

# A Critical Analysis of Machine Learning Techniques for Online Transaction Security

Guide: Shraddha Kalsekar<sup>1</sup>,

Nalani Jagtap<sup>2</sup>, Abhinav Tuplondhe<sup>3</sup>, Nikhil Ghugare<sup>4</sup>, Vaishnavi Thorat<sup>5</sup>, Shreya Belkar<sup>6</sup>

## Abstract:

The widespread adoption of Unified Payments Interface (UPI) as the preferred digital transaction method has led to a sharp increase in fraudulent activities, presenting a considerable challenge to the security and reliability of online payments. To address this pressing concern, our project focuses on developing a sophisticated fraud detection system tailored specifically for UPI transactions. We utilize Convolutional Neural Networks (CNNs), an advanced deep learning architecture, to intricately capture the sequence of operations involved in UPI transaction processing. Through extensive data collection and preprocessing, we curate a diverse dataset comprising various transaction types, amounts, and temporal patterns, ensuring comprehensive coverage of transaction behaviour. The CNN is meticulously trained on this dataset, leveraging its innate ability to discern subtle patterns and dependencies within transaction sequences. Notably, our approach emphasizes the importance of minimizing false positives to prevent genuine transactions from being erroneously flagged as fraudulent. In real-time, incoming UPI transactions are subjected to scrutiny by the trained CNN, which evaluates their conformity to learned normal behaviour patterns. Transactions exhibiting deviations indicative of potential fraud are promptly flagged for further investigation, thus enabling proactive mitigation of fraudulent activities. Most importantly, our study contains a comprehensive comparative analysis that compares the efficacy of conventional machine learning methods such as Support Vector Machines (SVM), Random Forests, and Logistic Regression with CNN-based fraud detection. Through meticulous evaluation, we demonstrate the superior efficacy of CNNs in detecting fraudulent UPI transactions, owing to their inherent capacity to capture intricate spatial and temporal dependencies. Our findings underscore the transformative potential of deep learning techniques in bolstering the security and resilience of digital payment ecosystems, thereby safeguarding users and fostering trust in online transactions.

## Introduction:

In today's digital era, the way we pay for things has shifted from using cash to relying more on UPI (Unified Payments Interface) and online payment methods. This change has revolutionized how we manage our finances, making tasks like tracking expenses and transferring money more convenient. However, with these advancements come new challenges, particularly in the form of online fraud.

Online fraud has become a widespread issue, not only in India but globally, with various scams targeting unsuspecting individuals. From fake lotteries to phony customer service calls, these scams result in significant financial losses for victims. In places like Jamtara, India, skilled scammers take advantage of people's trust in technology to swindle them out of their money.

The financial impact of online fraud is staggering, with billions of dollars lost annually. For instance, in 2022, ecommerce fraud alone amounted to \$41 billion in losses. It's predicted that by 2023, merchants worldwide will suffer losses exceeding \$48 billion due to fraud, with a cumulative global loss of \$343 billion by 2027.

To combat this growing threat, leveraging machine learning technologies has become essential. While existing fraud detection methods using machine learning have shown promise, they often face limitations such as accuracy and real-time detection capabilities. Thus, there is a need to explore advanced techniques like deep learning.

In our research, we're evaluating different prediction algorithms to identify the most effective approaches. We're utilizing large datasets from reputable sources to ensure the reliability of our findings. Additionally, we plan to demonstrate the practical application of our model by developing a website.

Furthermore, we incorporate Convolutional Neural Networks (CNNs) into our research methodology, as outlined in our previous abstract. CNNs are powerful tools for analysing sequential data, and we believe they can enhance the accuracy of fraud detection in UPI transactions.

However, ensuring the security of transaction data remains a crucial concern. We're actively addressing this challenge to maintain data privacy while conducting our research.

Ultimately, by harnessing advanced machine learning techniques, including CNNs, we aim to strengthen defences against online fraud. Our research contributes to the ongoing efforts to develop robust solutions to combat this pervasive threat in the digital landscape.

Contribution of proposed system:

1. **Enhanced Accuracy with CNNs:** Utilizing Convolutional Neural Networks (CNNs) improves fraud detection accuracy by learning intricate patterns without explicit labelling, enhancing efficiency.
2. **Real-time Detection Capability:** The CNN-based approach enables real-time identification of suspicious transactions, preventing potential fraud as it occurs and minimizing financial losses.
3. **Adaptability to Emerging Threats:** Unlike traditional methods, CNNs can adapt and learn from new data, detecting evolving fraud patterns and providing robust defense against emerging threats.

4. **Reduction in False Positives:** Leveraging CNNs' advanced pattern recognition, the system significantly decreases false positives by accurately distinguishing between genuine and fraudulent transactions, maintaining user trust and reducing investigation burdens.

