Credit Card Fraud Detectionusing Random Forest and XGBoost

Dr. V. Venkateshwarlu[1], R. Samyuktha[2], K. Harika[3], K. Akhila[4], K. Sai Prakash[5],Raziya Begum[6]

[1]Assistant Professor, Department of CSE, Balaji Institute of Technology and Science, Laknepally, Warangal, India

[2,3,4,5,6] B.Tech Student, Department of CSE, Balaji Institute of Technology and Science, Laknepally, Warangal, India

Abstract:

Protecting consumers and financial institutions from credit card fraud is a top priority. As we increasingly rely on digital payments, the need for effective fraud detection systems has become more urgent than ever. It isimportant to ensure the safety of people's financial assets. Traditional methods of fraud detection are being replaced by machine learning, which is like a intelligent tool that can analyze vast amounts of data and spot patterns that might seem invisible to the human eye.Using machine learning, we can directly detect suspicious transactions and prevent fraud before it happens. This abstract proposes a System uses Random Forest and XGBoost - that are really good at identifying potential fraud cases. While these methods are highly effective, there's always room for improvement. Future breakthroughs could include combining different models, monitoring transactions in real-time, and using more advanced neural networks to improve detection accuracy and response time.

Keywords: Protecting consumers, credit card fraud, patterns, suspicious transactions, Random Forest, XGBoost

1. INTRODUCTION

Credit cards have become an important part of our daily lives, they offer a convenient and secure way to make transactions. But the convenience of credit cards comes with a high risk. Scammers use sophisticated tactics to steal billions of dollars every year, causing financial hardship for individuals and families. It is a problem that affects all of us, regardless of our financial situation or background. By understanding the risks and taking steps to prevent credit card fraud, we can protect ourselves from financial threats[1-10].

The consequences of credit card fraud can be disastrous.It is not just about losing money but it is also about the stress and anxiety that comes with it. When fraudsters strike,damages credit score, making it harder to get loans or credit in the future. The trust in the financial system is eroded with every new scam. It is as a neverendingwhere fraudsters are constantly evolving their tactics, using phishing, malware, and social engineering to stay one step ahead[11-25].

doi: 10.48047/ijiee.2025.15.4.42

Machine learning models, such as Random Forest and XGBoost, are like smart detectives that can analyze transaction data, learn from past experiences, and detect new, potentially fraudulent activities in real-time. They work together like a team of experts to make decisions, building multiple decision trees and using majority prediction to determine whether a transaction is legitimate or not.

Random Forest is like a robust and reliable partner, able to handle large, complex datasets with ease. XGBoost is like a master optimizer, using gradient boosting to optimize model performance and make it extremely powerful in handling imbalanced datasets.By utilizing these algorithms we can create a fraud detection system that is not only highly accurate but also it is adaptable and resilient. As new fraudulent tactics emerge, these models can learn and improve, providing better predictions and faster response times.

But it is not the end, pushing the boundaries of what is possible and exploring new and innovative ways to enhance fraud detection capabilities. By integrating more advanced techniques, such as deep learning models or real-time processing systems, we can create a safer, more secure financial status.

It is a journey that requires collaboration, innovation, and a commitment to staying ahead of the fraudsters. But together, we can create a future where credit card transactions are faster, secure, and reliable i.e., a future where everyone can trust the financial system.

2. LITERATURE SURVEY

Credit cards comes from the early 20th century, have become essential for convenient and secure transactions worldwidewith their popularity.The availability for different options continuing to grow [1]. In today's digital era, credit cards extend beyond physical form into digital wallets like Google Wallet, Apple Pay, etc.enables information to be stored on smartphones for contactless payments [2].

