IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.134143

Application of Artificial Intelligence for Fraudulent Banking Operations Recognition

T.Sreekanth[1], S. Anil Reddy[2], U. Govardhan[3], M. Ramnath[4], Dr. G. Kishor Kumar[5]

Student, Computer Science and Engineering & Business Systems, RGMCET, Nandyala, India[1-4]

Computer Science and Engineering & Business Systems, RGMCET, Nandyala, India[5]

Abstract This research explores the application of artificial intelligence in detecting bank fraud, a problem exacerbated by the COVID-19 pandemic's shift to online operations and the proliferation of charitable funds used by criminals to deceive users. The study focuses on leveraging machine learning algorithms to analyze and identify fraudulent transactions in online banking. Its key contribution lies in developing machine learning models tailored for detecting fraudulent banking activities, along with preprocessing techniques to enhance data comparison and result selection. Additionally, the paper elaborates on methods to enhance detection accuracy, including managing highly imbalanced datasets, transforming features, and engineering new features. However, this paper proposes the use of Convolutional Neural Network (CNN) for UPI fraud detection. The CNN model is designed to analyze the spending profile of cardholders, thereby enhancing the accuracy of fraud detection. The Fraud Detection System (FDS) implemented in the bank monitors the spending patterns of cardholders, automatically blocking transactions deemed unusual and alerting the bank for further investigation. This approach minimizes the need for manual intervention and ensures swift action against fraudulent activities, thereby safeguarding users' financial assets.

Keywords: Transaction, Payment, UPI, Attackers, Fraudulent, Money, Datasets, Machine learning, recognition of fraudulent operations, Convolutional Neural Networks,

I. INTRODUCTION

The trend of online shopping is experiencing steady growth, as evidenced by an ACNielsen study from 2005, which revealed that approximately one-tenth of the global population engages in online shopping. Notably, Germany and Great Britain boast the highest numbers of online shoppers, with UPI Transaction emerging as the preferred mode of payment, constituting 59 percent of transactions. Barclaycard, the leading UPI Transaction company in the United Kingdom, reported approximately 350 million transactions annually towards the end of the previous century. Major retailers like Wal-Mart handle even larger volumes of UPI Transaction transactions, encompassing both online and in-store purchases.

However, with the global rise in UPI Transaction users, the potential for attackers to exploit UPI Transaction details for fraudulent activities is also escalating. In the United States alone, UPI Transaction fraud amounted to $2.7 billion in 2005, with an estimated increase to $3.0 billion in 2006, out of which $1.6 billion and $1.7 billion, respectively, were attributed to online fraud.Credit card transactions typically fall into two categories: physical card and virtual card. In physical card transactions, the cardholder physically presents the card to the merchant for payment. Perpetrating fraudulent transactions in this scenario requires the theft of the physical card. Failure by the cardholder to promptly report the loss of the card can result in significant financial losses for the UPI Transaction company. Conversely, virtual card-based purchases necessitate only key card details (such as card number, expiration date, and security code) for payment, commonly conducted online or over the phone. Fraudsters can exploit this by acquiring card details without the cardholder's knowledge. Detecting such fraud entails analyzing spending patterns associated with each card to identify any deviations from the norm. Leveraging existing purchase data to analyze cardholder behavior offers a promising avenue for reducing successful UPI Transaction frauds. Each cardholder can be represented by a set of behavioral patterns, encompassing typical purchase categories, time since last purchase, expenditure amounts, etc. Any deviation from these patterns signals a potential threat to the system. Various techniques for UPI Transaction fraud detection have emerged in recent years, aiming to address this growing challenge. The realm of UPI Transaction fraud detection has garnered significant attention in research circles, resulting in the exploration of various techniques, notably emphasizing data mining and neural networks. Ghosh and Reilly introduced a neural network-based approach for UPI Transaction fraud detection, utilizing a system trained on a substantial dataset containing labeled UPI Transaction account transactions, encompassing instances of fraud such as lost cards, stolen cards, application fraud, counterfeit fraud, mail-order fraud, and non-received issue (NRI) fraud. Syeda et al. implemented parallel granular neural networks (PGNNs) to enhance the speed of data mining and knowledge discovery in UPI Transaction fraud detection.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024
DOI: 10.17148/IJARCCE.2024.134143

