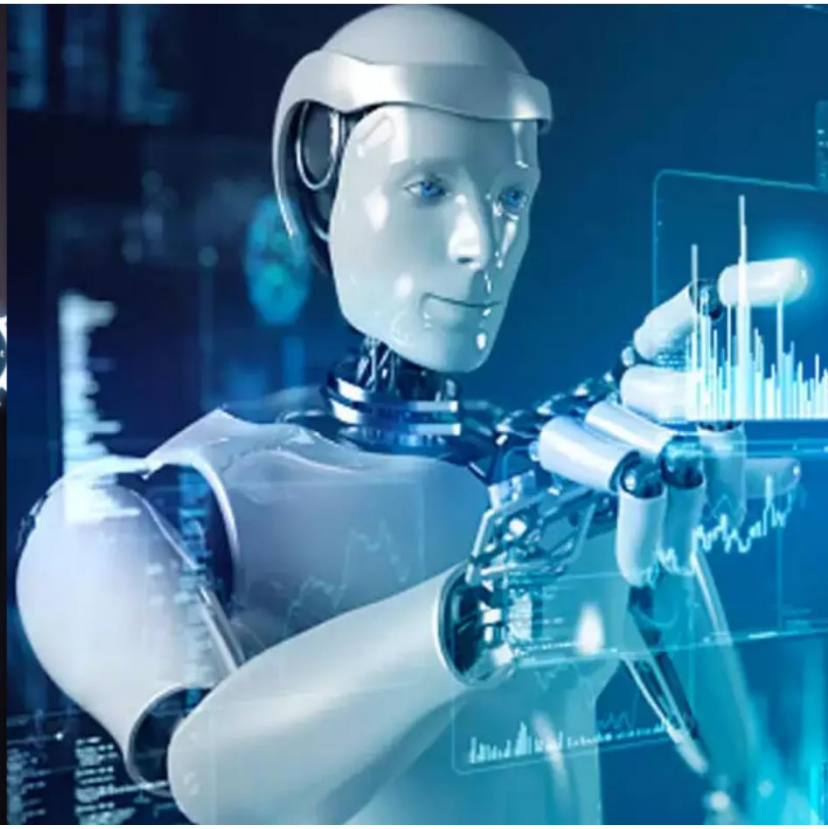
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

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



Impact Factor: 8.699

Volume 15, Issue 3, March 2026

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503252|

UPI Transaction Fraud Detection System using ML

Grandhi Anjana Priya[1], Bayapuri Srinivasa Prasanna Kumar[2], Bathini Venkata Rohit[3], Chukka Charan Bhuvanesh[4], M. Lakshmi Tirupatamma[5]

UG Students, Department of CSE (AI&ML), Kallam Haranadhareddy Institute of Technology, Guntur, India[1-4]

Assistant Professor, Department of CSE (AI&ML), Kallam Haranadhareddy Institute of Technology, Guntur, India[5]

ABSTRACT: With the rapid growth of digital payment systems, especially UPI transactions, the risk of financial fraud has also increased significantly. This project presents a UPI Fraud Detection System that uses machine learning techniques to identify fraudulent transactions. The system is built using the Random Forest algorithm, which analyzes transaction features such as transaction amount, time, authentication attempts, and user behavior to predict whether a transaction is fraudulent or legitimate.

The dataset is preprocessed and trained to achieve high accuracy in fraud detection. A user-friendly web application is developed using Flask, allowing users to input transaction details and receive real-time predictions. The proposed system aims to enhance digital payment security, reduce financial losses, and provide an efficient solution for fraud detection in online transactions.The system also incorporates data preprocessing techniques such as feature extraction and encoding to improve model performance. It handles both numerical and categorical data efficiently and ensures reliable predictions. The use of an ensemble learning approach helps in reducing overfitting and improving generalization on unseen data.

Furthermore, the application provides a probability-based output, indicating the percentage of fraud and legitimate transactions, which helps users better understand the prediction. The system is designed to be scalable and adaptable, allowing future integration with real-time banking systems and mobile applications.

I. INTRODUCTION

In recent years, digital payment systems have grown rapidly, with Unified Payments Interface (UPI) becoming one of the most widely used platforms for online transactions. While this advancement has made financial transactions faster and more convenient, it has also increased the risk of fraudulent activities. Detecting and preventing such frauds has become a critical challenge in ensuring the security of digital payments.

