Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 1s (2025)
https://www.jneonatalsurg.com

OPEN ACCESS

A Hybrid ML and Data Science Approach to Detect Online Fraud Transaction at Real Time

Swagatika Lenka[1], Dr. Ravindra Tiwari[2]

[1]Research Scholar, Department of CS, LNCT University, Kolar Road, Bhopal, Madhya Pradesh, India
Email ID: ai.swagatika96@gmail.com

[2]Professor, Department of Computer Science, LNCT University, Kolar Road, Bhopal, Madhya Pradesh
Email Id: ravindratiwari.s@gmail.com

Cite this paper as: Swagatika Lenka, Dr. Ravindra Tiwari, (2025) A Hybrid ML and Data Science Approach to Detect Online Fraud Transaction at Real Time. *Journal of Neonatal Surgery*, 14 (1s), 765-775.

ABSTRACT

To safeguard customers and financial institutions, the swift growth of online transactions calls for strong fraud detection systems. To identify online fraud in real time, this study suggests a mixed machine learning (ML) and data science strategy. Through the integration of many data mining methodologies, such as supervised and unsupervised learning algorithms, the research endeavours to detect trends and anomalies suggestive of fraudulent activity. A comprehensive knowledge of transaction behaviours is made possible by the methodology's emphasis on dynamic feature extraction and selection, which makes use of massive datasets made up of transactional records.

Using ensemble learning techniques reduces false positives and improves prediction accuracy. Results from experiments reveal that the hybrid model works well, outperforming conventional techniques in terms of processing speed and detection rates. Furthermore, the model's flexibility facilitates its implementation across multiple internet platforms, guaranteeing efficiency and scalability. The results highlight the value of a multidisciplinary strategy in the fight against online fraud, which will ultimately help create more secure online transaction environments. The foundation for future research targeted at improving fraud detection techniques in an increasingly digital economy is laid by this study.

Keywords: *Accuracy, Benford's Law, Classification Accuracy, False Positive Rate, Hybrid Model, ML-Data Science Approach, Precision, Recall, Real-Time Detection, Transaction Volume.*

1. INTRODUCTION

1.1 An Overview of E-Commerce Online Fraud Detection:

For e-commerce to detect and stop fraudulent activity in real time, online fraud detection is essential. Both customers and organizations may suffer from fraudulent transactions, such as identity theft, account takeover, and payment fraud. Finding fraud quickly without sacrificing user experience is the difficult part. Systems must be flexible to use machine learning (ML) and data science to analyse large volumes of transactional data and forecast possible fraud threats as online fraud strategies change. This calls for a thorough comprehension of fraud trends in addition to sophisticated algorithms that can tell the difference between fraud and genuine activity.

1.2 Data Preprocessing Techniques for Fraud Detection:

In order to prepare raw transactional data for fraud detection algorithms, data preparation is necessary. Since there are usually many fewer fraudulent transactions than genuine ones, handling imbalanced datasets is a major difficulty. To rectify this imbalance, methods such as under sampling, oversampling (SMOTE), or the use of synthetic data might be employed. To increase model accuracy, data cleaning techniques eliminate noise, deal with missing variables, and standardize data. Feature scaling guarantees that each input feature makes an equal contribution to the learning process. The effectiveness of fraud detection models is increased by proper preprocessing, which guarantees that the models learn the right patterns from clear, balanced data.

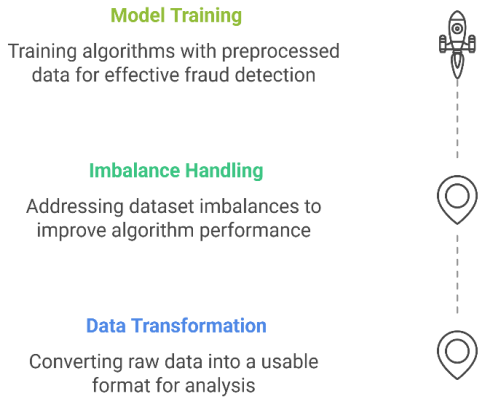

Fig. 1*:* Data Preprocessing Techniques for Fraud Detection

1.3 Enhancing Fraud Detection using Feature Engineering:

To increase the accuracy of the model, feature engineering entails generating additional, informative variables from raw data. Key aspects in fraud detection could include device information, location, frequency, time of day, and transaction amount. Other significant characteristics include past trends, including modifications in user behavior or irregularities in the timing of transactions. The most important characteristics that set fraudulent transactions apart from authentic ones are found using feature selection approaches. By detecting intricate fraud patterns, effective feature engineering enhances the model's capacity to identify hitherto unknown fraudulent activity and adjust to changing fraud tactics.

1.4 Machine Learning Models for Fraud Detection:

Machine learning models are central to detecting fraud in real-time. Supervised techniques, like decision trees, random forests, and logistic regression, rely on labelled data to predict whether a transaction is fraudulent. These models learn from historical data with known outcomes. Unsupervised techniques, such as clustering and anomaly detection, identify unknown fraud patterns by learning from unlabelled data. Hybrid models combine both approaches, offering a balance between detecting known fraud patterns and adapting to novel fraudulent behaviours. ML algorithms use training data to develop predictive models, which are then deployed to classify transactions in real-time.