Stolfo et al. proposed a UPI Transaction fraud detection system (FDS) leveraging Meta learning techniques to learn models of fraudulent transactions, wherein a metaclassifier is trained on the correlation of predictions from base classifiers. Aleskerov et al. introduced CARDWATCH, a database mining system utilizing neural learning modules for UPI Transaction fraud detection, interfacing with various commercial databases. Kim and Kim identified skewed data distribution and a mix of legitimate and fraudulent transactions as key challenges in UPI Transaction fraud detection, proposing a method to calculate fraud density and generate weighted fraud scores to mitigate misdetections. Fan et al. advocated for distributed data mining in UPI Transaction fraud detection, while Brause et al. developed an approach integrating advanced data mining techniques and neural network algorithms to achieve extensive fraud coverage.

II. LITERATURE SURVEY

UPI Fraud Detection:-

UPI fraud detection has drawn a lot of research interest and a number of techniques, with special emphasis on data mining and neural networks, have been suggested. Ghosh and Reilly have proposed UPI fraud detection with a neural network. They have built a detection system, which is trained on a large sample of labelled UPI account transactions. These transactions contain exam-ple fraud cases due to lost cards, stolen cards, application fraud, counterfeit fraud, mail-order fraud, and nonreceived issue (NRI) fraud. Recently, Syeda et al. have used parallel granular neural networks (PGNNs) for improving the speed of data mining and knowledge discovery process in UPI fraud detection. A complete system has been implemented for this purpose. Stolfo et al. suggest a UPI fraud detection system (FDS) using Metalearning techniques to learn models of fraudulent UPI transactions. Metalearning is a general strategy that provides a means for combining and integrating a number of separately built classifiers or models. A metaclassifier is thus trained on the correlation of the predictions of the base classifiers. The same group has also worked on a cost-based model for fraud and intrusion detection. They use Python agents for Metalearning (JAM), which is a distributed data mining system for UPI fraud detection. A number of important performance metrics like True Positive—False Positive (TP-FP) spread and accuracy have been defined by them. Aleskerov et al. present CARDWATCH, a database mining system used for UPI fraud detection. The system, based on a neural learning module, provides an interface to a variety of commercial databases. Kim and Kim have identified skewed distribution of data and mix of legitimate and fraudulent transactions as the two main reasons for the complexity of UPI fraud detection. Based on this observation, they use fraud density of real transaction data as a confidence value and generate the weighted fraud score to reduce the number of misdetections. Fan et al. suggest the application of distributed data mining in UPI fraud detection. Brause et al. have developed an approach that involves advanced data mining techniques and neural network algorithms to obtain high fraud coverage. Chiu and Tsai have proposed Web services and data mining techniques to establish a collaborative scheme for fraud detection in the banking industry. With this scheme, participating banks share knowledge about the fraud patterns in a heterogeneous and distributed environment. To establish a smooth channel of data exchange, Web services techniques such as XML, SOAP, and WSDL are used. Phua et al. have done an extensive survey of existing data-mining-based FDSs and published a comprehensive report. Prodromidis and Stolfo use an agent-based approach with distributed learning for detect-ing frauds in UPI transactions. It is based on artificial intelligence and combines inductive learning algorithms and Metalearning methods for achieving higher accuracy. Phua et al. suggest the use of metaclassifier similar to in fraud detection problems. They consider naive Bayesian C4.5, and Back Propagation neural networks as the base classifiers. A metaclassifier is used to determine which classifier should be considered based on skewness of data. Although they do not directly use UPI fraud detection as the target application, their approach is quite generic. Vatsa et al. have recently proposed a game-theoretic approach to UPI fraud detection. They model the interaction between an attacker and an FDS as a multistage game between two players, each trying to maximize his payoff. The problem with most of the abovementioned approaches is that they require labelled data for both genuine, as well as fraudulent transactions, to train the classifiers. Getting real-world fraud data is one of the biggest problems associated with UPI fraud detection. Also, these approaches cannot detect new kinds of frauds for which labelled data is not available. In contrast, we present a Hidden Markov Model (AUTO ENCODER, LOCAL OUTLIER FACTOR, KMEANS CLUSTERING)-based UPI FDS, which does not require fraud signatures and yet is able to detect frauds by considering a cardholder's spending habit. We model a UPI transaction processing sequence by the stochastic process of an AUTO ENCODER, LOCAL OUTLIER FACTOR, KMEANS CLUSTERING. The details of items purchased in individual transactions are usually not known to an FDS running at the bank that issues UPI to the cardholders. This can be represented as the underlying finite Markov chain, which is not observable. The transactions can only be observed through the other stochastic process that produces the sequence of the amount of money spent in each transaction. Hence, we feel that AUTO ENCODER, LOCAL OUTLIER FACTOR, KMEANS CLUSTERING is an ideal choice for addressing this problem. Another important advantage of the AUTO ENCODER, LOCAL OUTLIER FACTOR, KMEANS CLUSTERING-based approach is a drastic reduction in the number of False Positives (FPs)—transactions identified as malicious by an FDS although they are actually genuine. Since the number of genuine transactions is a few orders of magnitude higher than the number of malicious transactions, an FDS should be designed in such a way that the number of FPs is as low as possible.