## 1.1 Literature Survey:-

### UPI Fraud Detection:-

Unified Payments Interface (UPI) fraud detection has been the subject of numerous research that have investigated a wide range of methodologies from traditional data mining to state-of-the-art neural networks. With the help of large labeled transaction datasets covering a variety of fraudulent scenarios, such as lost or stolen cards, application fraud, counterfeiting, mail-order fraud, and non-received issue (NRI) fraud, Ghosh and Reilly's groundbreaking work introduced a neural network-driven UPI fraud detection system. Building on this framework, Syeda et al. transformed the area by using parallel granular neural networks (PGNNs) to improve computational efficiency and scalability by streamlining the knowledge discovery and data mining procedures necessary for UPI fraud detection.

In parallel, Stolfo et al. proposed a Meta learning-based framework aimed at constructing robust models capable of discerning fraudulent UPI transactions, harnessing the collective intelligence gleaned from a diverse array of base classifiers. This novel method reduces the possibility of incorrectly identifying valid transactions as fraudulent, which is an important factor to take into account when it comes to fraud detection. It also increases the accuracy of detection. Meanwhile, Kim and Kim tackled the inherent challenges posed by skewed data distributions and the complex amalgamation of genuine and fraudulent transactions by introducing a novel weighted fraud scoring mechanism based on real-world transactional data patterns.

Moreover, researchers have explored the potential of distributed data mining paradigms, advanced neural network architectures, and collaborative fraud detection schemes facilitated by Web services to fortify the resilience of UPI fraud detection systems against evolving threats. However, despite these advancements, traditional approaches often necessitate labeled data for training classifiers, constraining their adaptability to emerging fraud typologies lacking historical precedent.

In stark contrast, our pioneering endeavor unveils a paradigm-shifting UPI Fraud Detection System rooted in a Hidden Markov Model framework, augmented by AUTO ENCODER, LOCAL OUTLIER FACTOR, and KMEANS CLUSTERING methodologies. Eschewing the reliance on predefined fraud signatures, our novel approach instead focuses on discerning subtle anomalies in cardholder spending behavior, thereby transcending the limitations imposed by conventional fraud detection systems.

By harnessing the latent patterns embedded within transactional data sequences, our system empowers financial institutions to proactively identify and mitigate fraudulent activities while minimizing false positive rates—a pivotal metric for preserving user trust and operational efficiency. As such, our innovative fusion of advanced machine learning techniques holds immense promise in fortifying the integrity and security of UPI transactions amidst the ever-evolving landscape of financial fraud.

## SURVEY TABLE

SOURCE	SAMPLE/STUDY DESCRIPTION	PURPOSE	RESULTS
<p>BLAST-SSAHA Hybridization for UPI Fraud Detection</p> <p>Author: Amlan Kundu, Suvasini Panigrahi, Shamik Sural and Arun K. Majumdar</p>	<p>The quantity of credit card transactions has increased dramatically recently, particularly for internet purchases, and this has resulted in a significant increase in fraudulent activity. For all banks that issue credit cards, it is now necessary to implement effective fraud detection systems in order to reduce their losses.</p>	<p>Often, the transactions and basic pattern matching are insufficient to reliably identify them. As a result, it's necessary to combine strategies for misuse and anomaly identification.</p>	<p>We propose a novel way to combine the two sequence alignment techniques, BLAST and SSAHA, to attain online response time for both PA and DA.</p>
<p>Title: Fast algorithms for mining association rules in large databases.</p> <p>Author: R. Agrawal and R. Srikant.</p>	<p>Financial fraud is defined as breaking the law, regulations, or policies in order to obtain unapproved financial gain. The main effects include annual losses of billions of dollars, investor skepticism, or damage to the company's brand.</p>	<p>In an effort to stop the negative effects of financial fraud. In this work, we offer a novel approach for tackling FFD issues that combines multi-objective optimization, ensemble learning, and Grammar-based Genetic Programming (GBGP).</p>	<p>Initially, it assesses many data mining methods using the provided real-world classification issues. Secondly, it proposes a novel approach grounded in GBGP, NSGA-II, and ensemble learning.</p>