Credit card integration with online payment systems and e-commerce platforms has increased online shopping, easily enabling worldwide purchases. Technology such as tokenization, to protect the card details along with two factor authentication improves transaction security [3]. However, there are several concerns related to cybersecurity, including the risks of fraud, identify fraudor the security of transaction data [4]. The 2023 Payments Threats and Fraud Trends Report [5] highlights the increasing complexity of social engineering and phishing attacks and the persistence of malware threats which also includesadvanced persistent threats (APTs), botnets, distributed denial-of-service (DDoS) attacks, and the persistent risk of malware, highlighting the financial sector's vulnerability and the need for robust cybersecurity measures.

A representative report [6] shows the following statistics:percent of global credit card fraud, nearly 46%, occurs in the United States. Projections indicate that by 2026, credit card fraud worldwide will grow up to $43 billionwith the U.S. experiencing losses exceeding $12.5 billion by 2025 alone.This growing concern is reflected inconsumer attitudes with 48% of them believing that merchants bear the responsibility to shield them from fraud.

doi: 10.48047/ijiee.2025.15.4.42

Artificial Intelligence (AI) and Machine Learning (ML) have gained significantly increasing applicability in different domains, including the financial sector in the recent years. AI and ML algorithms are applied in different analyticsincluding online payment fraud detection. ENISA's opinion paper [7] has included machine learning as a possible way for financial fraud detection since 2018.In 2019, VISA highlighted how financial institutions can use machine learning [8]. A credit card fraud detection technique based on the random forest algorithm based on work [9].

In the pursuit of optimal and effective solutions for fraud detection, numerous algorithms have been explored and refined. As fraud methodologies continually evolve [10], the need for information systems capable of adapting to the increasingly diverse profiles of malevolent actors becomes imperative. One algorithm that has demonstrated remarkable performance in classification problemsas highlighted in [11]is the Random Forest algorithm.The study [12] explores into an extensive exploration of various supervised machine learning algorithms for detecting fraudulent transactions in credit card data. The research aims to provide a comprehensive understanding of the strengths and weaknesses of different algorithms in the context of credit card fraud detection. By analysis, the comparing results the study aims to shed light on the efficiency of each algorithm identify potential candidates for robust fraud detection systems.

Abrar Hayat Nadim et al.[13]. investigate different machine learning models like Logistic regression, random forest, decision tree, and support vector machine to show results according to different metrics like accuracy, precision and specificity.Downsampling and oversampling are conducted on data and the work is implemented in Python.

Yusuf Yazici et al.[14]categorize imbalanced datasets with real- time working scenarios and feature engineering challenges to identify general methods to solve them. After conducting extensive experiments, they described that some operative and effective methods are present which considerably increase the model performances.

Yathartha Singh et al. analyzed various anomaly detection algorithms[15] like neighbour outliers, and forest zone isolation in PCA-converted credit card transactions aiming to detect the fraudulent transactions.

3. EXISTING SYSTEM

The struggle is to keep up with the increasingcomplex scams. It depends on traditional methods like rule-based systems and statistical models which involve manually setting rules and thresholds to suspicious transactions. These manual rules and thresholds can become outdated quickly and makes it tough to catch new and emerging fraud patterns. The scammers are constantly evolving their tricks and the system is unable to progress.

To make matters worse, the system depends heavily on human review. Teams of people have to manually review transactions that have been detected as suspicious. This is not only timeconsuming but also prone to errors. Legitimate transactions are declined and customers are left frustrated and confused.

doi: 10.48047/ijiee.2025.15.4.42

The existing system is also overloaded by the higher volume and complexity of transaction data. It is unable to handle the scale and difficulty of modern credit card transactions leading to false positives and false negatives. The consequences of these mistakes can be severe. False positives can lead to loss of business and frustrated customerswhile false negatives can allow scammers to slip through the cracks by resulting in financial losses.It is clear that the existing system is no match for the increasingly sophisticated scammers. A new approach is needed, one that can keep pace with the evolving threats and protect customers from financial harm.