DOI: 10.17148/IJARCCE.2024.134143

Otherwise, due to the "base rate fallacy" effect, bank administrators may tend to ignore the alarms. To the best of our knowledge, there is no other published literature on the application of AUTO ENCODER, LOCAL OUTLIER FACTOR, KMEANS CLUSTERING for UPI fraud detection.

III. METHODOLOGY

a) Proposed Work:

Here we are introducing a project for the UPI fraud detection using Convolutional Neural Network (CNN). It is done on the basis of the spending profile of the card holder. The usual spending of the cardholder is being checked by the FDS (Fraud Detection system) in the bank .The system checks all the spending of the user. When it turns unusual the method blocks the transaction on the card. And it alerts the bank. It occurs automatically. It doesn't need any man power.

Advantages:

1) The main advantage is that the detection occurs much faster than any other method.
2) In all the existing systems the real card holder should checked for the Fraud detection. But in our method there is no need of the physical inconveniences of the card holder. All the checking and the detection occur automatically.
3) This project needs no man power for the detection.
4) This project provides most accurate method in UPI fraud detection.

Modules

There are five main types of modules in the fraud detection

- Register.
- Sign in.
- Security.
- User side.
- Purchase.

Module Description

1. Login: The Login module gives a login form to the user with a user name and password. The user can access the special features only when they enters correct user name and password

2. Register: In this module the hard holder registers new card. For this they are gives their personal details, UPI details. In this module the user also can fix security questions and answers for security purpose

3. Security: In this module we provide special features are the user can fix a spending limit, set security questions and answers. The purpose of this segment is this security questions arise when the user exceeds the spending limit. The user can access further only when they answer these questions correctly.

4. User side: This module is for the user to view the home page, purchase things or view reports. This report deals with what the user did with the UPI like purchase, transactions etc.

5. Purchase: In this module all the transaction process like purchase with the UPI occurs. The user submits the total amount to be credited after the completion of purchase. The transaction occurs only when the total amount is below the spending limit. If it exceeds the limit security questions are asked. The user can proceed only when the answers are correct. Otherwise the card will be blocked.

b) System Architecture:

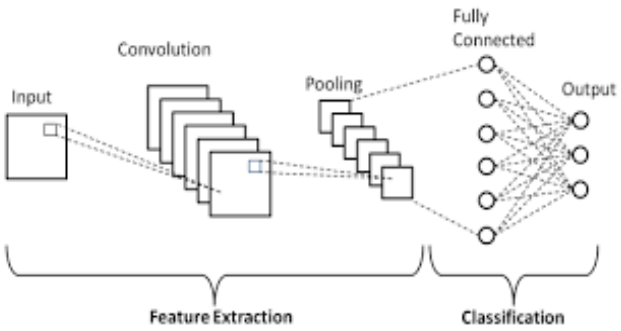


Fig1 Proposed Architecture

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.134143

Certainly! Here's how you can modify the implementation to include pooling layers and fully connected layers in a Convolutional Neural Network (CNN) for identifying numeric fraud values in binary format:

1. Data Preparation:
- Collect a dataset consisting of labeled numeric transactions, where each transaction is represented in binary format.
- Split the dataset into training, validation, and test sets.

