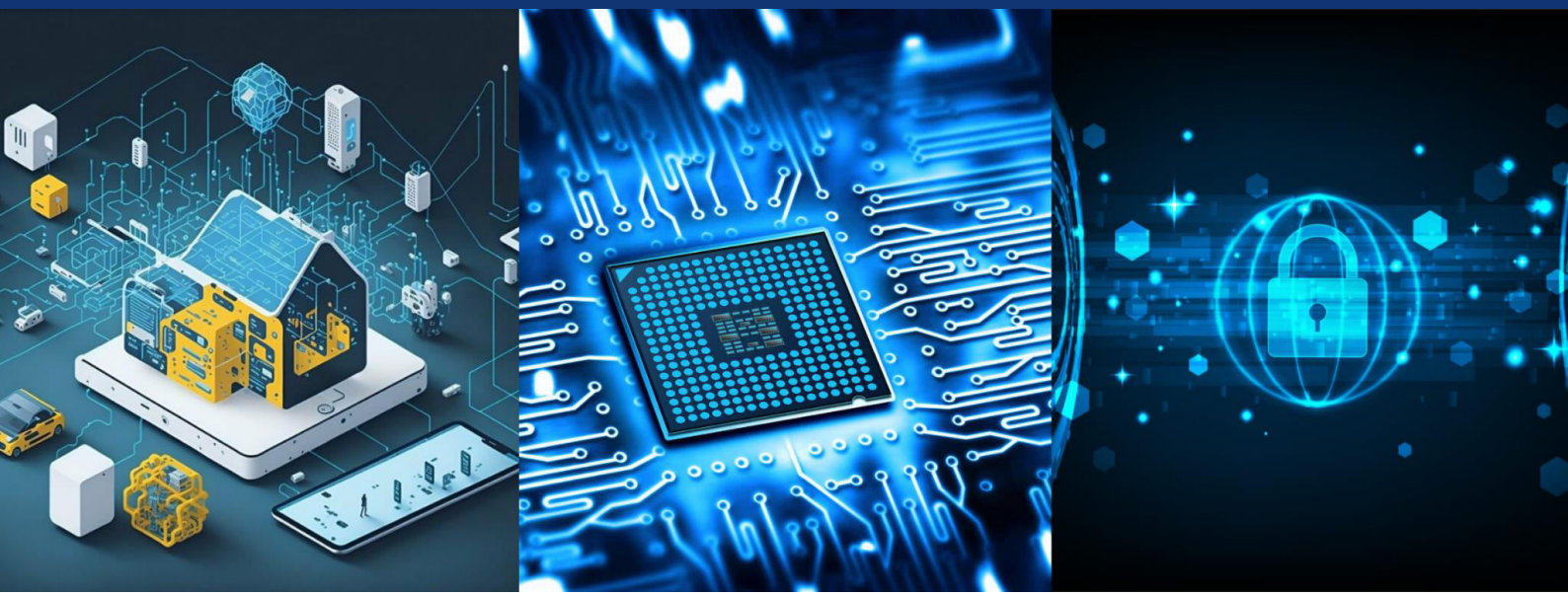
IJIRCCE

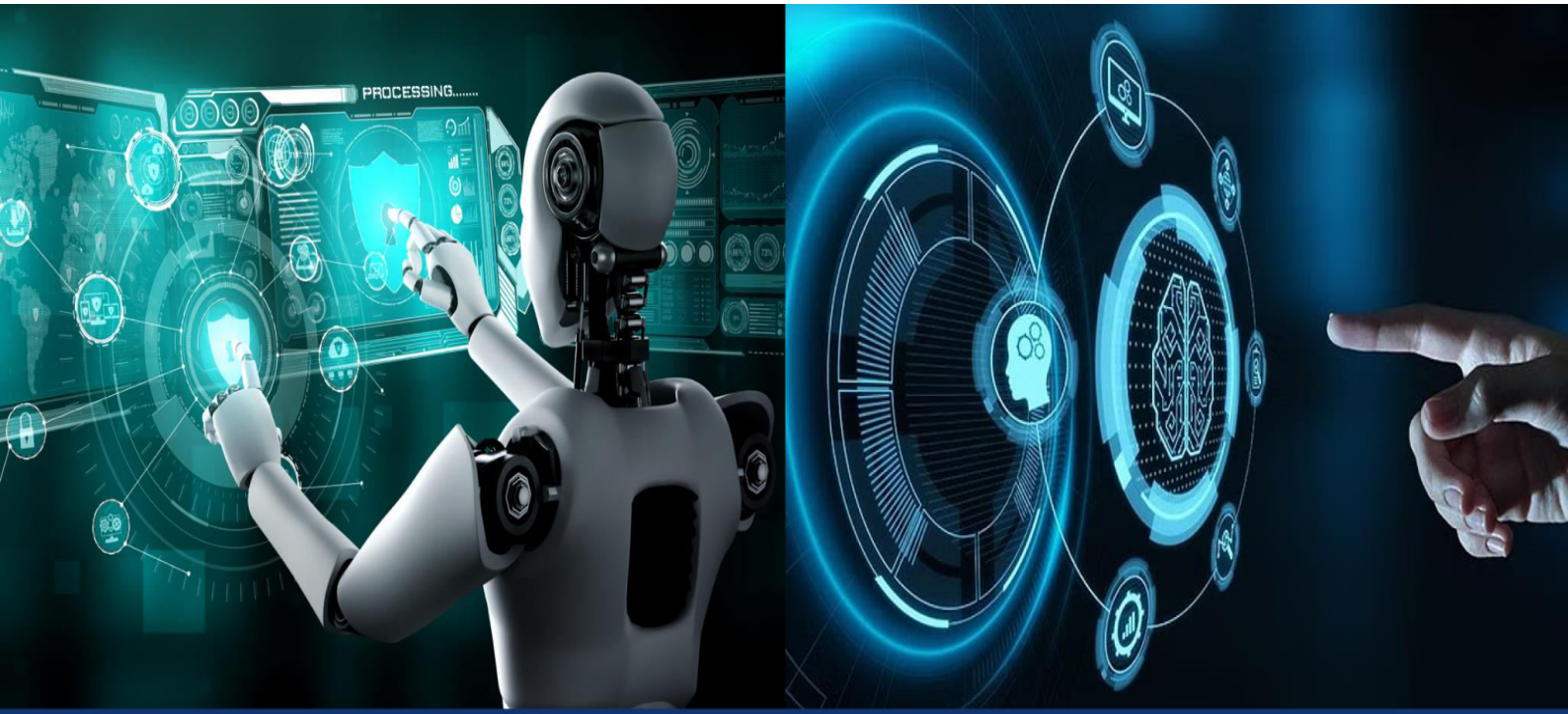
ISSN(O): 2320-9801

ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.771

Volume 14, Issue 4, April 2026

© 2026 IJIRCCE | Volume 14, Issue 4, April 2026 DOI: 10.15680/IJIRCCE.2026.1404084

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Credit Card Fraud Detection Using Machine Learning

Misba Tabassum[1], Shaheen M[1], Naziya Banu[1], Mehek S[1], Prof. Archana K N[2],

UG Students, Dept. of CSE, Jain Institute of Technology, Davangere, Karnataka, India[1]

Assistant Professor, Dept. of CSE, Jain Institute of Technology, Davangere, Karnataka, India[2]

ABSTRACT: Credit card fraud has become a major concern in the modern financial ecosystem due to the rapid growth of digital transactions and online payment systems. Financial institutions face significant losses every year as fraudsters continuously develop sophisticated techniques to exploit system vulnerabilities. Traditional fraud detection methods, which rely on static rule-based approaches, are often inefficient in identifying new and evolving fraud patterns, leading to delayed detection and increased financial risk.

This paper presents a machine learning-based credit card fraud detection system that leverages classification algorithms to identify fraudulent transactions in real-time. The proposed system utilizes algorithms such as Logistic Regression, Decision Tree, and Random Forest to analyze transaction data and classify them as legitimate or fraudulent. The dataset used in this study is highly imbalanced, with fraudulent transactions representing a very small portion of the total data. To address this issue, data preprocessing techniques such as normalization and resampling are applied to improve model

KEYWORDS: Credit Card Fraud Detection, Machine Learning, Classification Algorithms, Random Forest, Financial Security, Data Imbalance, Transaction Analysis

I. INTRODUCTION

The rapid advancement of digital technology has transformed the way financial transactions are conducted, with credit cards becoming one of the most widely used payment methods across the globe. The convenience and accessibility of credit card transactions have significantly contributed to the growth of e-commerce and online banking services. However, this rapid expansion has also led to an increase in fraudulent activities, posing serious challenges to financial institutions and customers.

Credit card fraud refers to the unauthorized use of a credit card or its information to perform transactions without the consent of the cardholder. Fraudsters employ various techniques such as phishing, card skimming, and identity theft to gain access to sensitive information.

Conventional fraud detection systems are primarily rule-based, relying on predefined rules and thresholds to identify suspicious transactions. While these systems are effective in detecting known fraud patterns, they lack the flexibility to adapt to new and unknown threats. This limitation highlights the need for intelligent systems that can learn from data and detect anomalies dynamically.