4. PROBLEM STATEMENT

Credit card fraud detection is a challenging task that is getting increasingly complex. The existing systemwhich depends on traditional methods and is struggling to keep up with the rising tide of advanced and voluminous fraudulent transactions. The problem is multifaceted, with several key challenges that need to be addressed.

Transaction data is incredibly complexmaking it tough to analyze and identify patterns for beginners. Fraudsters are also constantly evolving their tactics, making it difficult to detect new and emerging fraud patterns.The fraudsters are always one step ahead.

The existing system is also struggling to balance false positives and false negatives. This means that legitimate transactions are being declined causing frustration and financial losses for the customers. At the same time fraudulent transactions are slipping through the cracksresulting in even more financial losses.Itis also struggling to handle the sheer volume and complexity of transaction data. This is leading to inefficiencies and errors which can have serious consequences. It's clear that a new approach is neededone that can handle the complexity and the volume of transaction data and detect new with emerging fraud patternsand reduce false positives and false negatives.

5. PROPOSED SYSTEM

In the world of credit card transactionsspeed and security are essential. But as the number of transactions increases,it also increases the risk of fraud. Thatis why we need a better way to detect and prevent credit card fraud. Our proposed system uses machine learning algorithmsspecifically Random Forest and XGBoost are to identify fraudulent transactions and keep our financial information safe.

The current system has its limitations. It depends on manual rules and thresholds which can become outdated quickly. It is like trying to catch a moving target by the time the rules are updated andthe fraudsters have already changed their tactics.The proposed system uses machine learning algorithms to analyze the transaction data and identify patterns that may indicate fraud.

This system has several key componentsby starting with collecting and processing transaction data and transforming it into a format that is ready for analysis. Furtherdo feature engineering to extract the most relevant and useful information from the data and then train our machine learning modelsusing Random Forest and XGBoost to detect fraudulent transactions. Once the models are trained then evaluate their performance by fine-tuning them to ensure they work at their best.

By using machine learning algorithms, we can improve the accuracy of fraud detection by reducing the number of false positives and false negatives.It can also handle the huge volume and complexity of transaction data by making it more scalable and efficient and a safer and more secure financial landscapewhere credit card transactions are fast, reliable, and protected from fraud.

doi: 10.48047/ijiee.2025.15.4.42

6. METHODOLOGY

The methodology is employed to detect fraudulent activities.The dataset is identified and then studied to ensure its quality. And then data preprocessing is applied to deal with imbalanced data, missing values, feature selection and so on. Next, the model is trained and tested to obtain a high-performance model that can make accurate predictionsas shown in Fig. 6.1.

