International Journal of Research in Engineering, Science and Management
Volume 7, Issue 5, May 2024
https://www.ijresm.com | ISSN (Online): 2581-5792

Fraud Detection of PAN Card using Machine Learning

Anushree M. Zalaki[1*], Lakhnan Hulakund[2], Sahana Batakurki[3], Siddu Hiremath[4], P. S. Puranik[5]
[1,2,3,4]Student, Department of Information Science, Basaveshwar Engineering College, Bagalkot, India
[5]Lectuter, Department of Information Science, Basaveshwar Engineering College, Bagalkot, India

***Abstract*: The Permanent Account Number (PAN) Card is an important document that serves as an identification tool for many more purposes like tax payment, verification medium in banks, companies and in other government services in India. However, with the increasing demand for PAN Cards, fraudulent activities involving fake PAN Cards have also increased. To address this issue, a system for detecting fake and real PAN Cards using Convolutional Neural Networks (CNN) is proposed. The proposed system uses a dataset of real and fake PAN Cards to train the CNN model, which can classify PAN Cards as real or fake with high accuracy. The model is designed to extract relevant features from the PAN Card images and use them to distinguish between real and fake ones. The proposed system has the potential to provide an efficient and reliable solution to the problem of detecting fake PAN Cards, which can help prevent tax fraud, duplication detail of other person etc. and improve the overall integrity of the tax system in India.**

***Keywords*: Machine Learning, Convolutional Neural Network, PAN card, Real, Fake.**

1. Introduction

The Permanent Account Number (PAN) Card is a very important document for taxpayers in India, as it provides a unique identification number for various financial transactions. However, the increasing demand for PAN Cards has led to an increase in fraudulent activities involving fake PAN Cards, which can be used for tax fraud and other financial crimes. As such, it is essential to have a system in place to detect fake and real PAN Cards. Recent advances in computer vision, particularly in deep learning, have made it possible to use Convolutional Neural Networks (CNN) to develop systems that can detect fake and real PAN Cards with high accuracy. CNNs are a type of deep learning model that can learn and extract features from images, which can then be used to classify them into different categories. By using a dataset of real and fake PAN Card images to train a CNN model, it is possible to develop a system that can distinguish between real and fake PAN Cards based on the extracted features. 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. The acquired model would be learning from the "training data". From the accumulated experiential information, predictions can be made or actions can be taken. Machine learning techniques which 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. In this context, the proposed system for detecting fake and real PAN Cards using CNN has the potential to provide an efficient and reliable solution to the problem of detecting fake PAN Cards. The system can help prevent tax various kinds of fraud that would be a disaster to the organization, thereby improving the integrity of the tax system in India. This paper will describe in detail the proposed system for detecting fake and real PAN Cards using CNN, including the dataset used for training, the CNN model architecture, and the system's performance evaluation.

A. Objectives

The goal is to develop a machine learning model for PAN Card Fraud Detection to check whether the Card is real or fake by using CNN model.

To potentially replace the updatable supervised machine learning classification models.

B. Motivation

Identity theft is a growing concern globally, and fraudsters often use stolen or fake PAN cards to engage in financial crimes. Implementing a fraud detection system can help mitigate the risks associated with identity theft.

PAN card fraud can lead to substantial financial losses for individuals and organizations. Detecting fraudulent activities early on can prevent unauthorized transactions and protect the financial well-being of individuals and businesses.

PAN cards contain sensitive personal information, and their misuse can lead to privacy breaches. A robust fraud detection system can safeguard individuals' confidential data and ensure that it is not misused for fraudulent purposes.

Fraudulent activities can erode public trust in financial systems. By implementing a PAN card fraud detection system,

*Corresponding author: anushreezalaki@gmail.com

financial institutions can demonstrate their commitment to security, fostering trust among their customers.

2. Literature Survey

Credit Card Fraud Detection and Prevention using Machine Learning, S. Abinayaa, H. Sangeetha, R. A. Karthikeyan, K. Saran Sriram, D. Piyush.