Traditional fraud detection systems are mainly rule-based and often fail to identify new or complex fraud patterns. These systems depend on predefined rules, which makes them less effective in handling evolving fraud techniques. To overcome these limitations, machine learning approaches are increasingly being used, as they can learn patterns from data and adapt to new types of fraud.

This project focuses on developing a UPI Fraud Detection System using machine learning techniques. The system uses the Random Forest algorithm, an ensemble learning method, to analyze transaction-related features such as transaction amount, time, authentication attempts, and user behavior. By training the model on historical transaction data, the system can accurately predict whether a transaction is fraudulent or legitimate.

Additionally, a user-friendly web application is developed using Flask to make the system interactive and accessible. Users can input transaction details and receive real-time predictions along with fraud probability. The main objective of this project is to enhance digital payment security, reduce financial losses, and demonstrate the practical application of machine learning in fraud detection systems.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503252|

II. LITERATURE SURVEY

Financial fraud detection has become a significant area of research due to the rapid growth of digital payment systems. Earlier approaches mainly relied on rule-based systems, where predefined rules were used to identify suspicious transactions. Although these systems were simple to implement, they lacked flexibility and failed to detect new or evolving fraud patterns, making them less effective in modern financial environments.

With the advancement of data analytics, researchers began using machine learning techniques for fraud detection. These methods allow systems to learn patterns from historical transaction data and make predictions on new data. Algorithms such as Logistic Regression, Decision Trees, and Support Vector Machines have been widely applied for classification tasks in fraud detection, showing improved accuracy compared to traditional methods.

Further research highlighted the effectiveness of ensemble learning methods, particularly the Random Forest algorithm. Studies have shown that Random Forest performs better in handling large datasets and multiple features, while also reducing overfitting. It combines the predictions of multiple decision trees to produce more reliable and accurate results, making it a popular choice in financial fraud detection systems.

In recent years, researchers have also explored deep learning techniques such as Neural Networks and Long Short-Term Memory (LSTM) models. These approaches are capable of capturing complex patterns and sequential behavior in transaction data. However, they require large amounts of data and computational resources, which may not always be feasible for all applications.
Despite these advancements, challenges such as imbalanced datasets**,** real-time detection requirements, and evolving fraud techniques still exist. Based on the findings from previous studies, this project implements a machine learning-based UPI Fraud Detection System using the Random Forest algorithm, which provides an efficient and accurate approach for detecting fraudulent transactions in digital payment systems.

III. METHODOLOGY

The methodology of the proposed UPI Fraud Detection System involves a systematic process starting from data collection to deployment of the machine learning model. Each step is designed to ensure accurate prediction of fraudulent transactions and smooth integration with the web application.

1. Data Collection:
The dataset used in this project contains UPI transaction records with features such as user ID, transaction amount, transaction hour, authentication attempts, PIN entry method, and the target variable is_fraud. The dataset includes both fraudulent and legitimate transactions, which helps the model learn patterns effectively.

2. Data Preprocessing:
Before training the model, the dataset undergoes preprocessing:

- The timestamp is converted into useful features like transaction hour.
- A new feature is_night is created to identify transactions occurring at unusual hours.
- Unnecessary or leakage columns are removed to avoid biased predictions.
- Missing values, if any, are handled properly.

This step ensures that the data is clean and suitable for model training.

3. Feature Engineering:
Relevant features are selected to improve model performance:

- Transaction amount
- Transaction time (hour, night indicator)
- Authentication attempts
- PIN entry method
- User-related information

These features help in identifying abnormal transaction behavior.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503252|

4. Data Splitting:
The dataset is divided into:
- **Training set (80%)** – used to train the model
- **Testing set (20%)** – used to evaluate model performance

This helps in assessing how well the model performs on unseen data.

5. Model Training:
A Random Forest Classifier is used to train the model. It builds multiple decision trees and combines their predictions using majority voting. This improves accuracy and reduces overfitting.

6. Model Evaluation:
The trained model is evaluated using:
- Accuracy
- Classification Report (Precision, Recall, F1-score)
- Confusion Matrix

These metrics help in understanding how well the model detects fraudulent and legitimate transactions.

7. Model Saving:
The trained model is saved using the Pickle library as:
- fraud_model.pkl (model file)
- encoders.pkl (for categorical encoding)

This allows reuse of the model without retraining.

8. Web Application Development:
A web application is developed using Flask:
- Users enter transaction details through a web form
- The input data is processed and sent to the trained model
- The model predicts fraud probability
- The result is displayed on a separate result page

The proposed methodology ensures a complete pipeline from data preprocessing to deployment. By combining machine learning with a web interface, the system provides an efficient and user-friendly solution for detecting UPI fraud transactions.