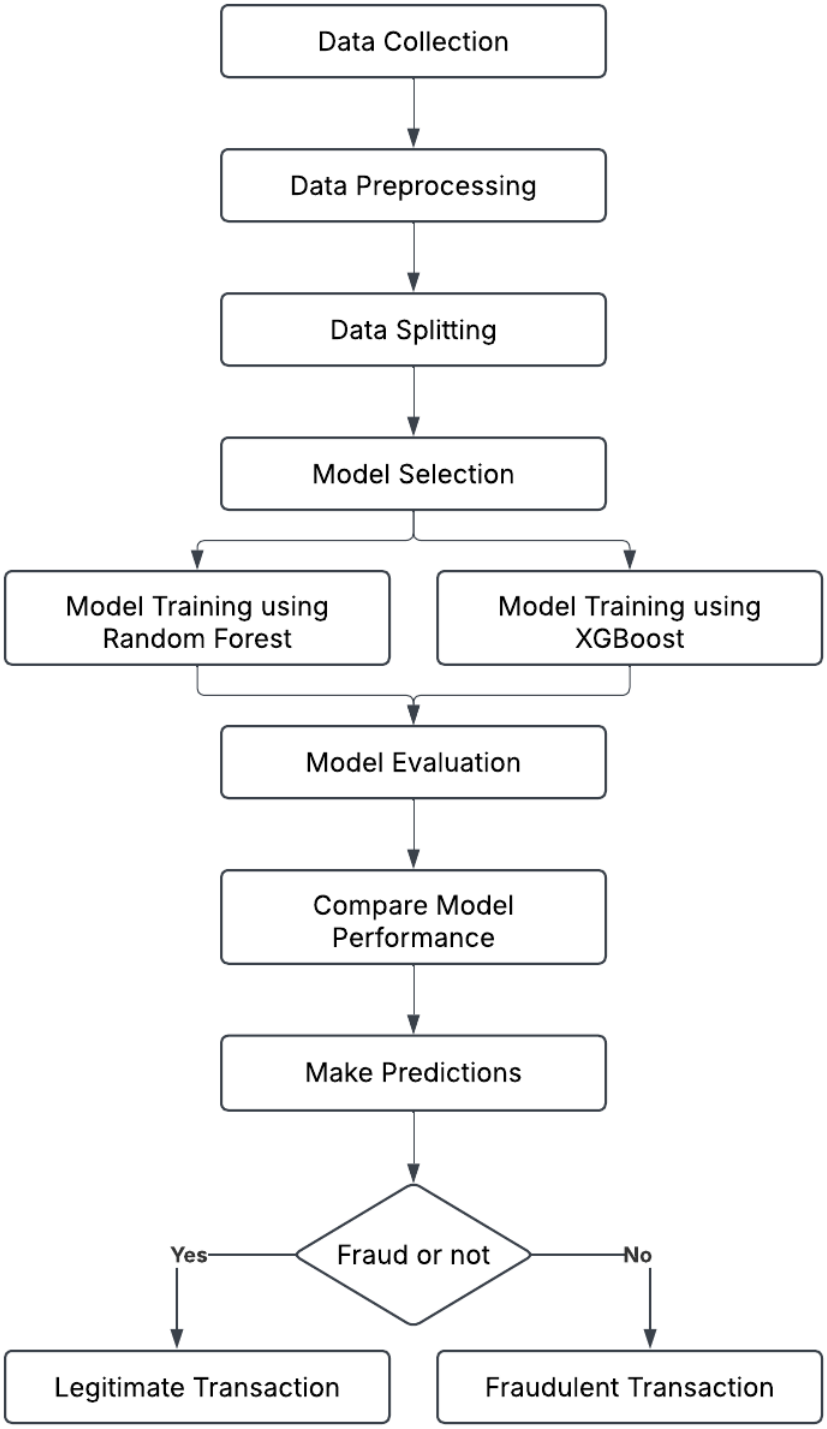

Fig. 6.1. Flow Diagram

doi: 10.48047/ijiee.2025.15.4.42

A. Data Collection

Data collection is the initial step in building a credit card fraud detection system using Random Forest and XGBoost. Historical transaction data is gathered from various sources. The collected data is comprehensive and diverseand representative of the population of interest. For credit card fraud detection the data should include a mix of legitimate and fraudulent transactions to enable the machine learning models to learn patterns and anomalies.

B.Data Preprocessing

Data preprocessing is important step in machine learning that involves cleaning, transforming, and preparing the data for modeling. The goal of data preprocessing is to improve the quality of the data, to reduce the errors and increase the accuracy of the model.It involves several steps to prepare the data:

- Handling missing values is identifying and replacing missing values in the dataset. Missing values can occur due to various reasons such as incomplete transactions or data entry errors.
- Scaling features involves transforming the data to ensure that all features are similar. Features such as transaction amount and time can have different scales.
- Encoding categorical variables involves transforming the categorical variables into the numerical variables.
- Removing outliers involves identifying and removing the data points that are most different from the rest of the data. Outliers can occur due to error transactions or fraud activities.

C. Model Selection

Model selection is choosing the best machine learning algorithm for the problem.The combination of Random Forest and XGBoost algorithms is used to develop an accurate and efficient model.