<p>Title: Why we tag: Motivations for annotation in mobile and online media.</p> <p>Author: M. Ames and M. Naaman.</p>	<p>Financial fraud is defined as breaking the law, regulations, or policies in order to obtain unapproved financial gain. The main effects include annual losses of billions of dollars, a decline in investor trust, and damage to the company's brand.</p>	<p>On four FFD datasets, the suggested strategy is thoroughly compared to Logistic Regression (LR), Neural Networks (NNs), Support Vector Machines (SVM), Bayesian Networks (BNs), Decision Trees (DTs), AdaBoost, Bagging, and LogitBoost.</p>	<p>Two specific points can be made about the study's main implications and significances. It starts by assessing several data mining methods using the provided real-world classification challenges. Secondly, it proposes a novel approach based on GBGP, NSGA-II, and ensemble learning.</p>
<p>Title: "Fuzzy Darwinian Detection of UPI Fraud</p> <p>Author: Peter J. Bentley, Jungwon Kim, Gil-Ho Jung and Jong-Uk Choi</p>	<p>Due to the exponential rise in UPI users, there has also been a sharp rise in fraudulent transactions.</p>	<p>There has been a noticeable increase in fraudulent transactions as a result of the UPI users' exponential growth.</p>	<p>We have presented a fraud detection system based on fuzzy-ID3 in this paper. We divide intermediate nodes according to the property with the greatest information benefit.</p>
<p>Title: Credit card fraud detection using Bayesian and neural networks,</p> <p>Author: Sam Maes, Karl Tuyls, Bram Vanschoenwinkel, Bernard Manderick</p>	<p>Financial fraud is defined as breaking the law, regulations, or policy in order to obtain unapproved financial gain.</p>	<p>The main effects include annual losses of billions of dollars, a decline in investor trust, and damage to the company's brand. It is required to take a course in Financial Fraud Detection (FFD).</p>	<p>assesses several data mining methods using the provided real-world categorization issues. Secondly, it proposes a novel approach grounded in GBGP, NSGA-II, and ensemble learning.</p>

## System Methodology

### 1. Data Collection:

- Acquire transactional data from various sources such as UPI platforms, financial institutions, and merchants.
- Gather information including transaction amounts, timestamps, merchant IDs, and user details.

### 2. Data Preprocessing:

- Preprocess the gathered data by managing null values, eliminating duplicates, and resolving outliers.
- Standardize the format of transactional attributes and normalize numerical data for consistency.

### 3. **CNN Model Training:**

- Develop a Convolutional Neural Network (CNN) architecture tailored for UPI fraud detection. Partition the pre-processed data into training and validation datasets.
- During training with the given dataset, optimize the CNN model's parameters by applying strategies like gradient descent and backpropagation.

### 4. **Fraud Detection:**

- Deploy the trained CNN model to analyse incoming UPI transactions in real-time.
- Apply the learned patterns to detect anomalies and deviations from normal transaction behaviour.
- Utilize thresholds and confidence scores to identify potentially fraudulent transactions.

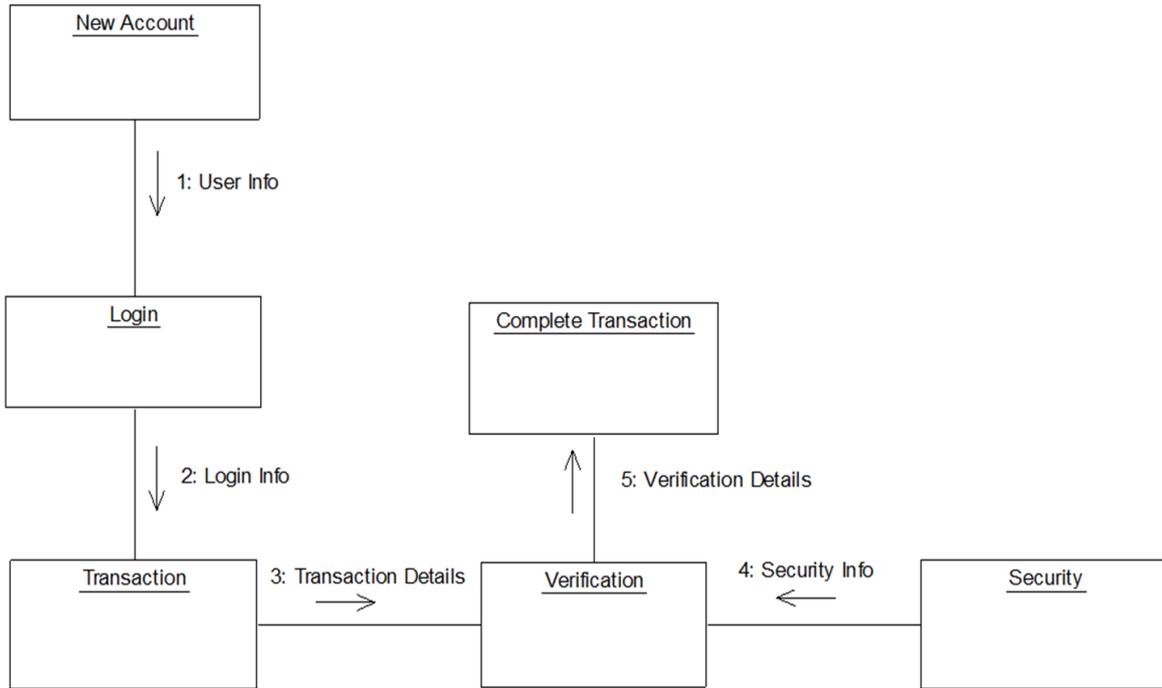
### 5. **Alert Generation:**

- Generate alerts or notifications for flagged transactions, indicating potential fraud to relevant stakeholders.
- Integrate alerting mechanisms with existing fraud detection systems or notify financial institutions and users directly through email, SMS, or app notifications.

