A Review: Fake Aadhar Card Detection Using Machine Learning

[1]Sakshi Devane, [2]Dr.Sandeep Chaware, [3]Shreni Pandit, [4]Manasi Kshirsagar, [5]Leena khairnar
[1]Department of Information Technology
[1]JSPM's Rajarshi Shahu College of Engineering, Tathawade, Pune, India.

sakshidevane33@gmail.com, shrenipandit14@gmail.com, manasikshirsagar19@gmail.com

Abstract –The idea behind Aadhaar was the requirement that each person have a distinct identity. The Indian government established the UIDAI distribution authority to carry out this and create user identities for each person using their biometric and demographic information. In the current situation, people are manipulating documents, abusing people's identification documents and photos in illegal ways, and several web portals demand time-consuming repetitive Aadhaar card submissions for different objectives. Therefore, the system assists in identifying such modified photos and documents and helps to decrease the number of papers that are submitted repeatedly in order to prevent such illegal use. The issue with the current approach is that it uses metadata from the specific document that has been altered by specific manipulation techniques, including splicing and coloring, utilizing basic editing tools like Photopea, Paint.NET, etc., to identify manipulated documents and photos.

Keywords: Biometric information, Demographic information, Identity verification, Convolutional Neural Networks (CNNs), Image processing.

I. INTRODUCTION

In order to verify the legitimacy of identity papers like Aadhaar cards, the suggested method uses Convolutional Neural Networks (CNNs) and image processing techniques to identify modified Aadhaar card images. Forged documents are frequently used illegally due to the increase in identity fraud. By detecting digitally altered Aadhaar cards and stopping unauthorized document usage, this technology addresses such fraudulent activities.

Nowadays, many people alter text fields, unique ID numbers, fonts, and background aspects in identity documents using programs like Adobe Photoshop and Photopea, making phony Aadhaar cards virtually identical to authentic ones. Often posted to several web portals for authentication, these changed photographs are hard to spot with a human eye. In addition, customers frequently have to deal with the inconvenience of constantly presenting their identity documents across many sites, which is ineffective and time-consuming.

Software that deletes or modifies metadata can be used to get around the metadata analysis i.e. the mainstay of traditional document verification systems. By using CNN-based deep learning models and Error Level Analysis (ELA) to identify image inconsistencies, the suggested approach improves security. The technology correctly determines if Aadhaar photos are real or false by examining textural changes, compression artifacts, and hidden manipulation evidence.

TABLE 1: Literature Review

Reference	Approach	Key Findings	Performance	Impact
[1]	Matching image with Aadhaar data with FRT	Enhances security, reducing fraud, and surveillance.	Effective in law enforcement, e-KYC, and access control	Strengthens authentication and reduces the risk for fraud.
[2]	CNN method to identify counterfeit PAN cards.	Identifies damaged PAN images and distinguishes authentic ones from altered versions.	Improve fraud detection rates while reducing the need for manual verification.	Enhances the security of document verification processes.
[3]	applies SVM RF and KNN classifiers with LBP, HOG, and LPQ to extract features.	The combined approach (LBP+HOG+LPQ) achieves 88% accuracy (RF, self-created dataset) and 78% accuracy (SVM, DEFECTO dataset)	Enhances detection accuracy for altered Aadhaar images.	Makes identity checking and fraud stopping systems stronger.
[4]	Using biometric and demographic data using NB, C4.5 Decision Tree, and Back Propagation.	Identify anomalies and avoiding identity duplication.	Correctly detects phoney Aadhaar cards	Enhanced identify verification systems based on aadhaar
[5]	Ensure PAN Card using deep learning, AI and image recognition.	High accuracy real-time detection.	Enhance the detection accuracy.	Increease tax compliance and financial security.
[6]	Uses LOF and Isolation Forest for anomaly detection.	effectively distinguishes fraudulent from legitimate transactions.	Reduces false positives with improved model accuracy.	Enhances fraud detection with big datasets and advanced techniques.
[7]	Credit card fraud detection using GA algorithm	The model effectively distinguishes between fraudulent and legitimate transactions.	enhance fraud detection accuracy through data preprocessing, visualization, and feature selection.	Improves fraud detection while preserving scalability and adaptability.
[8]	Annotated Aadhaar images are provided for ML training using a dataset-based methodology.	permits Aadhaar details to be detected, extracted, and verified for machine learning tasks.	excellent labelled photos that can be used for forgery detection and classification.	solid basis for developing models that distinguish between authentic and fraudulent Aadhaar.
[9]	Text extraction for identity documents using OCR.	Dependable text element extraction from documents, such as Aadhaar cards.	Consistent OCR results for a range of fonts and image types.	Essential for confirming the accuracy of extracted text in order to detect tampering.
[10]	A method for deep learning biometric identity verification.	Demonstrates efficient biometric recognition based on competition.	Scalable approach with high recognition accuracy.	Connects to security and identity verification to support ML justification.