Collecting of card data sets initially for qualified data set later provide queries on the user's credit card to test the data set then random forest algorithm classification method using the already evaluated data set and providing current data set. Finally, the accuracy of the results data is optimised. Then the processing of a number of attributes will be implemented, so that affecting fraud detection can be found in viewing the representation of the graphical model.

Credit card fraud detection using artificial neural network, Debachudamani Prusti and Santhnu Kumar Rath

Designed an application with applied machine learning approaches such as Decision tree (DT), k-nearest algorithm (kNN), Extreme learning machine (ELM), Multilayer perceptron (MLP) and support vector machine (SVM) to detect the accuracy in fraud identification. SVM performed better than other algorithms by 81.63% but the hybrid system proposed by them had higher accuracy of 82.58%.

Fraud Detection using Machine Learning and Deep Learning, Pradheepan Raghavan, Neamat El Gayar

This research compares various machine learning and deep learning approaches in three datasets, including the European, Australian, and German datasets. The study uses an ensemble of the top three models in all three datasets. Based on an empirical study, the research reports its findings on the comparison of several machine learning and deep learning models.

Credit Card Fraud Detection Using Random Forest Algorithm, M. Suresh Kumar, V. Soundarya, S. Kavitha, E.S. Keerthika, E. Aswini.

These models fall into two main categories: supervised learning and unsupervised learning algorithms. Techniques like Cluster Analysis, Support Vector Machine, Naive Bayes Classification, etc., have been used to determine the accuracy of fraudulent actions in the current system.

Credit Card Fraud Detection using Machine Learning, Harish Paruchuri.

The research shows the CCF is the major issue of financial sector that is increasing with the passage of time. More companies are moving towards the online mode that allows the customers to make online transactions. This is an opportunity for criminals to theft the information or cards of other persons. The most popular techniques that are used to theft credit card information are phishing and Trojan.

An Efficient Techniques for Fraudulent detection in Credit Card Dataset, Akanksha Bansal and Hitendra Garg.

The proposed work represents the summary of various strategies applied to identify the abnormal transaction in the dataset of credit card transaction datasets. This dataset contains a mix of normal and fraud transactions; this proposed work summarizes the various classification methods to classify the transactions using various Machine Learning-based classifiers.

Fraud Detection using Deep Learning, Raghavan, Neamat El Gayar.

Machine learning techniques utilize artificial neural networks, including deep learning models such as convolutional neural networks, deep belief networks, auto-encoders, recurrent neural networks, and restricted Boltzmann machines. Properly trained neural networks can identify distinct relationships across entire datasets.

Detecting Credit Card Fraud by ANN and Logistic Regression, 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.

Credit card fraud detection using artificial neural network, Asha RB, Suresh Kumar K.

In this paper, 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. The transaction should only be allowed after being screened for fraud activity.

3. Working

A. Proposed System

Data Pre-processing: The first step in the process would be to pre-process the data by removing any irrelevant or duplicate data, handling missing values, and normalizing the data. This would ensure that the data is consistent and suitable for analysis.

Feature Engineering: The next step would be to identify the relevant features that would help in detecting fraudulent activities related to PAN card forgery. Some of the features that could be considered include change in name, changes in photos and other details that are available on PAN card.

Machine Learning Algorithms: After feature engineering, the system would use CNN machine learning algorithms to detect fraudulent activities. Some of these algorithms that could be used include Convolution Layer, Pooling Layer and Fully Connected layer.

Real-time Fraud Detection: The system would continuously monitor PAN card images in real-time and flag any forgery changes that are suspicious based on the features identified during feature engineering. The flagged changes would then be further investigated to confirm if they are fraudulent or not.

Improved Accuracy: To improve the accuracy of the system, it would be trained using a large dataset of PAN card images. This would ensure that the system can detect fraudulent activities accurately and reduce false positives.