IV. EXPERIMENTAL RESULTS

The proposed UPI Fraud Detection System was implemented using a Random Forest machine learning model and evaluated on the given dataset. The model was trained using 80% of the data and tested on the remaining 20% to assess its performance on unseen transactions. The results indicate that the model performs effectively in classifying transactions as either fraudulent or legitimate.

To evaluate the model, a confusion matrix was used, which provides a clear representation of correct and incorrect predictions. The matrix shows that the model correctly identifies the majority of legitimate and fraudulent transactions, with very few misclassifications. The high number of true positives and true negatives, along with minimal false positives and false negatives, indicates strong predictive performance.

The experimental results show that the model achieves high accuracy with minimal misclassification, indicating its effectiveness in detecting fraudulent transactions. The probability-based output further enhances the system by providing confidence levels for each prediction, allowing users to better understand the results.

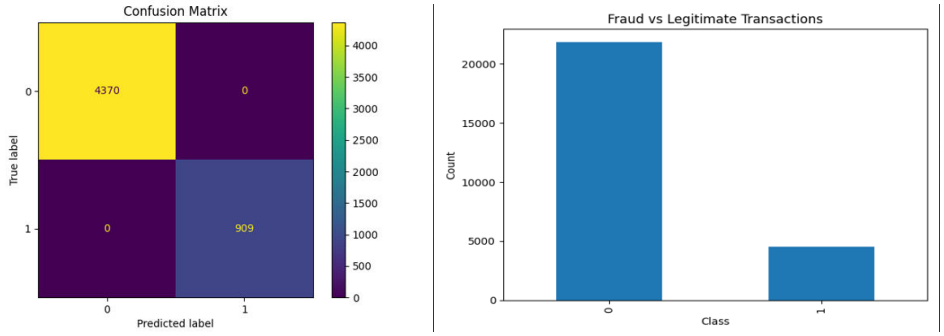
International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503252|



Web Application Results:

Furthermore, the developed Flask web application successfully integrates the trained model and allows users to input transaction details. The system processes the input data and displays the prediction results in a user-friendly interface, including visual indicators and probability scores. This demonstrates the practical applicability of the model in real-time scenarios.

Overall, the results confirm that the proposed system is accurate, efficient, and reliable in detecting fraudulent transactions. The combination of machine learning techniques and a web-based interface makes the system suitable for enhancing security in digital payment platforms.

Fig1: Home Page

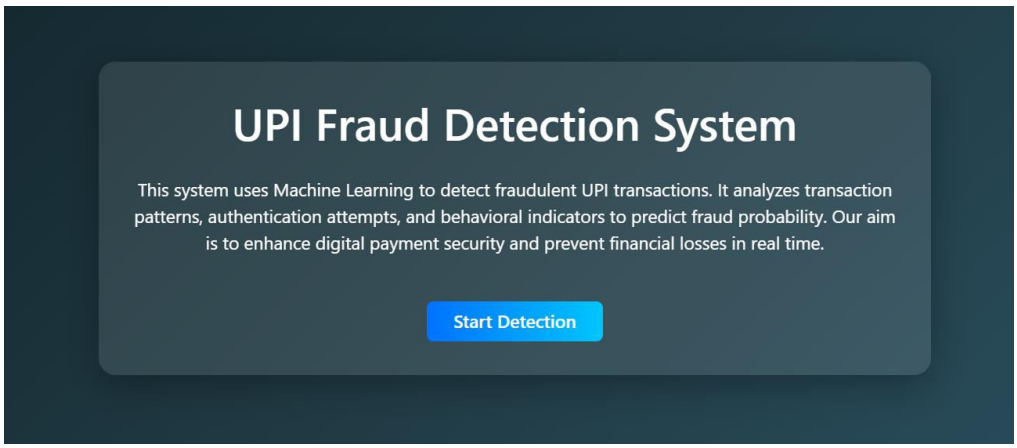
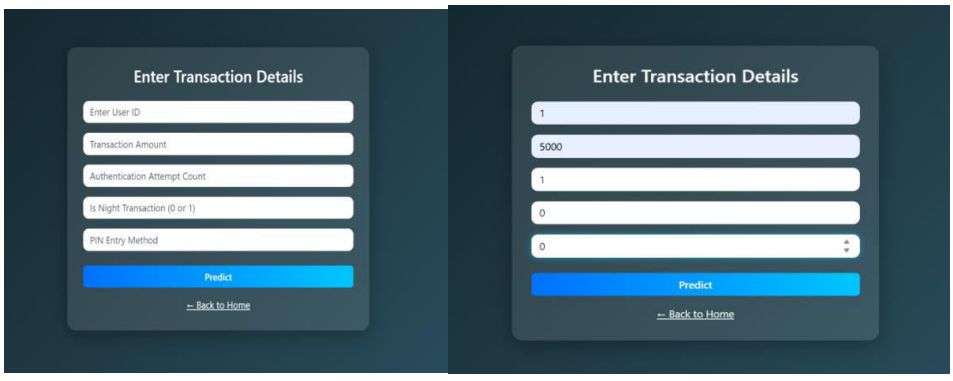