2. Preprocessing:
- Normalize the input data to ensure that features have similar scales.
- Convert the binary data into a suitable format for input into the CNN.

3. Model Architecture:
- Design a CNN architecture suitable for binary input data.
- The input layer will consist of binary feature vectors representing each transaction.
- Include convolutional layers with appropriate filters to extract features from the binary input.
- Add pooling layers after convolutional layers to reduce the dimensionality of the feature maps and capture the most important features.
- Flatten the output of the last convolutional layer to prepare it for input into fully connected layers.
- Add one or more fully connected layers to learn complex patterns in the data.
- Include activation functions like ReLU to introduce non-linearity into the model.
- Use dropout layers to prevent overfitting.
- The output layer should have a single neuron with a sigmoid activation function, as the task is binary classification (fraudulent or non-fraudulent).

4.Model Training:
- Compile the model with an appropriate loss function (e.g., binary cross-entropy) and optimizer (e.g., Adam).
- Train the model using the training data while monitoring performance on the validation set.
- Adjust hyperparameters such as learning rate, batch size, and number of epochs based on validation performance to prevent overfitting.

5. Evaluation:
- Evaluate the trained model using the test set to assess its performance on unseen data.
- Calculate metrics such as accuracy, precision, recall, and F1-score to evaluate the model's performance.
- Visualize the model's performance using confusion matrices and ROC curves.

6. Deployment:
- Once satisfied with the model's performance, deploy it for fraud detection in a real-world scenario.
- Integrate the model into the existing fraud detection system to automatically identify fraudulent transactions in binary format.

This approach leverages both convolutional layers for feature extraction and pooling layers and fully connected layers for learning complex patterns in the data. Adjust the architecture and hyperparameters as needed based on the characteristics of your dataset and the requirements of the fraud detection task.

IV. EXPERIMENTAL RESULTS

Accuracy: The accuracy of a test is its ability to differentiate the patient and healthy cases correctly. To estimate the accuracy of a test, we should calculate the proportion of true positive and true negative in all evaluated cases. Mathematically, this can be stated as:

Accuracy = TP + TN TP + TN + FP + FN.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.134143

F1-Score: F1 score is a machine learning evaluation metric that measures a model's accuracy. It combines the precision and recall scores of a model. The accuracy metric computes how many times a model made a correct prediction across the entire dataset.

$$mAP = \frac{1}{n}\sum_{k=1}^{k=n} AP_k$$

$$AP_k = \text{the AP of class } k$$

$$n = \text{the number of classes}$$

$$\text{F1 Score} = \frac{2}{\left(\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}\right)}$$

$$\text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Precision: Precision evaluates the fraction of correctly classified instances or samples among the ones classified as positives. Thus, the formula to calculate the precision is given by:

Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

Recall: Recall is a metric in machine learning that measures the ability of a model to identify all relevant instances of a particular class. It is the ratio of correctly predicted positive observations to the total actual positives, providing insights into a model's completeness in capturing instances of a given class.

mAP50: The mAP for object detection is the average of the AP calculated for all the classes. mAP@0.5 means that it is the mAP calculated at IOU threshold 0.5. The general definition for the Average Precision(AP) is finding the area under the precision-recall curve.

