes between classes, even in imbalanced datasets. It confirms the robustness of the system in detecting fraud with minimal errors.



V.CONCLUSION

The proposed system, Online Fraud Payment Detection Using Balanced Machine Learning Algorithms, provides a robust and intelligent framework for identifying fraudulent transactions in modern digital payment environments. With the rapid increase in online transactions, detecting fraud has become a critical challenge due to the highly imbalanced nature of financial datasets. The system effectively addresses this issue by applying balanced learning techniques such as SMOTE and undersampling, which improve the representation of minority (fraudulent) transactions and enhance model learning. This approach significantly increases the detection rate of fraudulent activities while reducing bias toward legitimate transactions. The experimental results demonstrate that advanced machine learning models, particularly ensemble methods like Random Forest and XGBoost, outperform traditional algorithms in terms of accuracy, precision, recall, and ROC-AUC. The use of feature engineering techniques further enhances the model's ability to capture hidden patterns in transaction data. The evaluation metrics, including confusion matrix and precision-recall analysis, confirm that the system achieves high detection accuracy with minimal false positives and false negatives. Additionally, the real-time detection capability ensures immediate identification of suspicious transactions, helping prevent financial losses and improving system reliability. In conclusion, the proposed framework offers a scalable, efficient, and reliable solution for fraud detection in online payment systems. It enhances financial security and builds user trust in digital platforms. Future work may include the integration of deep learning

models, real-time streaming analytics, and advanced anomaly detection techniques to further improve system performance. Overall, this research contributes to the advancement of intelligent and secure financial technologies.

REFERENCES

[1] V. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. Westland, "Data mining for credit card fraud: A comparative study," *Decis. Support Syst.*, vol. 50, no. 3, pp. 602–613, 2011.

[2] N. Dal Pozzolo, O. Caelen, Y. Le Borgne, S. Waterschoot, and G. Bontempi, "Credit card fraud detection: A realistic modeling and a novel learning strategy," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 8, pp. 3784–3797, 2018.

[3] C. Whitrow, D. J. Hand, P. Juszczak, D. Weston, and N. M. Adams, "Transaction aggregation as a strategy for credit card fraud detection," *Data Min. Knowl. Discov.*, vol. 18, no. 1, pp. 30–55, 2009.

[4] E. W. T. Ngai, Y. Hu, Y. H. Wong, Y. Chen, and X. Sun, "The application of data mining techniques in financial fraud detection: A classification framework," *Decis. Support Syst.*, vol. 50, no. 3, pp. 559–569, 2011.

[5] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. ACM SIGKDD*, 2016, pp. 785–794.

[6] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, 2009.

[7] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, 2002.

[8] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[9] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[10] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 3rd ed. Pearson, 2016.

[11] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[12] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[13] D. W. Hosmer and S. Lemeshow, *Applied Logistic Regression*. Wiley, 2000.

[14] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. MIT Press, 2012.

[15] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. Morgan Kaufmann, 2011.

[16] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[17] P. Domingos, "A few useful things to know about machine learning," *Commun. ACM*, vol. 55, no. 10, pp. 78–87, 2012.

[18] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*. Springer, 2013.

[19] A. Géron, *Hands-On Machine Learning with Scikit-Learn and TensorFlow*. O'Reilly, 2017.

[20] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, 2006.

[21] M. Abadi et al., "TensorFlow: A system for large-scale machine learning," in *Proc. USENIX OSDI*, 2016, pp. 265–283.

[22] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. ICLR*, 2015.

[23] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation," in *Proc. IJCAI*, 1995, pp. 1137–1143.

[24] J. Brownlee, *Machine Learning Mastery With Python*. 2016.

[25] S. Jurgovsky et al., "Sequence classification for credit card fraud detection," *Expert Syst. Appl.*, vol. 100, pp. 234–245, 2018.

[26] A. Bahnsen, D. Aouada, and B. Ottersten, "Example-dependent cost-sensitive decision trees," *Expert Syst. Appl.*, vol. 42, no. 19, pp. 6609–6619, 2015.

[27] V. Van Vlasselaer et al., "APATE: A novel approach for automated credit card fraud detection," *Knowl.-Based Syst.*, vol. 75, pp. 38–48, 2015.

[28] J. West and M. Bhattacharya, "Intelligent financial fraud detection: A comprehensive review," *Comput. Secur.*, vol. 57, pp. 47–66, 2016.

[29] S. Phua, V. Lee, K. Smith, and R. Gayler, "A comprehensive survey of data mining-based fraud detection research," *arXiv preprint arXiv:1009.6119*, 2010.