Fig 2: Index Page



International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

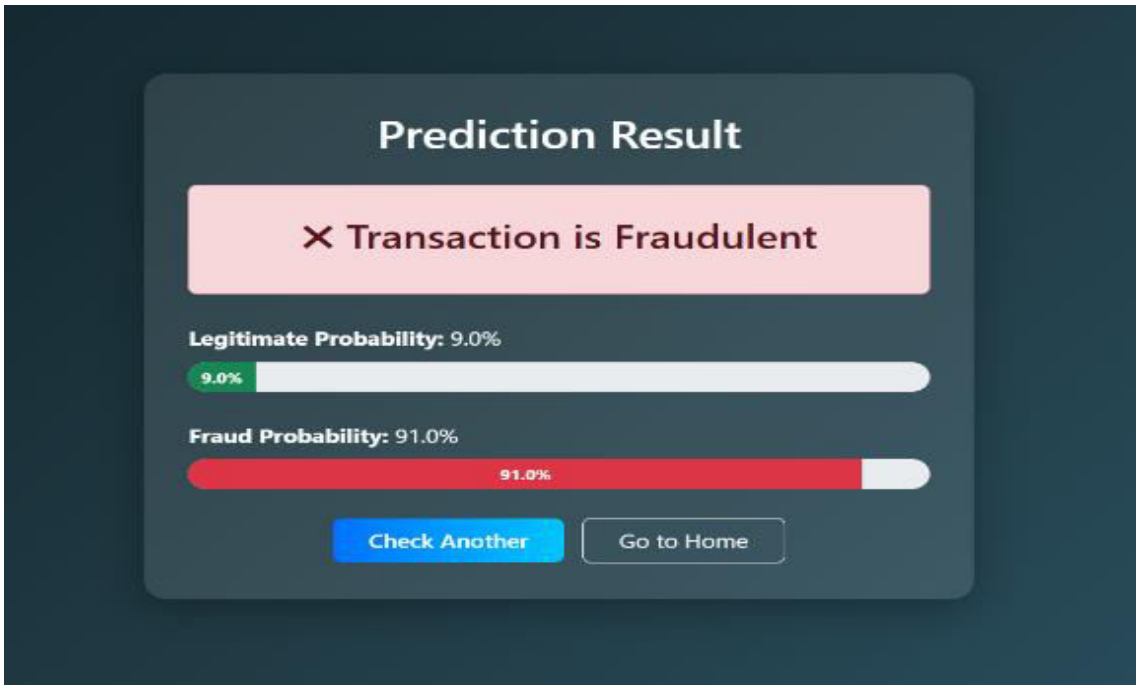
Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503252|

Fig 3: Result Page



V. ADVANTAGES AND LIMITATIONS

Advantages:

The proposed UPI Fraud Detection System offers several benefits in improving the security of digital transactions. Firstly, the use of the Random Forest algorithm provides high accuracy and reliable predictions, making the system effective in identifying fraudulent transactions. The ensemble nature of the algorithm helps in reducing overfitting and improves generalization on unseen data.

Secondly, the system is capable of handling both numerical and categorical data through appropriate preprocessing and encoding techniques. This allows it to analyze various transaction-related features such as amount, time, and authentication attempts, leading to better decision-making.

Another important advantage is the development of a user-friendly web application using Flask, which enables real-time prediction. Users can easily input transaction details and receive immediate results along with probability scores, enhancing usability and accessibility.

Additionally, the system provides probability-based outputs, which give a clear understanding of the confidence level of predictions. This helps users and organizations make informed decisions regarding transaction approval or rejection.

Finally, the system is scalable and flexible, allowing future improvements such as integration with real-time banking systems, mobile applications, and cloud platforms.