II. RELATED WORK

In recent years, significant research has been conducted in the field of credit card fraud detection using various computational techniques. Traditional approaches primarily relied on statistical methods and rule-based systems, which were limited in their ability to detect complex fraud patterns.

Machine learning techniques have gained popularity due to their ability to learn from data and adapt to new patterns. Logistic Regression has been widely used as a baseline model for binary classification tasks due to its simplicity and efficiency. Decision Trees provide an intuitive approach to classification by splitting data based on feature values, making them easy to interpret. However, they are prone to overfitting when dealing with complex datasets.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Support Vector Machines (SVM) have also been applied for fraud detection, offering high accuracy in certain cases. However, they require significant computational resources and are not suitable for large-scale datasets.

Random Forest, an ensemble learning technique, has emerged as one of the most effective methods for fraud detection. By combining multiple decision trees, it reduces overfitting and improves prediction accuracy.

III. PROPOSED SOLUTION

The proposed credit card fraud detection system is designed to intelligently identify fraudulent transactions using machine learning techniques combined with effective data preprocessing and decision-making strategies. The system follows a structured pipeline that integrates data handling, model training, prediction, and evaluation to ensure accurate and reliable fraud detection.

The overall objective of the proposed solution is to develop a robust and scalable system capable of handling large volumes of transaction data while maintaining high detection accuracy and minimizing false positives. The system is specifically designed to address the challenges associated with imbalanced datasets, evolving fraud patterns, and real-time processing requirements.

3.1 Data Acquisition and Preprocessing

The first stage of the system involves collecting transaction data from a reliable dataset.
Since real-world financial datasets often contain noise, missing values, and inconsistencies, preprocessing is a critical step in ensuring data quality. The preprocessing phase includes:

- Handling missing or null values
- Removing duplicate entries
- Normalizing numerical features to a standard scale
- Converting categorical data into numerical format (if applicable)

One of the major challenges in fraud detection is the class imbalance problem, where fraudulent transactions constitute a very small percentage of the dataset. To address this issue, techniques such as under sampling of majority class or oversampling of minority class (e.g., SMOTE) are applied. This helps the model learn patterns more effectively and improves detection performance.

3.2 Feature Selection and Engineering

Feature selection plays a crucial role in improving the efficiency and accuracy of the model. Not all features contribute equally to fraud detection, and some may introduce noise or redundancy.
In this system, important features are selected based on statistical analysis and their correlation with the target variable. Feature engineering techniques may also be applied to create new meaningful features, such as:

- Transaction frequency
- Average transaction amount
- Time gap between transactions

These derived features help in capturing user behavior more effectively, which improves the model's ability to detect anomalies.

3.3 Model Training and Classification

The core component of the proposed system is the machine learning model used for classification. Multiple algorithms are implemented and compared to identify the best-performing model.
Logistic Regression
Logistic Regression is used as a baseline model for binary classification. It estimates the probability of a transaction being fraudulent based on input features.
Decision Tree
Decision Tree is a supervised learning algorithm that splits the dataset into branches based on feature values. It is easy to interpret and can handle non-linear relationships.
Random Forest
Random Forest is an ensemble learning method that combines multiple decision trees to improve performance.
The advantages of Random Forest include:

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

- High accuracy
- Reduced overfitting
- Ability to handle large datasets
- Robustness to noise

Due to these properties, Random Forest is selected as the primary model in the proposed system.

3.4 Model Evaluation and Performance Metrics

To evaluate the effectiveness of the models, several performance metrics are used:

- Accuracy: Measures overall correctness of the model
- Precision: Indicates how many detected frauds are actually fraud
- Recall: Measures how many actual fraud cases are detected
- F1-score: Harmonic mean of precision and recall

Since fraud detection is a sensitive application, recall and precision are more important than accuracy, as missing a fraud case can lead to significant financial loss.

3.5 Fraud Detection Workflow

The complete workflow of the system can be summarized as follows:

1. Input transaction data
2. Perform preprocessing and normalization
3. Apply feature selection
4. Train machine learning models
5. Evaluate models using test data

This structured approach ensures efficient processing and accurate classification of transactions.

3.6 System Advantages

The proposed system offers several advantages:

- High detection accuracy
- Ability to handle imbalanced datasets
- Scalable for large transaction volumes
- Adaptable to new fraud patterns
- Suitable for real-time implementation

IV. SIMULATION RESULTS

The system was tested using a real-world dataset containing credit card transactions. The performance of different models was compared based on evaluation metrics.

The Random Forest model achieved the highest accuracy and recall, making it the most effective model for fraud detection. Logistic Regression provided a good baseline, while Decision Tree showed moderate performance.