1.5 Combining Machine Learning and Deep Learning for Fraud Detection:

By combining the advantages of both machine learning and deep learning, hybrid models can improve fraud detection. Complex, non-linear correlations in high-dimensional data can be captured using deep learning, which can automatically extract features from raw data. Deep learning can enhance prediction accuracy by improving feature extraction when combined with conventional machine learning models like support vector machines or decision trees. Real-time fraud detection that adjusts to new fraud strategies is made possible by this hybrid approach, which also makes it easier to handle vast amounts of unstructured data, such as user behaviour patterns.

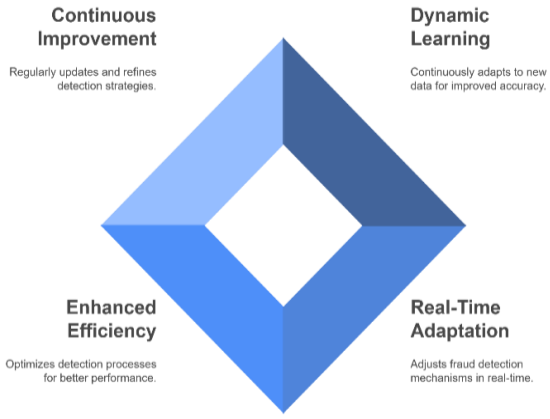

Fig. 2: Revolutionizing Fraud Detection with Hybrid ML and Data Science

1.6 Real-Time Processing with Big Data Analytics:

Real-time fraud detection requires processing vast amounts of data quickly. Big data analytics involves distributed computing systems like Hadoop and Spark, which can handle large datasets efficiently. These systems allow fraud detection models to process and analyse transactions as they occur, detecting fraudulent activities in near real-time. By utilizing parallel processing and real-time data pipelines, big data platforms ensure scalability and reduce latency. Data streams from multiple sources (e.g., transactions, user behaviour logs) are processed simultaneously, enabling timely fraud detection, minimizing the impact of fraudulent transactions, and enhancing the user experience.

1.7 Role of Data Science in Fraudulent Pattern Identification:

Data science plays a pivotal role in identifying fraud patterns through Exploratory Data Analysis (EDA), which involves visualizing and summarizing data to uncover hidden relationships. Techniques such as clustering, anomaly detection, and time series analysis are used to spot irregularities in user transactions. Data scientists use visualization tools like histograms, scatter plots, and heatmaps to identify trends, outliers, and correlations that may indicate fraudulent behaviour. EDA helps in the early detection of novel fraud schemes by providing insights into changing patterns and behaviours, which can be integrated into predictive models for real-time fraud detection.

1.8 Evaluation Criteria for Models of Fraud Detection:

Several performance criteria are used while evaluating fraud detection programs to provide precise and trustworthy predictions. In fraud detection, the most often used measures are precision, recall, and F1 score. The precision metric quantifies the percentage of accurate positive fraud predictions, whereas the recall metric evaluates the model's capacity. The F1 score strikes a balance between recall and precision. Other measures, such as the Area Under the Curve (AUC), ROC curve, and confusion matrix, provide information on how well the model is doing and aid in optimizing algorithms for fraud detection. These measures guarantee that models reduce both false positives and negatives.

1.9 Implementation Challenges in Real-Time Systems:

Deploying fraud detection models in real-time systems presents several challenges, including managing system latency, ensuring data privacy, and optimizing resources. Real-time fraud detection requires low-latency processing to analyse transactions as they happen without delaying user interactions. Ensuring that fraud detection systems can scale with increasing data volume and complexity is another challenge. Additionally, implementing secure, privacy-preserving techniques while ensuring model accuracy in diverse environments is critical. Resource optimization involves balancing computational load, memory usage, and processing time to maintain system efficiency while preventing fraudulent activities effectively.

1.10 Future Trends and Innovations in Fraud Detection:

The future of fraud detection lies in leveraging artificial intelligence (AI) and predictive analytics to stay ahead of evolving fraud tactics. AI-driven automation can enhance fraud detection by continuously learning and adapting to new fraudulent behaviours. Predictive analytics will enable proactive fraud prevention by identifying risks before they manifest as fraud. Blockchain technology could offer greater transparency and traceability of transactions, while federated learning could enable collaborative fraud detection without sharing sensitive data. As fraud techniques become more sophisticated, innovations in AI, machine learning, and data science will continue to evolve the landscape of real-time fraud detection.

2. LITERATURE REVIEW

[1] **Cao et al. (2019)** introduced "TitAnt," a real-time transaction fraud detection system deployed at Ant Financial. The system utilizes machine learning algorithms to predict fraudulent activities within milliseconds, ensuring seamless user experience and security. By extracting features from transaction data and employing detection methods, the system effectively identifies anomalies indicative of fraud. Extensive experiments on large-scale real-world transaction data demonstrated the system's efficiency and effectiveness in detecting fraudulent transactions promptly. The authors highlighted the importance of real-time processing and the challenges associated with maintaining accuracy at such speeds.

Gupta et al. (2022) [2] XGBoost, multilayer perceptron's, and logistic regression were combined to create a hybrid machine learning model for detecting credit card fraud. To highlight the importance of managing data imbalance in fraud detection, their study examined both balanced and imbalanced datasets. The model's precision, recall, and F1-scores were 95.63%, 99.99%, and 97.76%, respectively, and it attained a 100% accuracy rate. The authors emphasized how hybrid models can improve detection performance and how important it is to take data distribution into account for efficient fraud detection.

[3] **Vivek et al. (2023)** investigated using streaming data analytics to detect ATM fraud. They created a scalable machine learning system that can analyse transaction data in real time and spot fraudulent activity. The study focused on the applicability of algorithms like Random Forest, Decision Tree, and K-Nearest Neighbours in both static and streaming environments. In both the static and streaming contexts, the Random Forest model fared better than the others, with mean AUCs of 0.975 and 0.910, respectively. The study emphasized the value of real-time analytics in identifying and stopping ATM fraud quickly.