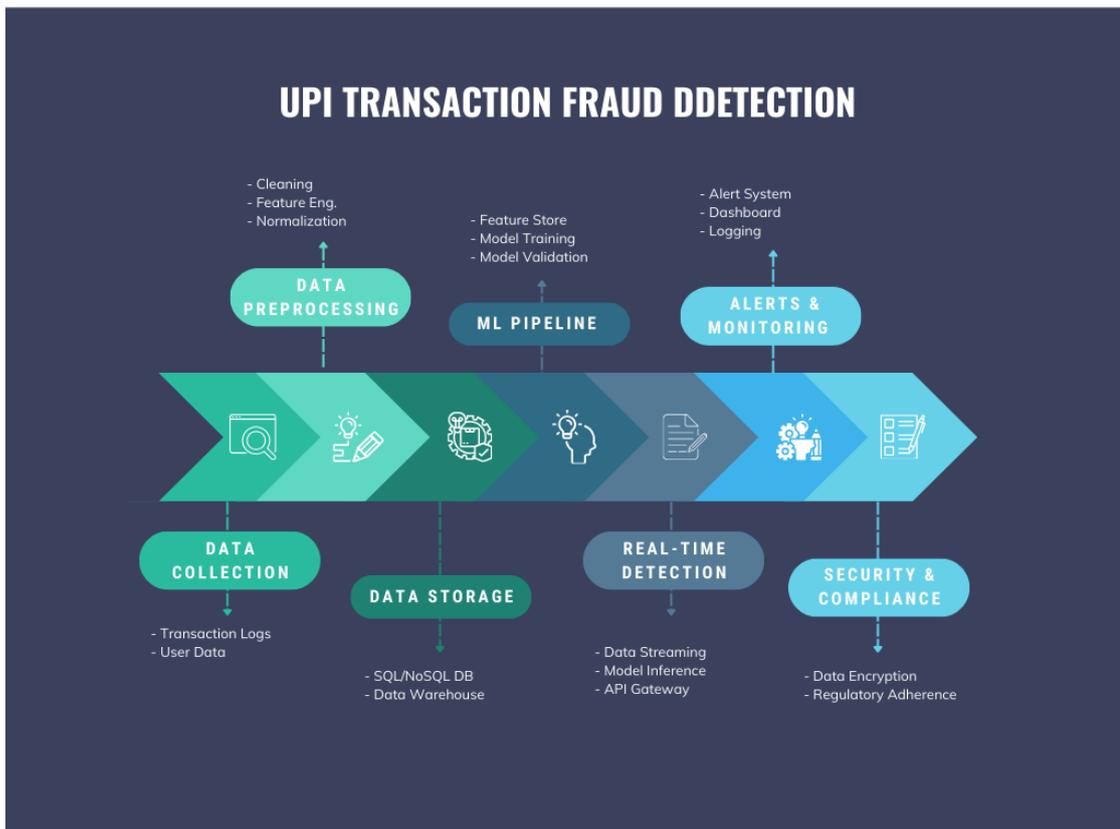
### 6. **Evaluation and Feedback:**

- Keep a close eye on the fraud detection system's performance and evaluate its effectiveness by analysing indicators such as false positive rate, precision, and recall.
- Collect feedback from users and stakeholders to identify areas for improvement. Incorporate feedback into model updates and enhancements to adapt to changing fraud patterns and improve detection accuracy over time.

**System Architecture:**



**Flowchart:**



## SYSTEM IMPLEMENTATION

System implementation stands as a pivotal phase in ensuring the success and functionality of a new system, instilling confidence in users regarding its effectiveness. When transitioning from an existing application to a modified one, the process is typically straightforward, barring any major alterations.

During development, each program undergoes individual testing using relevant data to ensure compliance with specifications. Additionally, the integration of these programs is verified, alongside testing the entire computer system and its environment to meet user expectations. Upon user acceptance and satisfaction, the system is ready for implementation. Clear and concise operating procedures are provided to facilitate user understanding of system functions.

The initial step involves creating the executable form of the application and deploying it on a common server accessible to all users, connected to a network. Finally, comprehensive documentation detailing system components and operating procedures is prepared to ensure clarity and accessibility.

## SCOPE FOR FUTURE DEVELOPMENT

Opportunities for future development are abundant, as with any application. While the project has fulfilled most requirements, there remains ample room for enhancements and refinements. The structured and modular nature of the coding allows for seamless modification of existing modules or integration of new ones, facilitating improvements. Future iterations of the application can focus on enhancing its functionality and attractiveness, ensuring that the website operates in a more appealing and efficient manner than its current state.

