International Journal of Science and Healthcare Research
Vol. 8; Issue: 2; April-June 2023
Website: ijshr.com
Case Study *ISSN: 2455-7587*

PAN Card Fraud Detection Using Machine Learning: A Case Study

Rohini Hanchate[1], Shreyas Yerole[2], Nazir Lalloti[3], Piyush Mahajan[4]

[1]Prof. Department of Computer Engineering, [2]Department of Computer Engineering,
[3]Department of Computer Engineering, [4]Department of Computer Engineering,
Nutan Maharashtra Institute of Engineering and Technology, SPPU University, Pune, India

Corresponding Author: Rohini Hanchate

DOI: https://doi.org/10.52403/ijshr.20230225

ABSTRACT

Artificial Intelligence research in the area of computer vision teaches machines to comprehend and interpret visual data. Machines can accurately identify and classify objects using digital images from cameras and videos, deep learning models, and then respond to what they "see". The detection of document manipulation is an area where these technologies are crucial. Everyone in India is given a PAN (Permanent Account Number), which is a 10-character alphanumeric code, especially those who pay taxes. Various organizations make use of identity documents like PAN cards. The PAN Card given by the employee or client may be genuine or fraudulent. This review paper's main goal is to use computer vision to determine whether the provided PAN card image is authentic or altered (fake). Our goal is to create a method that uses machine learning techniques to categorize PAN cards.

Keywords: Machine Learning, Convolutional Neural Network, Fraud, Image Processing

INTRODUCTION

In recent years, as there is a growing population and advancement of technology, most of us are having PAN cards as a criterion of eligibility, so the frauds associated with it are also rising gradually. In the present world, most of the enterprises from small to big industries are using PAN card for verification. Fraud is happening in all organizations such as the appliance industry, automobile industry, IT industry, banks and so on. Since, there is a need for efficient and effective algorithms to be developed that works significantly. We try to avoid the fraudster using PAN cards for his own benefit and wrongly using them. This can be avoided by using artificial intelligence and compared with few other machine learning algorithms. Machines can effectively detect and classify items using digital images from cameras and machine learning models. They may then respond to what they "see" in the world.

A subclass of artificial intelligence, machine learning is one of the most popular topics of this decade. To enhance their services, more and more businesses are looking to invest in machine learning. In order to enable the computer to carry out tasks without hard coding, machine learning combines several computer techniques with statistical modelling. From the "training data," the acquired model would be learning. From the accumulated experiential information, predictions can be made or actions can be taken. Machine learning techniques that use Artificial Neural Networks include deep learning models. There are numerous techniques, including convolutional neural networks, restricted Boltzmann machines, deep belief networks, auto-encoders, and recurrent neural networks. A properly trained CNN would be able to identify distinctive associations across the entire dataset.

As a result, bank transactions are a huge problem in today's technology society, where credit card fraud has emerged as the main problem. Sensitive data is lost as a result of several fraudulent transactions that are difficult for both the user and the banking authority to detect. Based on the behaviour of the transactions, a variety of models are used to identify fraud transactions. These models can be divided into two main categories: supervised learning and unsupervised learning algorithms. They have employed techniques like Cluster Analysis, Support Vector Machine, Naive Bayer's Classification, etc. in the current system to determine the accuracy of the fraudulent actions. This study uses the Random Forest Algorithm to determine the accuracy of fraudulent transactions.

LITERATURE REVIEW

Paper Name: Credit card fraud detection using artificial neural network
Author: Asha RB, Suresh Kumar K
Most people now use credit cards to buy their necessities due to technological advancements, which has led to a progressive increase in credit card fraud. In the modern world, credit cards are nearly universally accepted as a form of payment by businesses of all sizes. Every firm, including the banks, the automobile industry, the appliance industry, and others, suffers from credit card theft. Many techniques, including data mining, machine learning, and algorithmic methods, are used to spot fraud in credit card transactions, but they have not produced much of a payoff. Thus, it is necessary to build algorithms that are substantial in their effectiveness and efficiency.
By utilizing artificial neural network methods and comparing them to a few other machine learning techniques, we attempt to prevent the fraudster from using our credit card before the transaction is approved algorithms. Fraud is an offensive act that an uninvited person commits by deceiving innocent individuals. When someone uses a credit card fraudulently, the necessary login information from the cardholder is stolen and used by the fraudsters in an unlawful way, usually through phone calls or SMS messages. Certain software programs that are in the hands of fraudsters may potentially be used in this credit card fraud.
The process of detecting credit card fraud begins when the user or customer submits the required information to complete a credit card transaction. The transaction should only be allowed after being screened for fraud activity.