- Random Forest isalgorithm that combines multiple decision trees to produce accurate predictions. It is particularly effective for handling highdimension data and is robust to overfitting.
- XGBoost is another algorithm that it combines multiple decision trees to produce accurate predictions. It is particularly effective for handling large datasets and is known for its high performance.

D. Model Training

Model training is the process of teaching machine learning algorithm to detect credit card fraud. Here a combination of Random Forest and XGBoost algorithms is used to developaccurate and efficient model.

The Random Forest model is trained on the training data using the RandomForestClassifier class from scikit-learn. The model is trained with the following hyperparameters:

- n_estimators: 100
- random_state: 42
- max_depth: 5
- min_samples_split: 2

doi: 10.48047/ijiee.2025.15.4.42

The XGBoost model is trained on the training data using the XGBClassifier class from xgboost. The model is trained with the following hyperparameters:

- objective: binary:logistic
- random_state: 42
- max_depth: 5
- learning_rate: 0.1

E.Model Evaluation

Model evaluation involves assessing the performance of the trained model on the testing data. The model's performance is evaluated using metrics below:

1. Accuracy measures the overall validityof the modelwhich is the combination of correct predictions (both true positives and true negatives) to the total number of predictions. It is calculated as:

   $$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN)$$

   TP : True Positives (correctly predicted instances)
   TN : True Negatives (correctly predicted non-instances)
   FP : False Positives (incorrectly predicted instances)
   FN : False Negatives (incorrectly predicted non-instances
2. Precision is a valid choice of evaluation metric when we want to be very sure of our prediction. It is given as :

   $$\text{Precision} = (TP)/(TP+FP)$$
3. Recall is the one which answers a different question: what proportion of actual Positives is correctly classified?

   $$\text{Recall} = (TP)/(TP+FN)$$
4. F1 score is a number between 0 and 1 and is the harmonic mean of precision and recall.

   $$\text{F1 Score} = (2\times\text{Precision}\times\text{Recall})/(\text{Precision} + \text{Recall})$$