$$Recall = \frac{TP}{TP + FN}$$

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.134143

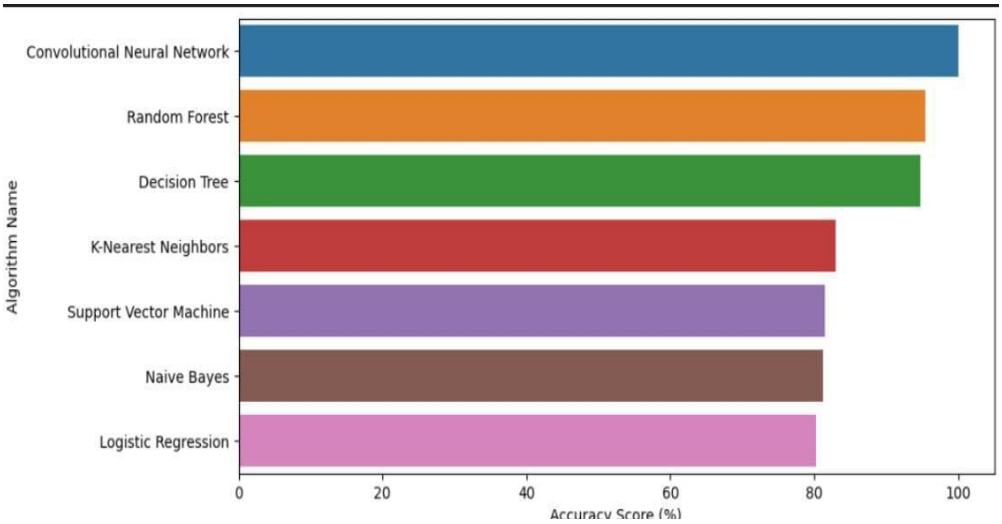
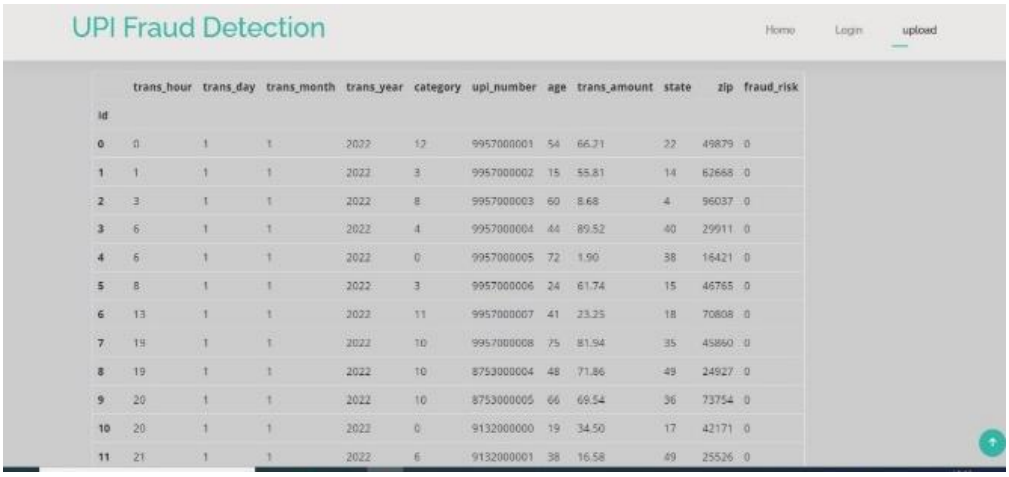
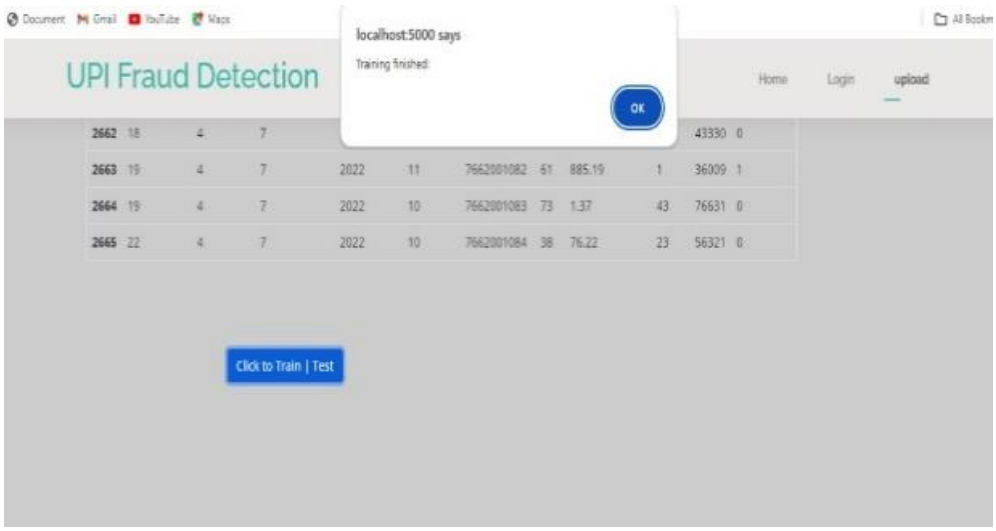


