PAN Card Fraud Detection Using Machine Learning

Anju Venkiteswaran Bhawana Thawarani

Rahul Trivedi

Abstract

The goal of this project is to utilize computer vision to detect tampering of PAN cards. It will assist organizations in verifying the authenticity of PAN cards submitted by employees, customers, or other individuals, ensuring the documents are genuine. For this project we will calculate structural similarity of original PAN card and the PAN card uploaded by user. We will also try to build a DeepLearning model to predict the fraud pancard.

1. 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.

1.1. Methodology

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: algorithm - fort We used CNN 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

1.2. Algorithm

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 pool-

ing 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.

1.3. ADVANTAGES

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

1.4. 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.

1.5. 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.

1.6. 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.

1.7. 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/ICCCT2.2019.88249 30 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

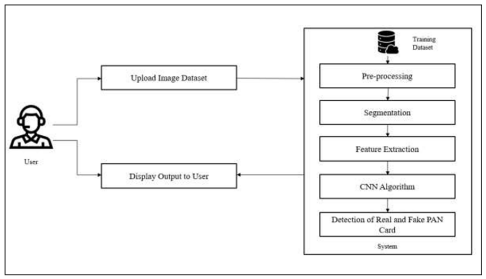

Figure 1. System Architecture