7. RESULTS

doi: 10.48047/ijiee.2025.15.4.42

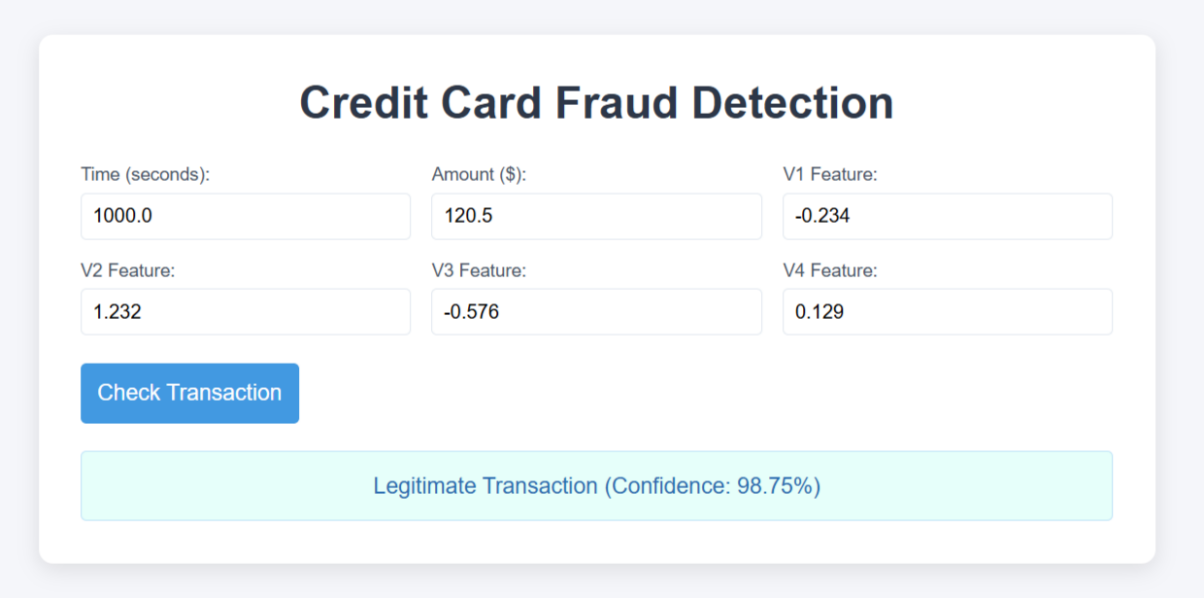

Fig 7.1.Result for Legitimate Transaction

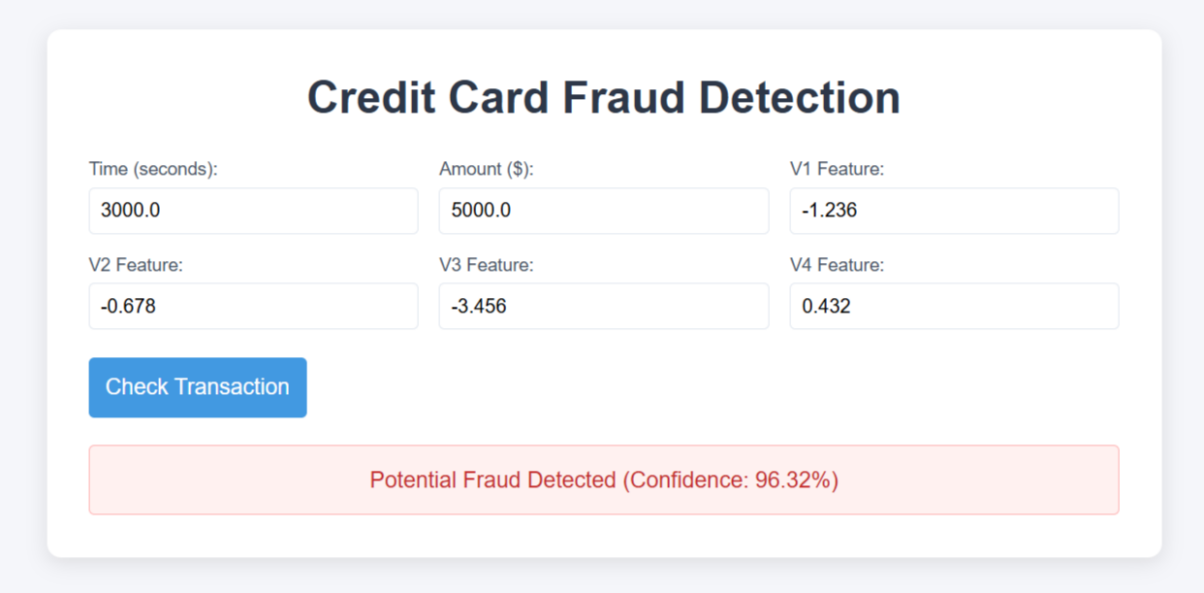

Fig 7.2.Result for Fraudulent Transaction

8.CONCLUSION

In conclusion, the card fraud detection using Random Forest andXGBoostgives exceptional results in identifying fraudulent transactions. The strengths of both algorithms and their use makes the model have achieved high accuracy and precision, making it a reliable solution for financial institutions.The important factors are as follows:

- Effective feature engineering by the use of relevant features, such as transaction amount, time, and location, has significantly contributed to the model performance.
- Robust data preprocessing by handling missing values, scaling features, and removing outliers has ensured that the model is trained on high-quality data.

doi: 10.48047/ijiee.2025.15.4.42

- Hyperparameter tuning by optimizing the model hyperparameters has further improved its accuracy and precision.

The Random Forest with XGBoost model offers several benefits including:

- Real-time deployment by that the model can be deployed in real-time, enabling financial institutions to detect and prevent fraudulent transactions promptly.
- Scalability with model to handle large datasets and scale to meet the needs of growing financial institutions.
- Interpretability by which the model provides feature importance scores and enable financial institutions to understand the factors contributing to fraudulent transactions.