Paper Name: Credit Card Fraud Detection Using Random Forest Algorithm
Author: M.Suresh Kumar, V.Soundarya, S.Kavitha, E.S. Keerthika, E.Aswini
The use of fraudulent credit cards is rising daily. Fraudulent use of credit cards is possible both online and offline. Real cards are needed for offline transactions, whereas online transactions only need virtual cards for fraudulent or unlawful activity. Hence, these credit card fraud activities may result in several fraudulent transactions without the actual users' knowledge. In order to conduct transactions, fraudsters are searching for sensitive data such as credit card numbers, bank account information, and other personal specifics. In the case of offline transactions, fraudsters must steal the user's credit card to complete the transactions, while for online purchases, they must steal the user's identity and login credentials.
As a result, bank transactions are a huge problem in today's technology society, where credit card fraud has emerged as the main problem. Sensitive data is lost as a result of several fraudulent transactions that are difficult for both the user and the banking authority to detect. Based on the behaviour of the transactions, a variety of models are used to identify fraud transactions. These models can be divided into two main categories: supervised learning and unsupervised learning algorithms. They have employed techniques like Cluster Analysis, Support Vector

Machine, Naive Bayer's Classification, etc. in the current system to determine the accuracy of the fraudulent actions. This study uses the Random Forest Algorithm to determine the accuracy of fraudulent transactions.

Paper Name: Fraud Detection using Machine Learning and Deep Learning
Author: Pradheepan Raghavan, Neamat El Gayar
One of the most popular topics of this decade is machine learning, which is a subset of artificial intelligence. A growing number of businesses are looking to invest in machine learning to enhance their offerings. Machine learning combines numerous computer algorithms with statistical modelling to enable the computer to carry out tasks without having to be explicitly programmed. Learning would take place using the "training data" and the acquired model. The knowledge of past experiences can be used to make predictions or take actions. Artificial neural networks are used in machine learning techniques, which includes deep learning models. Other approaches include convolutional neural networks, deep belief networks, auto-encoders, recurrent neural networks, and restricted Boltzmann machines. A properly trained NN would be able to identify distinct relationships across the entire dataset. Three data sets will be examined in this research. The European dataset, the Australian dataset, and the German dataset are those three. We want to compare several ML and DL approaches in this paper. In all 3 datasets, an ensemble of the top 3 models is also used. Based on an empirical study comparing several ML and deep learning models, we report our findings.

Paper Name: Detecting Credit Card Fraud by ANN and Logistic Regression
Author: Y. Sahin, E. Duman
In this paper, a system for detecting credit card fraud is created using a variety of ANN and LR techniques. This system attempts to identify and mark transactions as legal or normal by monitoring each account independently and using appropriate descriptors. The identification will be based on the classifier models' created suspicion scores. The classifier can determine if a new transaction is legitimate or fraudulent when it begins. The transaction record's date and hour contain information about when the transaction was made. The transaction type identifies whether this is a buy or a cash advance. The MCC code identifies the kind of retailer where the transaction was made. These are set codes provided by the VISA International Service Association members.
Many of these codes do, however, form organic groups. Instead of dealing with hundreds of codes, we divided them into 25 groups based on their characteristics and likelihood of being used for fraud. In some MCC codes, it is simple to change the goods or services purchased from merchant stores into cash. Transactions involving certain MCC codes are therefore more vulnerable to fraud and more dangerous than transactions involving other MCC codes. The MCC codes are grouped based on the quantity of fraudulent transactions associated with each MCC code as well as interviews conducted with data provider bank staff members with subject-matter expertise.

PROPOSED SYSTEM

A. System Architecture

The architecture diagram shows us the following process of fraud detection and how we can detect it and be sure to each PAN card fraudulent generating it for their purpose.