Fig 2 Accuracy Comparison Graph – Classification
On various Algorithms



Id	trans_hour	trans_day	trans_month	trans_year	category	upi_number	age	trans_amount	state	zip	fraud_risk
0	0	1	1	2022	12	9957000001	54	66.71	22	49879	0
1	1	1	1	2022	3	9957000002	15	55.81	14	62668	0
2	3	1	1	2022	8	9957000003	60	8.68	4	96037	0
3	6	1	1	2022	4	9957000004	44	89.52	40	29911	0
4	6	1	1	2022	0	9957000005	72	1.90	38	16421	0
5	8	1	1	2022	3	9957000006	24	61.74	15	46765	0
6	13	1	1	2022	11	9957000007	41	23.25	18	70808	0
7	19	1	1	2022	10	9957000008	75	81.94	35	45860	0
8	19	1	1	2022	10	8753000004	48	71.86	49	24927	0
9	20	1	1	2022	10	8753000005	66	69.54	36	73754	0
10	20	1	1	2022	0	9132000000	19	34.50	17	42171	0
11	21	1	1	2022	6	9132000001	38	16.58	49	25526	0

Fig 3 UPI transaction data set upload



2662	18	4	7							43330	0
2663	19	4	7	2022	11	7662001082	61	885.19	1	36909	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