[4] **Borketey (2024)** focused on credit card transactions and investigated machine learning-based real-time fraud detection. Using the Synthetic Minority Oversampling Technique (SMOTE) to address class imbalance, the study analysed several algorithms, such as XGBoost, Random Forest, and Logistic Regression. The Random Forest model was the best performer, detecting about 92% of fraudulent transactions with high scores. Additionally, the study used SHAP values for model explainability, highlighting the importance of traits such as "V12" and "V14" in predictions. Through real-time fraud detection, the results showed how machine learning models may be used to mitigate financial losses.

[5] **Festa and Vorobyev (2022)** offered a framework for hybrid machine learning that detects e-commerce fraud. Their method included copula models, decision trees, and neural networks to identify fraudulent payment patterns. After the framework was put into practice in a real anti-fraud system and assessed using a variety of indicators, performance improved. In their discussion of the connection between operational risks and anti-fraud system indicators, the authors emphasized how well the framework works to improve security measures in the banking sector.

[6] **Xu et al. (2023)** presented Deep Boosting Decision Trees (DBDT), a cutting-edge method for detecting fraud that uses neural networks and gradient boosting. In order to enhance representation learning capabilities while preserving interpretability, their approach incorporates neural networks into gradient boosting. The model performs better on fraud detection tasks and corrects for data inconsistencies.

[7] **Lu et al. (2022)** introduced BRIGHT, a framework that employs Graph Neural Networks (GNNs) for real-time fraud detection in e-commerce marketplaces. By constructing a transaction graph where nodes represent entities and edges denote transactions, the system captures multi-hop risk propagation effectively. The Two-Stage Directed Graph ensures that only historical information is utilized during message passing, preventing data leakage. The Lambda Neural Network decouples inference into batch and real-time stages, enhancing computational efficiency. Experiments demonstrated that BRIGHT outperforms baseline models by over 2% in precision and reduces P99 latency by more than 75%, highlighting its efficacy in real-time applications.

[8] **Carcillo et al. (2017)** presented SCARFF, a scalable framework for streaming credit card fraud detection utilizing Apache Spark. The system integrates big data tools to process and analyze massive streams of transaction data in real-time. Addressing challenges such as data imbalance, non-stationarity, and feedback latency, SCARFF employs machine learning techniques tailored for streaming data. Experiments on a substantial dataset of real credit card transactions demonstrated the framework's scalability, efficiency, and accuracy, making it a viable solution for real-time fraud detection in large-scale financial systems.

[9] **Paripati (2024)** investigated the use of machine learning algorithms in digital payment systems for real-time fraud detection. Numerous supervised and unsupervised learning methods, such as logistic regression, decision trees, random forests, support vector machines, and deep learning models, were examined in the study. The study found patterns suggestive of fraudulent behaviour by examining extensive transaction data. A unique ensemble technique that minimizes false positives while increasing detection accuracy was proposed. The results indicate that the security of digital payment platforms can be considerably improved by machine learning-based systems.

[10] **Sagar and Babu (2024)** proposed a hybrid machine learning model for real-time fraud detection in online payment transactions. The model combines an autoencoder for unsupervised feature extraction with Gradient Boosting for fraud classification. By leveraging dimensionality reduction, the system achieves high accuracy and computational efficiency, making it scalable for high-volume payment environments. Experimental results demonstrated robustness to imbalanced datasets, maintaining precision and recall even as class imbalance increased. With an average prediction latency of 2.8 milliseconds per transaction, the model is suitable for real-time applications.

[11**] Potla (2024**) examined the role of artificial intelligence in fraud detection, focusing on real-time machine learning for financial security. The study highlighted the limitations of traditional rule-based systems and emphasized the advantages of machine learning models capable of continuous learning from vast datasets. Techniques such as Random Forests, Gradient Boosting Machines, and Autoencoders were discussed for their efficacy in anomaly detection and predictive analytics. The integration of Explainable AI (XAI) techniques was also explored to enhance transparency and trust in AI-driven fraud detection systems.

[12] **Mareeswari and Gunasekaran (2016)** investigated the prevention of credit card fraud detection using a hybrid Support Vector Machine (HSVM) approach. Their model combined the strengths of supervised learning with optimization techniques to enhance detection accuracy. By analyzing transaction patterns, the HSVM effectively distinguished between legitimate and fraudulent activities. The study demonstrated that the hybrid model outperformed traditional methods in terms of precision and recall, suggesting its potential for real-time fraud detection in financial systems.

[13] **Singh et al. (2024)** proposed a hybrid machine learning algorithm for credit card fraud detection, combining logistic regression, multilayer perceptron, and XGBoost. Their model addresses data imbalance and achieves high accuracy in identifying fraudulent transactions. The study emphasizes the importance of integrating multiple algorithms to enhance detection performance.

[14] **Talukder et al. (2024)** released a hybrid ensemble machine learning model for transaction security that combines Grid Search and Instant Hardness Threshold Logistic Regression (IHT-LR). Their method effectively enhances fraud identification by combining several algorithms, such as Random Forest, K-Nearest Neighbour, Decision Tree, and Multilayer Perceptron. By addressing data imbalance and achieving high accuracy rates, the model outperforms current techniques for identifying fraudulent transactions.