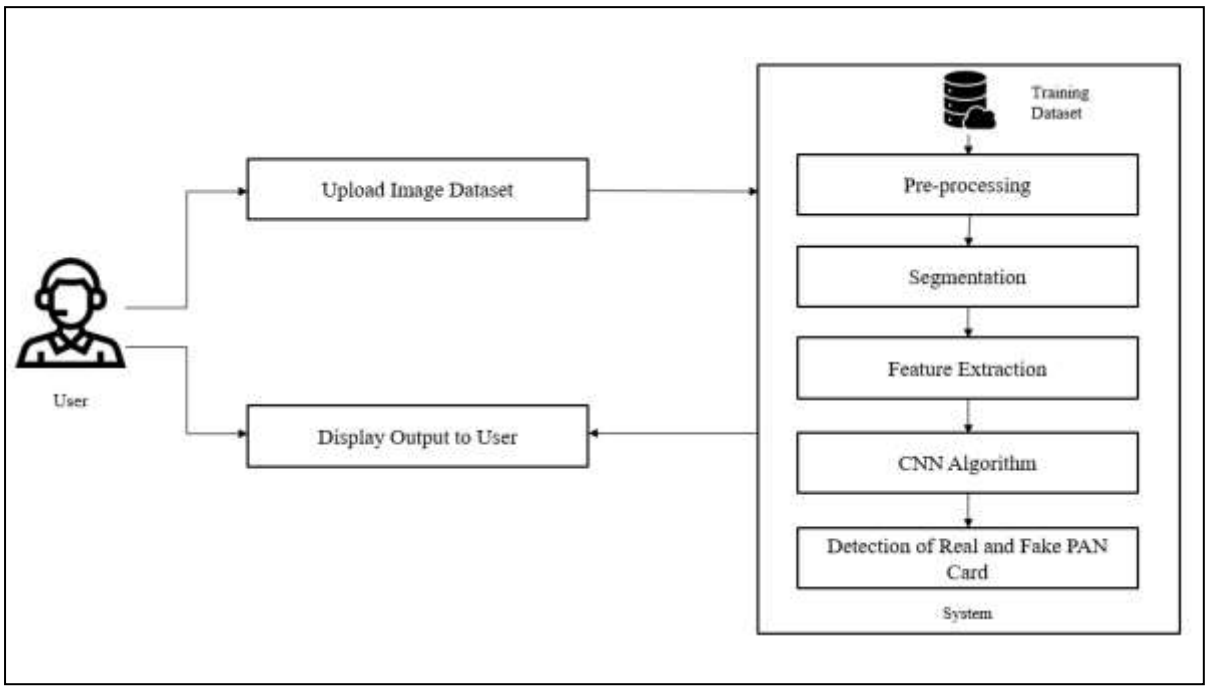

Fig.1 System Architecture

B. Methodology

This project can be implemented in following steps:

- Data Collection: -

1. First, we provide image dataset to the machine. Dataset is of images of PAN Card. We have a set of legitimate PAN Card image and a set of tampered PAN Card image. We must modify or prepare that dataset, for that next step is pre-processing.
2. Pre-processing: - In Pre-processing phase, in that removing the noisy and blur part of the dataset, and rescale, resize the image dataset.
3. After pre-processing of dataset, next phase is trained that dataset. For that, dataset goes through feature extraction classification.

- Train the dataset: - In this process we train the dataset by following steps:

1. Feature extraction: - In Feature extraction extract the features like edges, size etc. from dataset. Extract the features for classifications. After Feature Extraction next step is segmentation.
2. Segmentation: - In segmentation we divide image in multiple parts. Then after the all steps done, next phase is classification. We used classifier for the classification.
3. Classification: - We used CNN algorithm fort the classification. Classification is process of categorizing and labelling groups of pixels or vectors within an image based on specific rules. After all the training phase done Machine create a model i.e., trained model. It is 80% model.

C. Algorithm

CNN Algorithm: - Convolutional Neural Networks specialized for applications in image and video recognition. CNN is mostly utilized for image analysis tasks like segmentation and object detection. The Four types of layers in Convolutional Neural Networks are:

1. Convolutional Layer: - In a typical neural network each input neuron is connected to the next hidden layer. In CNN, only a small region of the input layer neurons connects to the neuron hidden layer.
2. Pooling Layer: - The pooling layer is used to reduce the dimensionality of the feature map. There will be multiple

activation pooling layers inside the hidden layer of the CNN.