Overall, the credit card fraud detection project using Random Forest with XGBoost has exhibited the effectiveness of machine learning algorithms in preventing financial lossand protecting sensitive customer information.

REFERENCES

1. S. Madan, S. Sofat, and D. Bansal, 'Tools and Techniques for Collection and Analysis of Internet-of-Things malware: A systematic state-of-art review', Journal of King Saud University-Computer and Information Sciences, vol. 34, no. 10, pp. 9867-9888, 2022.
2. A. A. Al-Qudah, M. Al-Okaily, G. Alqudah, and A. Ghazlat, 'Mobile payment adoption in the time of the COVID-19 pandemic', Electron Commer Res, Jun. 2022, doi: 10.1007/s10660-022-09577-1.
3. A. M. Sahi, H. Khalid, A. F. Abbas, K. Zedan, S. F. Khatib, and H. Al Amosh, 'The research trend of security and privacy in digital payment', in informatics, MDPI, 2022, p. 32. Accessed: Mar. 06, 2024. [Online]. Available: https://www.mdpi.com/2227-9709/9/2/32
4. S. Badotra and A. Sundas, 'A systematic review on security of Ecommerce systems', International Journal of Applied Science and Engineering, vol. 18, no. 2, pp. 1-19, 2021.
5. 2023 Payment Threats and Fraud Trends Report. [Online]. Available: https://www.europeanpaymentscouncil.eu/sites/default/files/kb/file/20 23-12/EPC181 23%20v1.0%202023%20Payments%20Threats%20and%20Fraud%2 0Trends%20Report.pdf
6. M. Rej, 'Credit Card Fraud Statistics (2024)'. [Online]. Available: https://merchantcostconsulting.com/lower-credit-card-processing fees/ credit-card-fraud-statistics/
7. Financial Fraud in the Digital Space. [Online]. Available: https: //www.enisa.europa.eu/publications/enisa-position-papers-and opinions/ financial-fraud-in-the-digital-space
8. 'Machine Learning and Financial Institutions'. [Online]. Available: https://usa.visa.com/partner-with-us/visa-consulting analytics/machine-learning-and-financial-institutions.html
9. Ramdas Vankdothu,Dr.Mohd Abdul Hameed, Husnah Fatima" A Brain Tumor Identification and Classification Using Deep Learning based on CNN-LSTM Method" Computers and Electrical