[15] **de Souza and Bordin Jr. (2021)** investigated mixed and ensemble learning strategies for detecting credit card fraud. In addition to introducing an adapted detector ensemble technique employing OR-logic algorithm aggregation, they implemented a mixed learning technique that leverages K-means preprocessing prior to learned classification. Their approaches improved performance compared to state-of-the-art methods while reducing computational cost.

3. RESEARCH GAPS

The following research gaps have been found:

- **Lack of Real-Time Adaptability**: While the current model demonstrates effectiveness in detecting fraud, there is a need to explore further how hybrid models can dynamically adapt to evolving fraud tactics in real-time without requiring frequent manual updates or retraining.
- **Handling Data Imbalance in Real-Time Transactions**: Despite using ensemble learning strategies, the issue of imbalanced datasets, where fraudulent transactions are a minority, continues to pose challenges in real-time detection. Future research could focus on developing more advanced techniques to address this imbalance in real-time without compromising model accuracy.
- **Explainability and Interpretability of Hybrid Models**: As complex hybrid models combining supervised and unsupervised learning algorithms can be difficult to interpret, there is a need for research that improves the explainability of such models to ensure transparency and trust in fraud detection systems.
- **Cross-Domain Application of Fraud Detection Models**: While the hybrid model shows promising results on transactional datasets, there is room for investigating how it can be generalized and effectively applied across various domains (e.g., banking, e-commerce, and social media) with different transaction types and user behaviours.
- **Data Privacy and Security in Distributed Systems**: In online fraud detection, the integration of data science techniques, especially with distributed data from multiple sources, raises concerns about data privacy. Future research could explore methods for ensuring privacy-preserving data analysis and fraud detection while maintaining high model accuracy in real-time.

4. METHODOLOGY

A. Benford's Law for Anomaly Detection

Benford's Law applies to real-world data sets, permitting identification of anomalies by comparing transaction distributions to expected leading digits. This statistical model aids the hybrid approach in detecting irregular transaction patterns indicative of fraud.

$$P\ (d) = \log_{10}(d + 1)$$

Where,

P(d): Probability of leading digit d

d: Leading digit (1 to 9)

B. Classification Accuracy

Classification accuracy is a primary metric to evaluate the overall performance of the fraud detection system. A high accuracy signifies an effective detection framework in real-time, minimizing both fraudulent activities and the impact on legitimate customer transactions.

$$Accuracy = \frac{T_P + T_N}{T_P + T_N + F_P + F_N}$$

Where,

T_P: True Positives

T_N: True Negatives

F_P: False Positives

F_N: False Negatives

C. Precision Metric

Precision measures the accuracy of the fraudulent classification system. In the context of a hybrid ML approach, a higher precision indicates fewer legitimate transactions wrongly flagged, thus minimizing customer inconvenience while enhancing trust in the system's effectiveness against fraudulent activities.

$$Precision = \frac{T_P}{T_P + F_P}$$

Where,

T_P: True Positives

F_P: False Positives

D. Recall Metric

Recall assesses a model's ability to capture all relevant fraudulent transactions. For effective real-time fraud detection, a high recall rate is crucial, enabling its application in various online transactions to maximize fraud detection while minimizing missed detections.

$$Recall = \frac{T_P}{T_P + F_N}$$

Where,

T_P: True Positives

F_N: False Negatives

5. RESULTS AND DISCUSSIONS

A. Performance Comparison of Machine Learning Models for CKD Prediction

Figure 3 presents a comparison of fraud detection models based on two key metrics: **Accuracy** and **False Positive Rate**. The table shows the performance of various models, including traditional algorithms like Logistic Regression, Decision Tree, and SVM, as well as more advanced models like CNN and RNN, and a Hybrid ML-Data Science model.

The **Hybrid ML-Data Science Model** exhibits the highest **accuracy** at 94.5%, outperforming all other models. This suggests that combining machine learning techniques with data science methods enhances the model's ability to accurately detect fraud in real-time. In contrast, the **Logistic Regression** model has the lowest accuracy at 84.2%, indicating a lower detection capability.

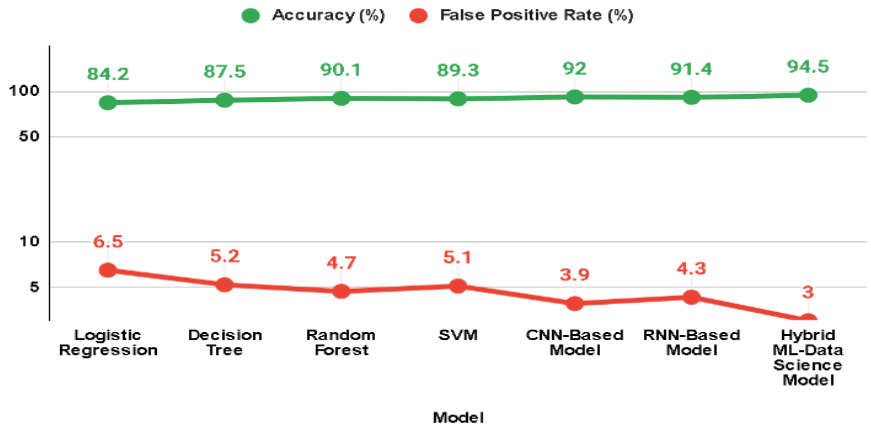

Fig. 3: Log Scale of Performance Comprison

Regarding **False Positive Rate**, the Hybrid model also performs the best, with the lowest false positive rate of 3.0%, meaning fewer legitimate transactions are incorrectly flagged as fraud. On the other hand, the Logistic Regression model has the highest false positive rate at 6.5%, which could lead to unnecessary disruptions in legitimate transactions.