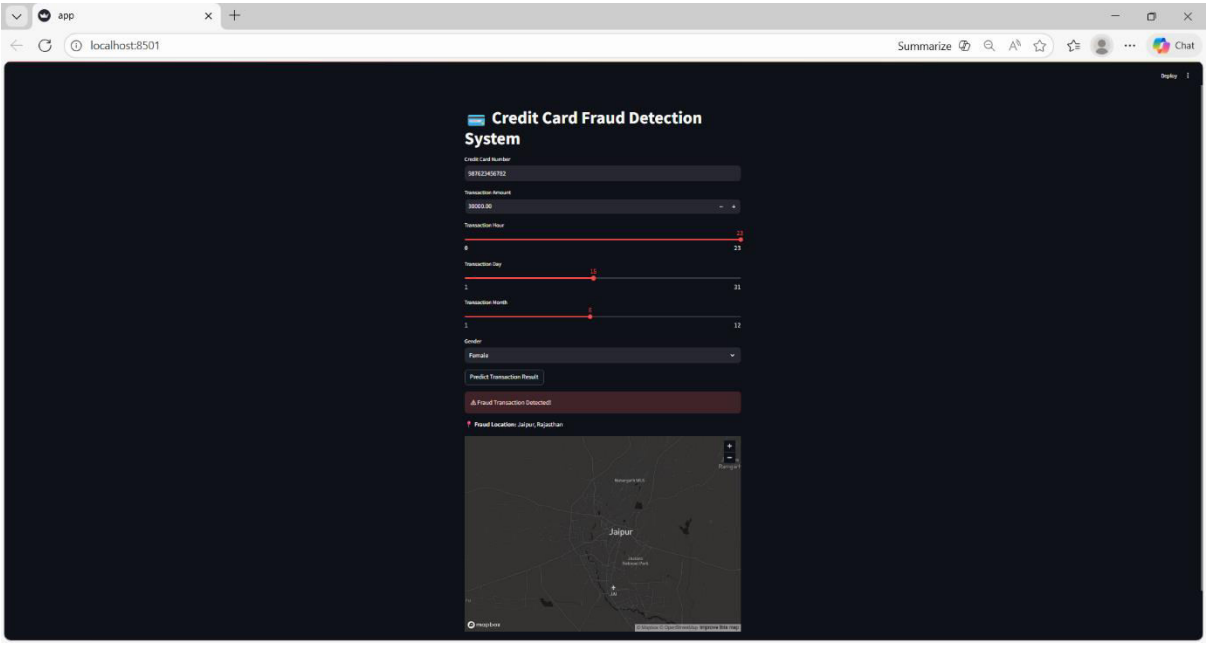
The results demonstrate that ensemble learning techniques are more effective in handling complex and imbalanced datasets.

© 2026 IJIRCCE | Volume 14, Issue 4, April 2026 DOI: 10.15680/IJIRCCE.2026.1404084

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



V. CONCLUSION AND FUTURE WORK

This paper presents a machine learning-based system for credit card fraud detection, which aims to improve the accuracy and efficiency of identifying fraudulent transactions. The proposed approach utilizes classification algorithms such as Logistic Regression, Decision Tree, and Random Forest to analyze transaction data and distinguish between legitimate and fraudulent activities. By applying data preprocessing techniques and handling class imbalance, the system enhances its ability to detect rare fraud cases effectively.

In future work, the system can be further improved by integrating advanced techniques such as deep learning models to capture more complex patterns in transaction data. Additionally, implementing real-time fraud detection using APIs and cloud-based systems can make the solution more practical for real-world applications. Other enhancements may include reducing false positive rates, incorporating explainable AI for better transparency, and using larger datasets to improve model performance. These improvements can help in developing a more robust and scalable fraud detection system.

REFERENCES

1. 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. https://doi.org/10.1016/j.eswa.2014.02.026
2. S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. Westland, “Data mining for credit card fraud: A comparative study,” *Decision Support Systems*, vol. 50, no. 3, pp. 602–613, 2011. https://doi.org/10.1016/j.dss.2010.08.008
3. 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,” *Decision Support Systems*, vol. 50, no. 3, pp. 559–569, 2011. https://doi.org/10.1016/j.dss.2010.08.006
4. J. West and M. Bhattacharya, “Intelligent financial fraud detection: A comprehensive review,” *Computers & Security*, vol. 57, pp. 47–66, 2016. https://doi.org/10.1016/j.cose.2015.09.005
5. V. Jurgovsky et al., “Sequence classification for credit-card fraud detection,” *Expert Systems with Applications*, vol. 100, pp. 234–245, 2018. https://doi.org/10.1016/j.eswa.2018.01.037

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details