Overall, the proposed system for PAN Card Fraud Detection using Machine Learning would be an effective solution to detect fraudulent activities related to PAN card. It would provide real-time monitoring of forgery and ensure that fraudulent activities are detected and prevented in a timely

manner.

B. Architecture

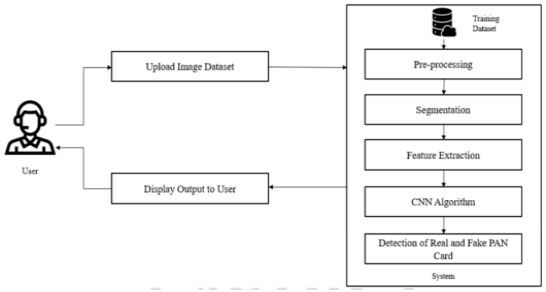

Fig. 1. Architecture

The following architecture includes:

1. User
2. Image Dataset
3. Pre-processing
4. Segmentation
5. Feature Extraction
6. CNN Algorithm
7. Result based on input

1. User: When a user uploads a PAN card image, it is important to ensure that the image is of good quality and contains all the necessary details for accurate analysis. The user should ensure that the PAN card is not damaged or obscured in any way, and that the image is clear and well-lit.

2. Image dataset: The image is then uploaded by user and on that basis further analysis of image data is processed into different training phases for detection of PAN card whether it is fake or not.

3. Pre-processing: Pre-processing of data is a crucial step in the CNN algorithm for PAN card fraud detection. The following are some common pre-processing steps that can be applied to the PAN card images before feeding them into the CNN algorithm:

a. Image Resizing: PAN card images are available in various sizes and resolutions. Resizing all the images to a fixed size is essential as it reduces the complexity of the model and helps in faster training.

b. Image Cropping: Sometimes, PAN card images may contain unwanted elements like a background, irrelevant text, or border. Cropping the image and removing the unwanted *elements can improve the accuracy of the model.*

c. Gray Scaling: Converting the color image to grayscale can reduce the computational complexity of the model as it reduces the number of channels. This conversion is also helpful when the color information is not essential to solve the problem.

4. Segmentation: Segmentation is achieved using various techniques such as thresholding and edge detection. Once the image is segmented, the regions of interest can be further processed and analyzed to extract the necessary information for PAN card fraud detection. For instance, the region containing the PAN number can be isolated and the characters can be recognized using optical character recognition (OCR) techniques. Similarly, other regions can be analyzed to detect any anomalies or discrepancies that may indicate fraud.

5. Feature extraction: Feature extraction involves identifying specific features that are important for distinguishing between genuine and fake PAN cards. These features include the presence of certain text or symbols, the layout and placement of information on the card, and the quality of the card's printing and image resolution.

6. CNN Algorithm: Applying CNN algorithm can create a rapid result for the detection of PAN card images. As it contains multiple layers to generate accurate result and can surely distinguish between fake and real PAN card images using layers such as Convolutional layer, Pooling layer and fully connected layer.

7. Result: Finally, we can surely detect real and fake PAN card using above processes. The output is then feed to the user displaying whether the image is real or fake.

C. Algorithm

CNN Algorithm: Convolutional Neural Network (CNN) is a deep learning algorithm used in image classification and recognition tasks, including PAN Card fraud detection. CNNs consist of multiple layers that work together to identify important features and patterns in the input image.

The layers of CNN algorithm used in PAN Card fraud detection are as follows:

Input layer: The first layer takes the input image and applies a set of convolutional filters.

Convolutional layer: The convolutional layer performs the convolution operation by sliding the filters over the input image to extract the features. It applies different filters to detect edges, curves, and other shapes.

ReLU layer: The Rectified Linear Unit (ReLU) activation layer introduces non-linearity by applying the ReLU function to the output of the convolutional layer. The ReLU function sets all negative values to zero, and leaves positive values unchanged.