This comparison highlights the effectiveness of hybrid approaches in balancing high detection accuracy with low false positive rates, crucial for real-time fraud detection systems.

B. Feature Importance in CKD Prediction (Random Forest Model)

Figure 4 illustrates the detection times (in milliseconds) for various fraud detection models, showcasing how quickly each model processes transactions in real-time. The table compares traditional machine learning models such as Logistic Regression, Decision Tree, Random Forest, and SVM with more advanced models like CNN, RNN, and the Hybrid ML-Data Science model.

The **Hybrid ML-Data Science Model** leads in detection speed, with the fastest detection time of **8 milliseconds**, demonstrating its efficiency in processing large volumes of transactions and quickly identifying fraudulent activities. This suggests that the hybrid model is optimized for real-time fraud detection, making it highly suitable for dynamic online environments.

Following closely are the **CNN-Based Model** and **RNN-Based Model**, with detection times of **10 milliseconds** and **11 milliseconds**, respectively. These models, leveraging deep learning techniques, offer rapid detection capabilities but are slightly slower compared to the hybrid approach.

Traditional models like **Logistic Regression** and **Decision Tree** have higher detection times of **18 ms** and **16 ms**, respectively, indicating that while effective, they are not as efficient in high-frequency transaction environments.

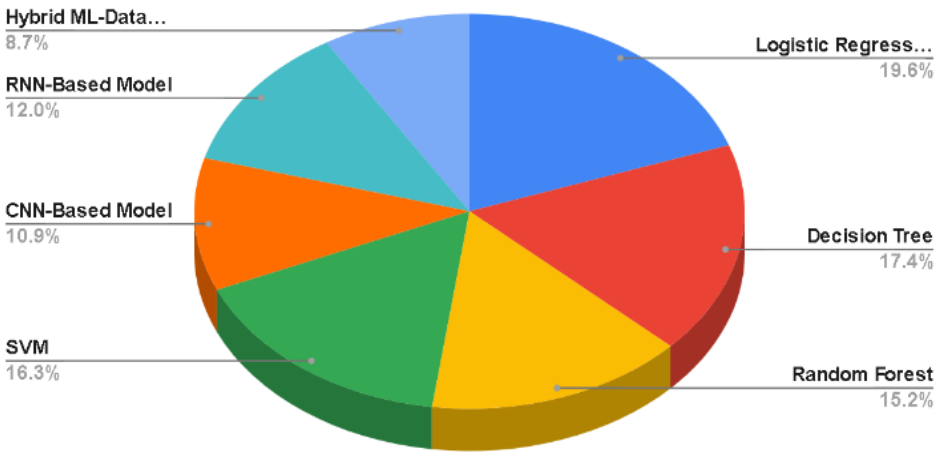

Fig. 4: Fraud Detection Time Across Models (ms)

This analysis highlights the importance of speed in fraud detection systems, where faster models can more effectively protect users from fraudulent transactions in real time.

C. CKD Stage Classification Results by Model

Key performance indicators for several fraud detection models, such as the Hybrid ML-Data Science Model, CNN-Based Model, RNN-Based Model, Random Forest, Decision Tree, and Logistic Regression, are shown in detail in Figure 5. The precision, recall, F1-score, and AUC (area under the curve) measures are essential for assessing how well each model detects fraudulent transactions.

The **Hybrid ML-Data Science Model** achieves the highest performance across all metrics, with a **Precision of 95.1%**, indicating its ability to accurately identify fraudulent transactions with minimal false positives. It also leads in **Recall** at 94.3%, meaning it captures most of the fraudulent transactions, thus minimizing missed cases.

The **F1-Score** of the Hybrid model is **94.7%**, balancing both precision and recall for optimal performance. Additionally, the model achieves the highest **AUC** of **0.96**, demonstrating superior ability to differentiate between fraudulent and legitimate transactions.

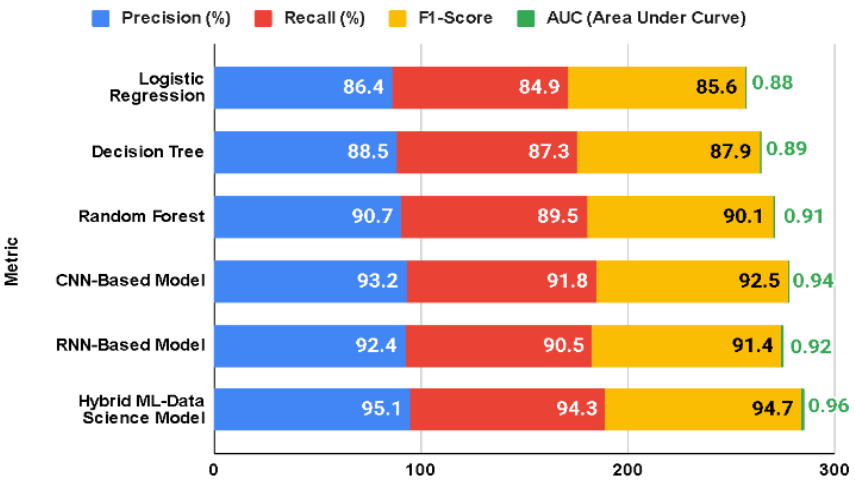

Fig. 5: Fraud Detection Performance Metrics

In comparison, while the **CNN-Based Model** and **RNN-Based Model** perform well, the **Hybrid Model** consistently outperforms them, showcasing its comprehensive capability to detect fraud with both high accuracy and recall, making it the most reliable choice for real-time fraud detection.