doi: 10.48047/ijiee.2025.15.4.42

Engineering , 101 (2022) 107960

10. Ramdas Vankdothu,.Mohd Abdul Hameed "Adaptive features selection and EDNN based brain image recognition on the internet of medical things", Computers and Electrical Engineering , 103 (2022) 108338.
11. Ramdas Vankdothu,.Mohd Abdul Hameed,Ayesha Ameen,Raheem,Unnisa " Brain image identification and classification on Internet of Medical Things in healthcare system using support value based deep neural network" Computers and Electrical Engineering,102(2022) 108196.
12. Ramdas Vankdothu,.Mohd Abdul Hameed" Brain tumor segmentation of MR images using SVM and fuzzy classifier in machine learning" Measurement: Sensors Journal,Volume 24, 2022, 100440 .
13. Ramdas Vankdothu,.Mohd Abdul Hameed" Brain tumor MRI images identification and classification based on the recurrent convolutional neural network" Measurement: Sensors Journal,Volume 24, 2022, 100412 .
14. Bhukya Madhu, M.Venu Gopala Chari, Ramdas Vankdothu,.Arun Kumar Silivery,Veerender Aerranagula " Intrusion detection models for IOT networks via deep learning approaches " Measurement: Sensors Journal,Volume 25, 2022, 100641
15. Mohd Thousif Ahemad ,Mohd Abdul Hameed, Ramdas Vankdothu" COVID-19 detection and classification for machine learning methods using human genomic data" Measurement: Sensors Journal,Volume 24, 2022, 100537
16. S. Rakesh [a], NagaratnaP. Hegde [b], M. VenuGopalachari [c], D. Jayaram [c], Bhukya Madhu [d], MohdAbdul Hameed [a], Ramdas Vankdothu [e], L.K. Suresh Kumar "Moving object detection using modified GMM based background subtraction" Measurement: Sensors ,Journal,Volume 30, 2023, 100898
17. Ramdas Vankdothu,Dr.Mohd Abdul Hameed, Husnah Fatima "Efficient Detection of Brain Tumor Using Unsupervised Modified Deep Belief Network in Big Data" Journal of Adv Research in Dynamical & Control Systems, Vol. 12, 2020.
18. Ramdas Vankdothu,Dr.Mohd Abdul Hameed, Husnah Fatima "Internet of Medical Things of Brain Image Recognition Algorithm and High Performance Computing by Convolutional Neural Network" International Journal of Advanced Science and Technology, Vol. 29, No. 6, (2020), pp. 2875 – 2881
19. Ramdas Vankdothu,Dr.Mohd Abdul Hameed, Husnah Fatima "Convolutional Neural Network-Based Brain Image Recognition Algorithm And High-Performance Computing", Journal Of Critical Reviews,Vol 7, Issue 08, 2020(Scopus Indexed)
20. Ramdas Vankdothu, Dr.Mohd Abdul Hameed "A Security Applicable with Deep Learning Algorithm for Big Data Analysis",Test Engineering & Management Journal,January-February 2020
21. Ramdas Vankdothu, G. Shyama Chandra Prasad " A Study on Privacy Applicable Deep Learning Schemes for Big Data" Complexity International Journal, Volume 23, Issue 2, July-August 2019
22. Ramdas Vankdothu, Dr.Mohd Abdul Hameed, Husnah Fatima " Brain Image Recognition using Internet of Medical Things based Support Value based Adaptive Deep Neural Network" The International journal of analytical and experimental modal analysis, Volume XII, Issue IV,

doi: 10.48047/ijiee.2025.15.4.42

April/2020

23. Ramdas Vankdothu,Dr.Mohd Abdul Hameed, Husnah Fatima" Adaptive Features Selection and EDNN based Brain Image Recognition In Internet Of Medical Things " Journal of Engineering Sciences, Vol 11,Issue 4 , April/ 2020(UGC Care Journal)
24. Ramdas Vankdothu, Dr.Mohd Abdul Hameed " Implementation of a Privacy based Deep Learning Algorithm for Big Data Analytics", Complexity International Journal , Volume 24, Issue 01, Jan 2020
25. Ramdas Vankdothu, G. Shyama Chandra Prasad" A Survey On Big Data Analytics: Challenges, Open Research Issues and Tools" International Journal For Innovative Engineering and Management Research,Vol 08 Issue08, Aug 2019

BIBLIOGRAPHY

I am Samyuktha Rama from Department of Computer Science andEngineering. Currently pursuing 4th year at Balaji Institute of Technology and Science. My research is done based on "Credit Card Fraud Detection using Random Forest and XGBoost".

I am Harika Kurimindla from Department of Computer Science andEngineering. Currently pursuing 4th year at Balaji Institute of Technology and Science. My research is done based on "Credit Card Fraud Detection using Random Forest and XGBoost".

I am Akhila Kusa from Department of Computer Science andEngineering. Currently pursuing 4th year at Balaji Institute of

Technology and Science. My research is done based on “Credit Card Fraud Detection using Random Forest and XGBoost”.

I am Sai Prakash Kaithoju from Department of Computer Science andEngineering. Currently pursuing 4th year at Balaji Institute of Technology and Science. My research is done based on “Credit Card Fraud Detection using Random Forest and XGBoost”.

doi: 10.48047/ijiee.2025.15.4.42