3. Flatten: - Flattening is converting the data into a 1- dimensional array for inputting it to the next layer. We flatten the output of the convolutional layers to create a single long feature vector.
4. Fully-Connected layer: - Fully Connected Layers form the last few layers in the network. The input to the fully connected layer is the output from the final Pooling or Convolutional Layer, which is flattened and then fed into the fully connected layer.

ADVANTAGES

- Applicable for both low and high pixel images.
- It automatically detects the important features without any human supervision.

DISADVANTAGES

- Lot of training data is required to be effective.
- Sometime it is hard to classify images with different positions.
- If the layout of PAN card in future may get changed then it is hard to identify based on the older layout of the PAN card.

CONCLUSION

A system is proposed with the innovative method to solve the problem. Fraud detection for PAN card is the system based on CNN algorithm. Thus, an efficient and highly accurate system is proposed.

FUTURE SCOPE

A system is proposed with the innovative method to solve the problem. Fraud detection for PAN card is the system based on algorithm. Thus, an efficient and highly accurate system is proposed. This system can be used to detect future fraud detection on Aadhaar card, various other documents and useful to detect duplicate currency notes. This project can be implemented in different organizations where customers need to provide any kind of id in order to get themselves verified. With the help of this project each organization can find out whether the ID is original or fake. Similarly, this can be used for any type of ID like Aadhar, voter id, etc.

Declaration by Authors

Acknowledgement: The authors would like to thank the publishers and researchers for making their resource available and, we are thankful to our teachers for their guidance. We would like to express our gratitude to our guide Professor Rohini Hanchate for her encouraging support and guidance in carrying out this work. We express our sincere thanks to Nutan Maharashtra Institute of Engineering and Technology Pune for permitting us to take our work this further.

Source of Funding: None

Conflict of Interest: The authors declare no conflict of interest.

REFERENCES

1. Asha RB, Suresh Kumar KR, ”Credit card fraud detection using artificial neural network”,pp. 35–41, 2021, doi: https://doi.org/10.1016/j.gltp.2021.01.006.
2. M.Suresh Kumar, V.Soundarya, S.Kavitha, E.S. Keerthika, E.Aswini, ”Credit Card Fraud Detection Using Random Forest Algorithm”, 2019, doi: https://doi.org/10.1109/ICCCT2.2019.8824930
3. Pradheepan Raghavan, Neamat El Gayar, ”Fraud Detection using Machine Learning and Deep Learning”, December 2019, doi: https://doi.org/10.1109/ICCIKE47802.2019.9004231
4. Badal Soni, Pradip K. Das, Dalton Meitei Thounaojam. CMFD: a detailed review of block based and key feature based techniques in image copy-move forgery detection, 2018. IET Image Processing 12:2, pages 167- 178.
5. Francisco Cruz, Nicolas Sidere, Mickael Coustaty, Vincent Poulain ¨ D’Andecy, and Jean-Marc Ogier. Local binary patterns for document forgery detection. In Document Analysis and Recognition (ICDAR), 2017 14th IAPR International Conference on, volume 1, pages 1223– 1228. IEEE, 2017.

6. Y. Sahin, E. Duman, ”Detecting Credit Card Fraud by ANN and Logistic Regression”, 2011, doi: https://doi.org/10.1109/INISTA.2011.5946108
7. He, Zhiwei, et al. "A new automatic extraction method of container identity codes." IEEE Transactions on intelligent transportation systems 6.1 (2005): 72-78.
8. S. Shang, N. Memon, and X. Kong, “Detecting documents forged by printing and copying,” EURASIP Journal on Advances in Signal Processing, vol. 2014, no. 1, p. 140, 2014.
9. Ulutas, G., Muzaffer, G.: ‘A new copy move forgery detection method resistant to object removal with uniform background forgery’, Math. Probl. Eng., 2016, 2016, pp. 1–19
10. Bashar, M., Noda, K., Ohnishi, N., et al.: ‘Exploring duplicated regions in natural images’, IEEE Trans. Image Process., 2016, 99, pp. 1–40

How to cite this article: Rohini Hanchate, Shreyas Yerole, Nazir Lalloti et.al. PAN Card fraud detection using machine learning: a case study. *International Journal of Science & Healthcare Research.* 2023; 8(2): 214-219. DOI: *https://doi.org/10.52403/ijshr.20230225*

******