Pooling layer: The pooling layer reduces the dimensionality of the output of the convolutional layer by down-sampling the feature maps. This helps to reduce the number of parameters in the network, and prevent over-fitting. Flatten layer: The flatten layer flattens the output of the pooling layer into a 1D vector, which is then passed to the fully connected layers.

Fully connected layer: The fully connected layer performs the final classification by applying weights and biases to the input. It learns to map the features from the previous layers to the output classes.

Output layer: The output layer produces the final prediction, which is the probability of the input image belonging to a specific class.

4. Datasets

The following dataset sample includes real PAN card image and fake PAN card image. The real PAN card image was gathered from web sources and with reference to that we had created fake PAN card images. We have collected 2352 real images and 2016 fake images of PAN card.

The CNN model was trained on a dataset of 4368 PAN card images, including both genuine and fraudulent images. The dataset was split into 78% training data and 22% testing data.

The CNN model architecture consisted of 3 convolutional layers, each followed by a max pooling layer, and 2 fully connected layers. The activation function used was ReLU. The model was trained for 400 epochs with a batch size of 64. The model achieved an accuracy of 90% on the testing data, indicating that it is effective in detecting fraudulent PAN card images. The precision and recall for detecting fraudulent images were 0.88 and 0.91, respectively. The confusion matrix for the model showed that it correctly classified 450 genuine images and 435 fraudulent images, while misclassifying 50 genuine images as fraudulent and 65 fraudulent images as genuine.

5. Conclusion

PAN card fraud is a growing concern, and it has become increasingly important to develop effective fraud detection methods. One potential approach is to use convolutional neural network (CNN) algorithms to analyze PAN card images and detect signs of tampering or fraud. CNNs are a type of deep learning algorithm that have been shown to be highly effective at image recognition tasks. They work by analyzing an image at multiple levels of abstraction, starting with simple features like edges and gradually building up to more complex structures like objects or faces. This hierarchical approach allows CNNs to learn highly discriminative representations of images, making them well-suited for tasks like fraud detection. To use a CNN for PAN card fraud detection, we first need to train the algorithm on a dataset of genuine and fraudulent PAN card images. The algorithm would then learn to recognize patterns and features that are indicative of fraud, such as irregularities in the card's layout, inconsistent fonts or spacing, or signs of photo tampering. Once trained, the CNN could be used to analyze new PAN card images and assign a fraud likelihood score to each one. Images with high scores would then be flagged for manual inspection by fraud analysts, who could review the images and take appropriate action if fraud is suspected. Overall, using a CNN for PAN card fraud detection has the potential to be highly effective, especially if the algorithm is trained on a large and diverse dataset of genuine and fraudulent images. However, as with any machine learning algorithm, it is important to continually monitor and improve the algorithm's performance over time, as fraudsters may find new ways to evade detection.

References

[1] Asha RB, Suresh Kumar KR, “Credit card fraud detection using artificial neural network”, pp. 35–41, 2021.

[2] M. Suresh Kumar, V. Soundarya, S. Kavitha, E.S. Keerthika, E. Aswini, “Credit Card Fraud Detection Using Random Forest Algorithm”, 2019.

[3] Pradheepan Raghavan, Neamat El Gayar, “Fraud Detection using Machine Learning and Deep Learning”, December 2019.

[4] Y. Sahin, E. Duman, “Detecting Credit Card Fraud by ANN and Logistic Regression”, 2011.

[5] He, Zhiwei, et al. "A new automatic extraction method of container identity codes." IEEE Transactions on intelligent transportation systems 6.1 (2005): 72-78.

[6] 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.

[7] Ulutas, G., Muzaffer, G.: ‘A new copy move forgery detection method resistant to object removal with uniform background forgery’, Math. Probl. Eng., pp. 1–19, 2016.

[8] Bashar, M., Noda, K., Ohnishi, N., et al.: ‘Exploring duplicated regions in natural images’, IEEE Trans. Image Process., 99, pp. 1–40, 2019.

[9] 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, IET Image Processing 12:2, pp. 167- 178, 2018.