Fig 4 Trained the dataset

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024

DOI: 10.17148/IJARCCE.2024.134143

Fig 5 Enter the Upi transaction details

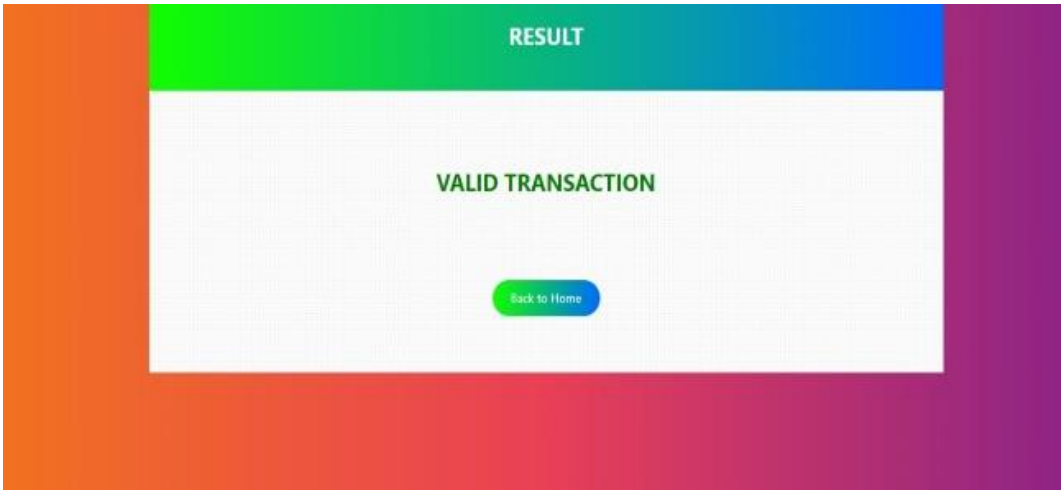


Fig 5 Predicted Results as valid transaction

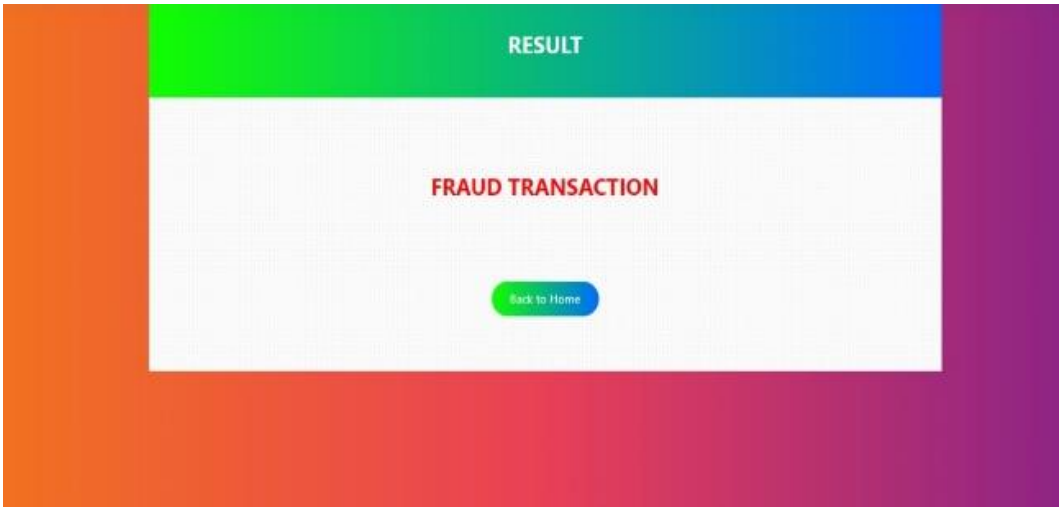


Fig 6 Predicted Results as Fraud transaction

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 4, April 2024
DOI: 10.17148/IJARCCE.2024.134143

V. CONCLUSION

In this project, we have proposed an application of CNN in UPI fraud detection. The different steps in UPI transaction processing are represented as the underlying stochastic process of an CNN. We have used the ranges of transaction amount as the observation symbols, whereas the types of item have been considered to be states of the CNN. We have suggested a method for finding the spending profile of cardholders, as well as application of this knowledge in deciding the value of observation symbols and initial estimate of the model parameters. It has also been explained how the CNN can detect whether an incoming transaction is fraudulent or not. Experimental results show the performance and effectiveness of our system and demonstrate the usefulness of learning the spending profile of the cardholders. Comparative studies reveal that the Accuracy of the system is close to 80 percent over a wide variation in the input data. The system is also scalable for handling large volumes of transactions.

Our study's conclusion is that it's critical for E-Banking users to take precautions to safeguard their personal data and to be knowledgeable about any hazards involved with online banking. Using strong passwords that are only known to you, updating software and security protocols, and exercising caution when exchanging personal information or clicking on links from untrusted sources are some examples of how to do this. Even with all the danger, there were still methods to succeed. We referred to it as security precautions. Physical access control, human aspect: awareness, and antiviral are examples of security measures. Limiting access connections to computer networks, system files, and data is facilitated by physical access control. Therefore, phishing scams can be avoided. Human aspect: By safeguarding our personal information and reporting the loss right away so the bank will repay your account, we can stop the criminal from carrying out his activity even if he is able to steal money from your account. Next, antivirus programmed have unique signatures that provide security and block access to harmful assaults.

VI. FUTURE SCOPE

Every application has its own merits and demerits. The project has covered almost all the requirements.Further requirements and improvements can easily be done since the coding is mainly structured or modular in nature. Changing the existing modules or adding new modules can append improvements. Further enhancements can be made to the application, so that the web site functions very attractive and useful manner than the present one.

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. [7] Simon Haykin, “Neural Networks: A Comprehensive Foundation,” 2nd Edition, pp.842, 1999.