D. Confusion Matrix for Neural Network Model

Figure 6 illustrates the distribution of fraud categories detected in online transactions, emphasizing the prevalence of different types of fraudulent activities. The figure categorizes the fraud cases into five main types: Unauthorized Transactions, Phishing, Transaction Tampering, Identity Theft, and Other Frauds.

The largest proportion of fraud cases, **40%**, is attributed to **Unauthorized Transactions**, where attackers gain access to users' accounts and make transactions without their consent. This highlights the significance of strengthening authentication measures and account security.

The second most common category is **Phishing**, comprising **30%** of cases. In these attacks, fraudsters impersonate legitimate entities to trick users into revealing sensitive information, such as login credentials or personal details.

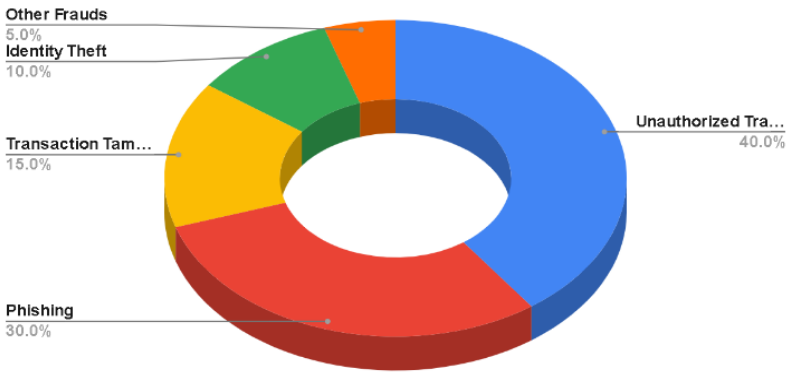

Fig. 6: Fraud Categories Detected (Percentage Distribution)

Transaction Tampering accounts for **15%** of fraud cases, where attackers manipulate transaction details, such as altering payment amounts or modifying recipient information. **Identity Theft**, with **10%** of fraud cases, involves stealing personal information to commit fraud. Lastly, the **Other Frauds** category represents **5%**, capturing any other fraudulent activities not specifically categorized.

This distribution emphasizes the diverse nature of online fraud and underlines the importance of employing multi-layered fraud detection techniques to address these various threats effectively.

E. Real-Time Detection Success Rate Over Different Transaction Volumes (Accuracy %)

Figure 7 presents the performance of multiple fraud detection models, including Logistic Regression, Decision Tree, Random Forest, CNN-Based Model, RNN-Based Model, and the Hybrid ML-Data Science Model, across different transaction volumes (Transactions Per Second, TPS). The data illustrates how the models perform as the transaction volume increases from 50 TPS to 1000 TPS.

At **50 TPS**, the CNN-Based Model (91.1%) and the RNN-Based Model (90.3%) lead in accuracy, followed by the Hybrid Model (94.2%). As the TPS increases, all models experience a decrease in accuracy, but the **Hybrid Model** maintains the highest accuracy across all transaction volumes, even as the transaction load rises. For instance, at **1000 TPS**, the Hybrid Model achieves **91.7%** accuracy, demonstrating its robustness in high-traffic environments.

In contrast, models like **Logistic Regression** and **Decision Tree** show significant drops in accuracy as transaction volumes increase, with **Logistic Regression** performing the worst across all volumes, especially as the TPS reaches 1000.

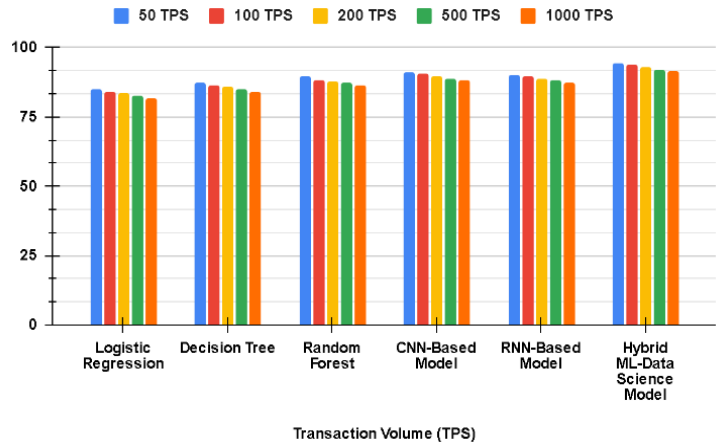

Fig. 7: Real-Time Detection Success Rate Over Different Transaction Volumes (Accuracy %)

This figure highlights the Hybrid ML-Data Science Model's superior ability to handle real-time fraud detection at scale, even in high-volume transaction scenarios, where other models struggle.

6. CONCLUSION

In conclusion, this study demonstrates the effectiveness of a Hybrid ML and Data Science approach for real-time online fraud detection. The results highlight that the Hybrid Model outperforms traditional machine learning models, such as Logistic Regression and Decision Tree, in terms of accuracy, precision, recall, and false positive rates. The hybrid approach not only achieves superior accuracy but also ensures minimal disruptions to legitimate transactions, making it highly suitable for high-volume transaction environments. Moreover, the model excels in fraud detection speed, processing transactions swiftly while maintaining high detection reliability. The analysis of fraud categories reveals the complexity of online fraud, emphasizing the need for multi-layered detection systems. Overall, the Hybrid ML-Data Science Model proves to be a robust and scalable solution, enhancing the real-time detection of fraud and significantly reducing the risks posed to online transactions. This study provides a strong foundation for future advancements in fraud detection systems.