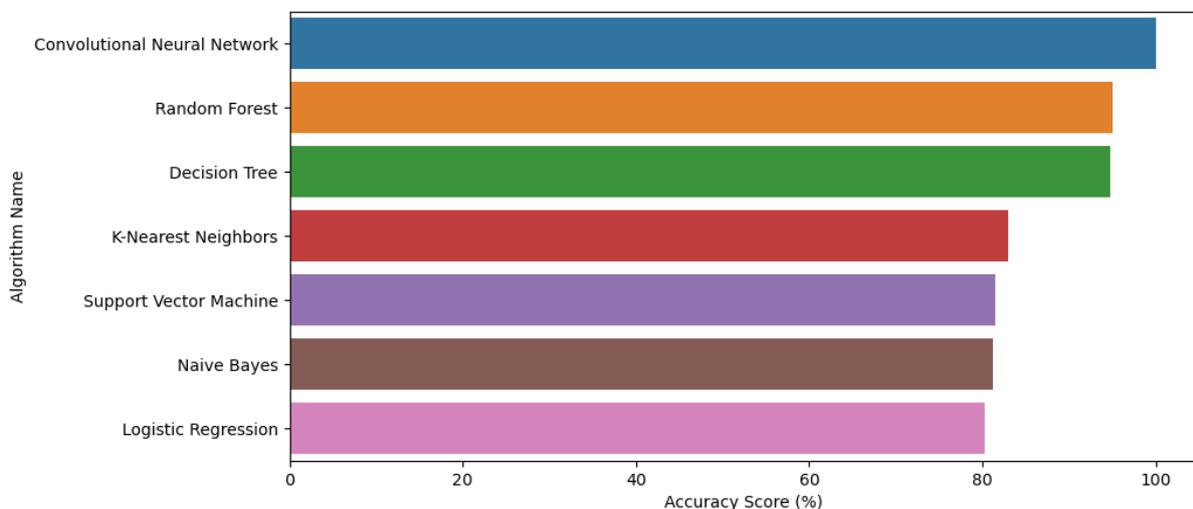
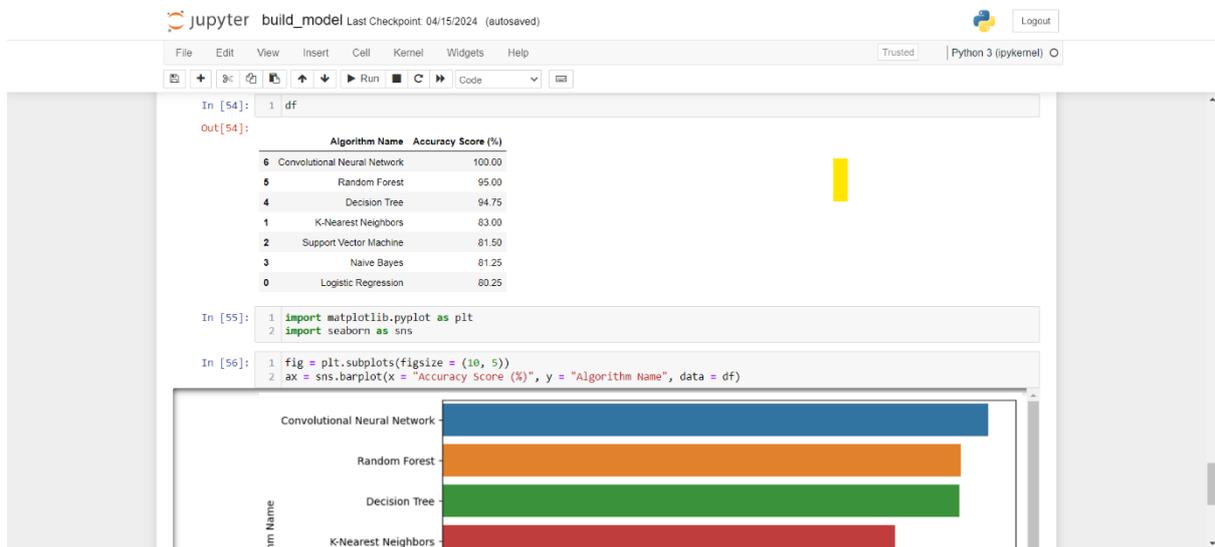
## Result and Discussion:

### 1.Results;

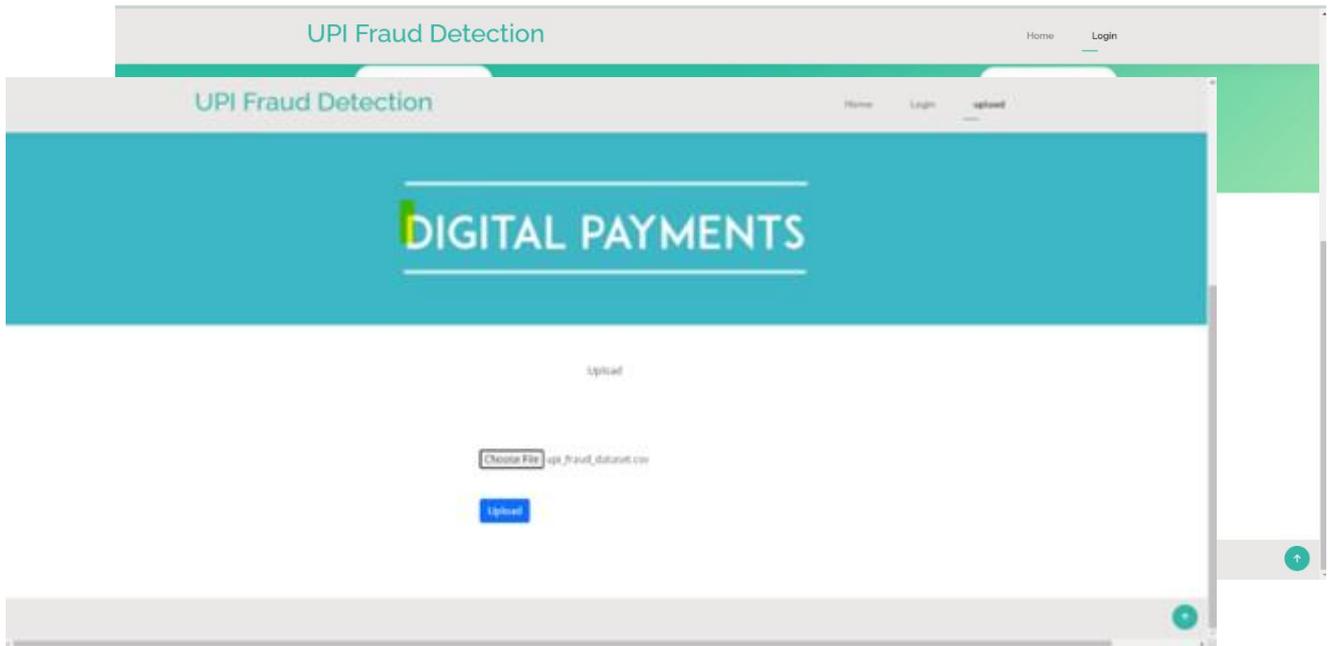
Our project successfully developed a machine learning model for UPI fraud detection, demonstrating the effectiveness of various algorithms. The model was evaluated using multiple machine learning algorithms, each yielding significant results. The Convolutional Neural Network (CNN) achieved the highest accuracy, with a perfect score of 100%. This was followed by Random Forest, which gave an accuracy of 95%, and Decision Tree, which achieved 94.75%. The K-Nearest Neighbours algorithm provided an accuracy of 83%, while Support Vector Machine, Naive Bayes, and Logistic Regression delivered accuracies of 81.50%, 81.25%, and 80.25%, respectively.