Limitations:

Despite its advantages, the proposed system has certain limitations. One of the main limitations is that the model is trained on a simulated or limited dataset, which may not fully represent real-world transaction patterns. This can affect the model's performance when applied to real-time data.

Another limitation is the issue of imbalanced data, where the number of legitimate transactions is significantly higher than fraudulent ones. This may lead to biased predictions if not handled properly.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503252|

The system currently operates in an offline environment and is not integrated with actual banking or UPI platforms. Therefore, it does not support real-time transaction monitoring in a live environment.

The use of Label Encoding for categorical features may introduce unintended relationships between values, which can slightly affect the model's performance. Furthermore, the system lacks explainability, meaning it does not provide clear reasons for why a transaction is classified as fraudulent or legitimate, which is important in financial applications.

Additionally, while Random Forest provides good accuracy, it may not capture **very** complex or evolving fraud patterns compared to advanced deep learning models.

Lastly, the system depends on the selected features, and limited feature engineering may reduce the model's ability to detect sophisticated fraud activities.

VI. FUTURE SCOPE

The proposed UPI Fraud Detection System provides a strong foundation for identifying fraudulent transactions using machine learning. However, there are several areas where the system can be further enhanced to improve its performance and real-world applicability.

In the future, the system can be extended to support real-time fraud detection by integrating it with live UPI and banking transaction systems through secure APIs. This would allow transactions to be analyzed instantly before completion, helping to prevent fraud proactively.

The implementation of advanced machine learning and deep learning algorithms, such as Neural Networks and LSTM models, can further improve the system's ability to detect complex and evolving fraud patterns. These models can capture deeper relationships within transaction data and enhance prediction accuracy.

The system can also be improved by incorporating additional features such as user location, device information, IP address, and transaction history. This enhanced feature set would provide better insights into user behavior and improve fraud detection performance.

Another important future enhancement is the development of a mobile application interface, allowing users to access the system easily and receive real-time alerts for suspicious transactions. This would increase usability and accessibility for end users.

Furthermore, deploying the system on cloud platforms would enable it to handle large-scale transaction data efficiently and provide better scalability and performance. Continuous model training using updated transaction data can also be implemented to ensure that the system adapts to new fraud techniques over time.

VII. CONCLUSION

The proposed UPI Fraud Detection System successfully demonstrates the use of machine learning techniques to identify fraudulent transactions in digital payment systems. By utilizing the Random Forest algorithm, the system effectively analyzes transaction features such as transaction amount, time, authentication attempts, and user-related information to classify transactions as fraudulent or legitimate.

The model achieved high accuracy with minimal misclassification, as observed from the evaluation metrics and confusion matrix. This indicates that the system is reliable and capable of detecting fraud patterns efficiently. The use of proper data preprocessing and feature engineering further contributed to improving the model's performance.

In addition to the machine learning model, a Flask-based web application was developed to provide a user-friendly interface. This allows users to input transaction details and receive real-time predictions along with probability scores, making the system practical and accessible.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503252|

Overall, the project highlights the importance and effectiveness of machine learning in enhancing digital payment security. Although the system has certain limitations, it provides a strong foundation for future improvements and real-world implementation in fraud detection systems.

REFERENCES

[1] Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. C. (2011). *Data mining for credit card fraud: A comparative study.* Decision Support Systems, 50(3), 602–613.
[2] Phua, C., Lee, V., Smith, K., & Gayler, R. (2010) . *A comprehensive survey of data mining-based fraud detection research.* arXiv preprint arXiv:1009.6119.
[3] Dal Pozzolo, A., Boracchi, G., Caelen, O., Alippi, C., & Bontempi, G. (2015). *Credit card fraud detection: A realistic modeling and a novel learning strategy.* IEEE Transactions on Neural Networks and Learning Systems.
[4] Breiman, L. (2001). *Random Forests.* Machine Learning Journal, 45(1), 5–32.
[5] Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques.* Morgan Kaufmann.
[6] Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow.* O'Reilly Media.
[7] Pedregosa, F., et al. (2011). *Scikit-learn: Machine Learning in Python.* Journal of Machine Learning Research, 12, 2825–2830.
[8] Grinberg, M. (2018) *Flask Web Development: Developing Web Applications with Python.* O'Reilly Media.
[9] Chollet, F. (2017). *Deep Learning with Python.* Manning Publications.
[10] Kaggle Dataset, "**UPI Fraud Detection Dataset" :**
https://www.kaggle.com/datasets/omsshete/upi-fraud-detection-dataset

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details