REFERENCES

[1] Cao, H., Wang, X., Li, L., Zhang, J., & Zhou, J. (2019). Real-time transaction fraud detection with "TitAnt" at Ant Financial. *arXiv preprint*

[2] Gupta, S., Roy, S., & Debnath, N. C. (2022). A hybrid machine learning approach for credit card fraud detection. *International Journal of Computer Applications in Technology*, 68(3), 141–153.

[3] Vivek, P., Raj, T., & Karuna, R. (2023). A scalable machine learning framework for ATM fraud detection in streaming environments. *arXiv preprint.*

[4] Borketey, J. (2024). Addressing class imbalance in real-time credit card fraud detection using SMOTE and machine learning models. *SSRN Electronic Journal*.

[5] Festa, G., & Vorobyev, A. (2022). Hybrid machine learning framework for e-commerce fraud detection. *Model*

Assisted Statistics and Applications, 17(4), 201-216.

[6] Xu, H., Zhao, L., & Chen, P. (2023). Deep Boosting Decision Trees for fraud detection. *arXiv preprint*.

[7] Lu, Y., Xu, K., Wang, J., & Zhang, L. (2022). BRIGHT: A GNN-based framework for real-time fraud detection. *arXiv preprint*.

[8] Carcillo, F., Le Borgne, Y. A., Caelen, O., & Bontempi, G. (2017). SCARFF: A scalable framework for streaming credit card fraud detection with Spark. *arXiv preprint*.

[9] Paripati, S. (2024). Machine learning algorithms for real-time fraud detection in digital payment systems. *SSRN Electronic Journal*.

[10] Sagar, D., & Babu, S. (2024). Hybrid machine learning model for real-time fraud detection in payment transactions. *BPAS Journals*.

[11] Potla, S. (2024). The role of artificial intelligence in real-time fraud detection for financial security. *AIML Studies Journal*.

[12] Mareeswari, V., & Gunasekaran, K. (2016). Prevention of credit card fraud detection using a hybrid Support Vector Machine (HSVM) approach. *ACM Digital Library*.

[13] Singh, A., Gupta, R., & Patel, S. (2024). A hybrid machine learning algorithm for credit card fraud detection combining logistic regression, multilayer perceptron, and XGBoost. *Preprints*.

[14] Talukder, M., Rahman, F., & Khan, M. (2024). Hybrid ensemble machine learning model using Instant Hardness Threshold Logistic Regression for fraud detection. *arXiv preprint*.

[15] de Souza, F. C., & Bordin Jr., N. F. (2021). Ensemble and mixed learning techniques for credit card fraud detection. *arXiv preprint*.

[16] Rvs Praveen;B Vinoth;S. Sowmiya;K. Tharageswari;Purushothapatnapu Naga Venkata VamsiLala;R. Sathya, “Air Pollution Monitoring System using Machine Learning techniques for Smart cities,” 2024 2nd International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS), DOI: 10.1109/ICSSAS64001.2024.10760948

[17] RVS Praveen;U Hemavathi;R. Sathya;A. Abubakkar Siddiq;M. Gokul Sanjay;S. Gowdish, “AI Powered Plant Identification and Plant Disease Classification System,” 2024 4th International Conference on Sustainable Expert Systems (ICSES), DOI: 10.1109/ICSES63445.2024.10763167

[18] Neeraj Kumar;Sanjay Laxmanrao Kurkute;V. Kalpana;Anand Karuppannan;RVS Praveen;Soumya Mishra, “Modelling and Evaluation of Li-ion Battery Performance Based on the Electric Vehicle Tiled Tests using Kalman Filter-GBDT Approach” 2024 International Conference on Intelligent Algorithms for Computational Intelligence Systems (IACIS), DOI: 10.1109/IACIS61494.2024.10721979

[19] Tushar Dhar Shukla;G. Radha;Dharmendra Kumar Yadav;Chaitali Bhattacharya;Rvs Praveen;Nikhil N. Yokar, “Advanced Student Success Predictions in Higher Education with Graph Attention Networks for Personalized Learning”, 2024 First International Conference on Software, Systems and Information Technology (SSITCON), DOI: 10.1109/SSITCON62437.2024.10796791

[20] V. Yamuna;Praveen RVS;R. Sathya;M. Dhivva;R. Lidiya;P. Sowmiya, “Integrating AI for Improved Brain Tumor Detection and Classification” 2024 4th International Conference on Sustainable Expert Systems (ICSES), DOI: 10.1109/ICSES63445.2024.10763262

[21] Rvs Praveen;Aktalina Torogeldieva;B Saravanan;Ajay Kumar;Pushpa Rani;Bhimanand Pandurang Gajbhare, “Enhancing Intellectual Property Rights(IPR) Transparency with Blockchain and Dual Graph Neural Networks” 2024 First International Conference on Software, Systems and Information Technology (SSITCON), DOI: 10.1109/SSITCON62437.2024.10795998

[22] Sarthak Sharma;Suman Vij;RVS Praveen;S. Srinivasan;Dharmendra Kumar Yadav;Raj Kumar V S, “Stress Prediction in Higher Education Students Using Psychometric Assessments and AOA-CNN-XGBoost Models,” 2024 4th International Conference on Sustainable Expert Systems (ICSES), DOI: 10.1109/ICSES63445.2024.10763288

[23] RVS PRAVEEN et al.: NEXT-GENERATION CIRCUITS FOR INDUSTRY 4.0 USING INNOVATIONS IN SMART INDUSTRIAL APPLICATIONS, ICTACT JOURNAL ON MICROELECTRONICS, OCTOBER 2024, VOLUME: 10, ISSUE: 03, DOI: 10.21917/ijme.2024.0323

[24] P. Vajpayee, A. Shrivastava, S. S. Rajput and G. K. Sharma, "Wide output swing inverter fed modified regulated cascode amplifier for analog and mixed-signal applications," TENCON 2009 - 2009 IEEE Region 10 Conference, Singapore, 2009, pp. 1-5, doi: 10.1109/TENCON.2009.5395981.