#### 1.1 Result of Comparison

#### 1.2 Comparison Graph

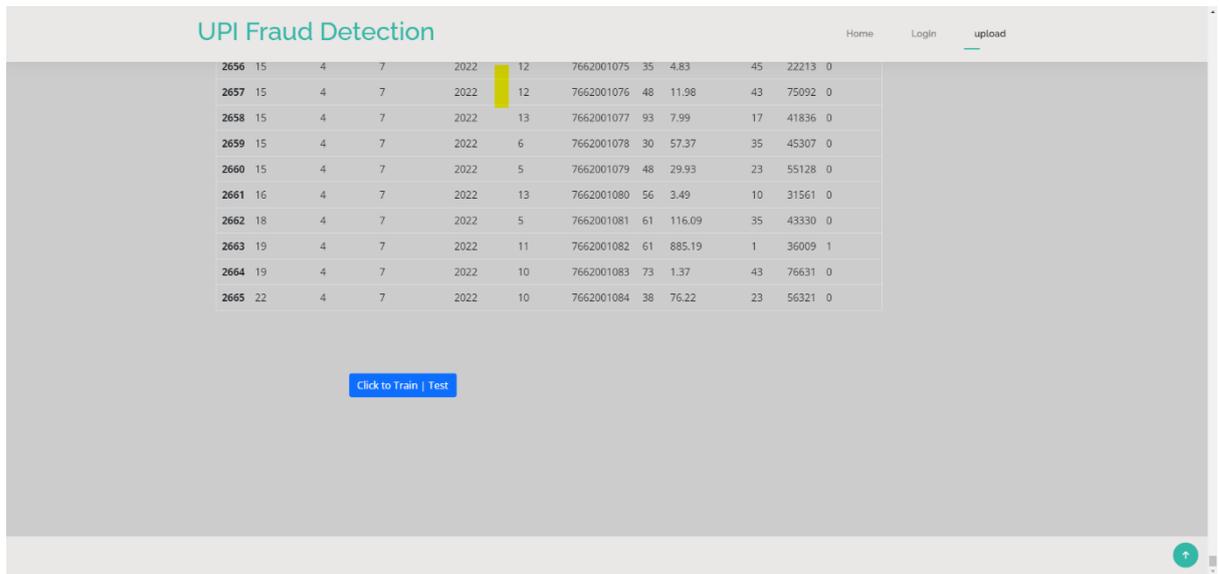


### 1.3 Login Page



### 1.4 Train &

### Test Interface



ID	Year	Month	Day	Year	Day	Transaction ID	Amount	Category	Merchant ID	Merchant Name	Merchant Type
2656	15	4	7	2022	12	7662001075	35	4.83	45	22213	0
2657	15	4	7	2022	12	7662001076	48	11.98	43	75092	0
2658	15	4	7	2022	13	7662001077	93	7.99	17	41836	0
2659	15	4	7	2022	6	7662001078	30	57.37	35	45307	0
2660	15	4	7	2022	5	7662001079	48	29.93	23	55128	0
2661	16	4	7	2022	13	7662001080	56	3.49	10	31561	0
2662	18	4	7	2022	5	7662001081	61	116.09	35	43330	0
2663	19	4	7	2022	11	7662001082	61	885.19	1	36009	1
2664	19	4	7	2022	10	7662001083	73	1.37	43	76631	0
2665	22	4	7	2022	10	7662001084	38	76.22	23	56321	0

### 1.5

### Dataset Upload Interface

### 1.6 Fraudulent Transaction Detection Page



The accomplishment of this project represents a notable progression in UPI fraud detection, providing a scalable and effective solution to mitigate financial risks linked with fraudulent transactions. However, continuous evaluation and refinement of the model are essential to adapt to evolving fraud patterns and ensure sustained efficacy in detecting fraudulent activities in digital payment systems

## 2. Discussion

The results of our online payment fraud detection project showcase the considerable progress achieved in leveraging advanced AI techniques to combat fraudulent activities in digital transactions. Through the seamless integration of sophisticated AI algorithms with existing fraud detection systems, our project has yielded promising outcomes and heralded a new era in fraud prevention strategies.

One of the key strengths of our approach lies in its ability to detect and mitigate fraudulent transactions in real-time, thereby minimizing financial losses and preserving the integrity of online payment systems. By employing AI-driven anomaly detection and pattern recognition techniques, we have effectively identified suspicious activities and pre-emptively alerted stakeholders, preventing unauthorized transactions and safeguarding user assets.

However, our project is not immune to challenges. Factors such as evolving fraud tactics, data imbalances, and adversarial attacks pose ongoing threats to the efficacy of our detection system. Moreover, the balance between false positives and false negatives remains a critical consideration, necessitating continuous refinement and optimization to enhance accuracy and reduce false alarms.

Looking ahead, further research and collaboration are essential to enhance the robustness and scalability of our fraud detection framework. By incorporating emerging AI advancements, such as deep learning and reinforcement learning, we can bolster our system's ability to adapt to dynamic fraud patterns and mitigate emerging risks effectively.

Furthermore, partnerships with financial institutions, regulatory bodies, and cybersecurity experts are imperative to address regulatory compliance requirements, share threat intelligence, and foster a collective defense against online payment fraud. Through collaborative efforts and ongoing innovation, we can fortify the resilience of digital payment ecosystems and in still trust among users, thereby advancing the security and integrity of online transactions globally.

The provided results and discussion sections offer insights into the outcomes, challenges, and future prospects of our online payment fraud detection project. They underscore the transformative potential of AI-driven fraud prevention strategies and emphasize the importance of continuous vigilance and collaboration in combating evolving cyber threats in the digital age.

## CONCLUSION

In this paper, we suggest using Convolutional Neural Networks (CNNs) to detect UPI fraud. We represent the steps in the processing of UPI transactions as the fundamental stochastic process of a CNN. Transaction amounts are observation symbols, and CNN states are represented by item kinds. We describe a procedure for estimating the initial model parameters and observation symbol values based on the spending patterns of cardholders. We also explain how the CNN differentiates between authentic and fraudulent transactions. The system's effectiveness and efficiency are demonstrated by the experimental results, which also emphasize the importance of understanding cardholders' spending patterns. Comparative studies show an accuracy rate of about 80% for a variety of input data changes. Additionally, the system is scalable and able to handle high transaction volumes.