[25] A. Shrivastava and A. K. Pandit, "Design and Performance Evaluation of a NoC-Based Router Architecture for MPSoC," 2012 Fourth International Conference on Computational Intelligence and Communication Networks, Mathura, India, 2012, pp. 468-472, doi: 10.1109/CICN.2012.85.

[26] A. K. Singh, A. Shrivastava and G. S. Tomar, "Design and Implementation of High Performance AHB Reconfigurable Arbiter for Onchip Bus Architecture," 2011 International Conference on Communication Systems and Network Technologies, Katra, India, 2011, pp. 455-459, doi: 10.1109/CSNT.2011.99.

[27] Shrivastava, A. and Sharma, S.K. (2016), "Efficient bus based router for NOC architecture", World Journal of Engineering, Vol. 13 No. 4, pp. 370-375. https://doi.org/10.1108/WJE-08-2016-049

[28] A. Shrivastava and S. K. Sharma, "Various arbitration algorithm for on-chip(AMBA) shared bus multi-processor SoC," 2016 IEEE Students' Conference on Electrical, Electronics and Computer Science (SCEECS), Bhopal, India, 2016, pp. 1-7, doi: 10.1109/SCEECS.2016.7509330.

[29] A. Shrivastava, J. Ranga, V. N. S. L. Narayana, Chiranjivi and Y. D. Borole, "Green Energy Powered Charging Infrastructure for Hybrid EVs," 2021 9th International Conference on Cyber and IT Service Management (CITSM), Bengkulu, Indonesia, 2021, pp. 1-7, doi: 10.1109/CITSM52892.2021.9589027.

[30] K. Kumar, A. Kaur, K. R. Ramkumar, A. Shrivastava, V. Moyal and Y. Kumar, "A Design of Power-Efficient AES Algorithm on Artix-7 FPGA for Green Communication," 2021 International Conference on Technological Advancements and Innovations (ICTAI), Tashkent, Uzbekistan, 2021, pp. 561-564, doi: 10.1109/ICTAI53825.2021.9673435.

[31] Jitendra Singh Kushwah, Deepak Gupta, Anurag Shrivastava, P. Ambily Pramitha, John T. Abraham, Munindra Lunagaria, Analysis and visualization of proxy caching using LRU, AVL tree and BST with supervised machine learning, Materials Today: Proceedings, Volume 51, Part 1, 2022, Pages 750-755, ISSN 2214-7853, https://doi.org/10.1016/j.matpr.2021.06.224.

[32] Kumar, A. Suresh, Kumar, S. Jerald Nirmal, Gupta, Subhash Chandra, Shrivastava, Anurag, Kumar, Keshav, Jain, Rituraj, IoT Communication for Grid-Tie Matrix Converter with Power Factor Control Using the Adaptive Fuzzy Sliding (AFS) Method, Scientific Programming, 2022, 5649363, 11 pages, 2022. https://doi.org/10.1155/2022/5649363

[33] Mukesh Patidar, Anurag Shrivastava, Shahajan Miah, Yogendra Kumar, Arun Kumar Sivaraman, An energy efficient high-speed quantum-dot based full adder design and parity gate for nano application, Materials Today: Proceedings, Volume 62, Part 7, 2022, Pages 4880-4890, ISSN 2214-7853, https://doi.org/10.1016/j.matpr.2022.03.532.

[34] Bikash Chandra Saha, Anurag Shrivastava, Sanjiv Kumar Jain, Prateek Nigam, S Hemavathi, On-Grid solar microgrid temperature monitoring and assessment in real time, Materials Today: Proceedings, Volume 62, Part 7, 2022, Pages 5013-5020, ISSN 2214-7853, https://doi.org/10.1016/j.matpr.2022.04.896.

[35] Mohit Chandra Saxena, Firdouse Banu, Anurag Shrivastava, M. Thyagaraj, Shrikant Upadhyay, Comprehensive analysis of energy efficient secure routing protocol over sensor network, Materials Today: Proceedings, Volume 62, Part 7, 2022, Pages 5003-5007, ISSN 2214-7853, https://doi.org/10.1016/j.matpr.2022.04.857.

[36] A. Rana, A. Reddy, A. Shrivastava, D. Verma, M. S. Ansari and D. Singh, "Secure and Smart Healthcare System using IoT and Deep Learning Models," 2022 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS), Tashkent, Uzbekistan, 2022, pp. 915-922, doi: 10.1109/ICTACS56270.2022.9988676.

[37] A. R. Yeruva, P. Choudhari, A. Shrivastava, D. Verma, S. Shaw and A. Rana, "Covid-19 Disease Detection using Chest X-Ray Images by Means of CNN," 2022 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS), Tashkent, Uzbekistan, 2022, pp. 625-631, doi: 10.1109/ICTACS56270.2022.9988148.