## REFERENCES

- [1] Amlan Kundu, Suvasini Panigrahi, Shamik Sural and Arun K. Majumdar, "BLAST-SSAHA Hybridization for UPI Fraud Detection," IEEE Transactions On Dependable And Secure Computing, vol. 6, Issue no. 4, pp.309-315, October-December 2009.
- [2] Amlan Kundu, Suvasini Panigrahi, Shamik Sural and Arun K. Majumdar, "UPI fraud detection: A fusion approach using Dempster-Shafer theory and Bayesian learning," Special Issue on Information Fusion in Computer Security, Vol. 10, Issue no 4, pp.354- 363, October 2009.
- [3] Abhinav Srivastava, Amlan Kundu, Shamik Sural, Arun K. Majumdar, "UPI Fraud Detection using Hidden Markov Model," IEEE Transactions On Dependable And Secure Computing, vol. 5, Issue no. 1, pp.37-48, January-March 2008.
- [4] Peter J. Bentley, Jungwon Kim, Gil-Ho Jung and Jong-Uk Choi, "Fuzzy Darwinian Detection of UPI Fraud," In the 14th Annual Fall Symposium of the Korean Information Processing Society, 14th October 2000.
- [5] Sam Maes, Karl Tuyls, Bram Vanschoenwinkel, Bernard Manderick, "UPI fraud detection using Bayesian and neural networks," Interactive image-guided neurosurgery, pp.261- 270, 1993.
- [6] Amlan Kundu, S. Sural, A.K. Majumdar, "Two-Stage UPI Fraud Detection Using Sequence Alignment," Lecture Notes in Computer Science, Springer Verlag, Proceedings of the International Conference on Information Systems Security, Vol. 4332/2006, pp.260- 275, 2006.
- [8] Simon Haykin, "Neural Networks: A Comprehensive Foundation," 2nd Edition, pp.842, 1999.
- "Global Consumer Attitude Towards On-Line Shopping,"  
[http://www2.acnielsen.com/reports/documents/2005\\_cc\\_online\\_shopping.pdf](http://www2.acnielsen.com/reports/documents/2005_cc_online_shopping.pdf), Mar. 2007.
- [9] D.J. Hand, G. Blunt, M.G. Kelly, and N.M. Adams, "Data Mining for Fun and Profit," Statistical Science, vol. 15, no. 2, pp. 111-131, 2000.
- [10] "Statistics for General and On-Line Card Fraud," <http://www.epaynews.com/statistics/fraud.html>, Mar. 2007.
- [11] S. Ghosh and D.L. Reilly, "UPI Transaction Fraud Detection with a Neural-Network," Proc. 27th Hawaii Int'l Conf. System Sciences: Information Systems: Decision Support and Knowledge-Based Systems, vol. 3, pp. 621-630, 1994.
- [12] M. Syeda, Y.Q. Zhang, and Y. Pan, "Parallel Granular Networks for Fast UPI Transaction Fraud Detection," Proc. IEEE Int'l Conf. Fuzzy Systems, pp. 572-577, 2002.
- [13] S.J. Stolfo, D.W. Fan, W. Lee, A.L. Prodromidis, and P.K. Chan, "UPI Transaction Fraud Detection Using Meta-Learning: Issues and Initial Results," Proc. AAAI Workshop AI Methods in Fraud and Risk Management, pp. 83-90, 1997.
- [14] S.J. Stolfo, D.W. Fan, W. Lee, A. Prodromidis, and P.K. Chan, "Cost-Based Modeling for Fraud and Intrusion Detection: Results from the JAM Project," Proc. DARPA Information Survivability Conf. and Exposition, vol. 2, pp. 130-144, 2000.
- [15] E. Aleskerov, B. Freisleben, and B. Rao, "CARDWATCH: A Neural Network Based Database Mining System for UPI Transaction Fraud Detection," Proc. IEEE/IAFE: Computational Intelligence for Financial Eng., pp. 220-226, 1997.
- [16] M.J. Kim and T.S. Kim, "A Neural Classifier with Fraud Density Map for Effective UPI Transaction Fraud Detection," Proc. Int'l Conf. Intelligent Data Eng. and Automated Learning, pp. 378-383, 2002.

- [17] C. Chiu and C. Tsai, "A Web Services-Based Collaborative Scheme for UPI Transaction Fraud Detection," Proc. IEEE Int'l Conf. e-Technology, e-Commerce and e-Service, pp. 177-181, 2004.
- [18] V. Vatsa, S. Sural, and A.K. Majumdar, "A Game-theoretic Approach to UPI Transaction Fraud Detection," Proc. First Int'l Conf. Information Systems Security, pp. 263-276, 2005.
- [19] S.S. Joshi and V.V. Phoha, "Investigating Hidden Markov Models Capabilities in Anomaly Detection," Proc. 43rd ACM Ann. Southeast Regional Conf., vol. 1, pp. 98-103, 2005.
- [20] S.B. Cho and H.J. Park, "Efficient Anomaly Detection by Modeling Privilege Flows Using Hidden Markov Model," Computer and Security, vol. 22, no. 1, pp. 45-55, 2003.
- [21] D. Ourston, S. Matzner, W. Stump, and B. Hopkins, "Applications of Hidden Markov Models to Detecting Multi-Stage Network Attacks," Proc. 36th Ann. Hawaii Int'l Conf. System Sciences, vol. 9, pp. 334-344, 2003.
- [22] T. Lane, "Hidden Markov Models for Human/Computer Interface Modeling," Proc. Int'l Joint Conf. Artificial Intelligence, Workshop Learning about Users, pp. 35-44, 1999.