III. METHODOLOGY

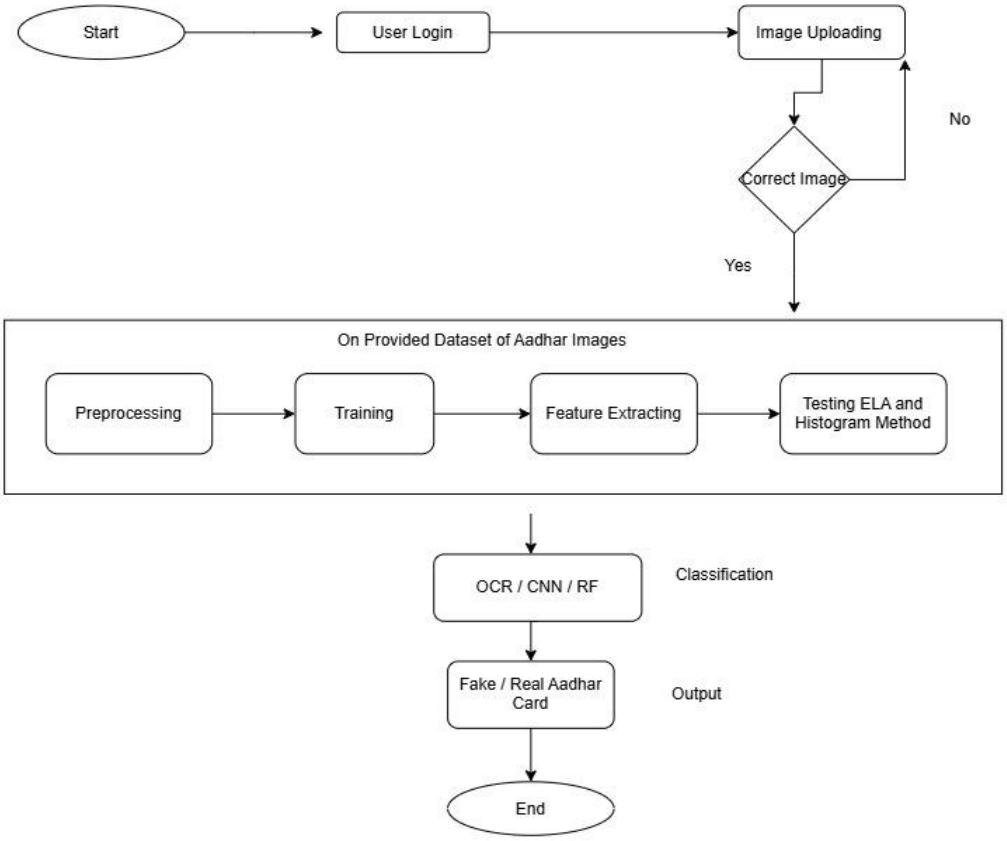

Fig.1. System Framework

This project can be implemented in the following steps:

- User Login:

When a user enters the system, he/she needs to enter the login credentials such as username and password for authorized access.

- Uploading Image:

The user has an option to upload an image of the Aadhaar card. After uploading the image, the system checks for file type, resolution, and clarity before proceeding.

- Correct Image:

It is the decision step of the proposed system. The system evaluates whether the uploaded image meets the required conditions or not.

If the Image is blurry, cropped, or low-quality images i.e. the image is incorrect, then the system asks for a reuploading the image.
If the image is correct i.e. not blurry, the process moves to the next step.

- Dataset:

The dataset used for the Fake Aadhaar Card Detection project consists of a total of 4,000 images, categorized into two classes: Fake and Real. To ensure reliable model training and performance evaluation, the 2,000 images in each class are further split into training and testing subsets.

fake Class:1,500 images in a training set
Testing Collection: 500 images
True Class:1,500 images in a training set
Testing Collection: 500 images

- Preprocessing:

From fig. 1 Preprocessing is carried out on the given dataset of Aadhaar Card photos, which includes both authentic and fraudulent photographs. By emphasizing language and security features, the image is improved in this step to eliminate noise, fix distortions, and make it easier to read.

- Training:

A labeled dataset of actual and false Aadhaar card photographs is used to train a machine learning model to identify various image patterns.

- Features Extraction:

In order to differentiate between authentic and fraudulent Aadhaar cards, this process takes crucial information from the image, such as edges, texture, size, and other distinctive components.

- Testing:

Using the below two primary techniques, the authenticity of the Aadhaar card is validated:

Error Level Analysis (ELA):

In this testing it checks for manipulated areas in the images. Fake Aadhaar cards may have inconsistent compression levels, which is highlighted by this method.

Histogram Method

In this method it analyzes the color distribution and pixel intensity of the image. Fake cards may have inconsistent color gradients due to alterations of image.

- Classification:

The extracted features are passed to a classifier for final decision-making.

Algorithms Used for Classification:

OCR (Optical Character Recognition): Unlike authentic UIDAI-issued cards, fake Aadhaar cards may have improper typefaces or alignments. By comparing with the official Aadhaar template, OCR finds font discrepancies.

OCR is able to read text from the Aadhaar QR code and compare it with the information on a printed card. The Aadhaar card can be fraudulent if the scanned text and the content of the QR code are different.

CNN: It uses image-based characteristics to identify Aadhar card tampering and fraud. CNN automatically identifies patterns on the card, including missing security features, pixel irregularities, text distortions, and changes made to the QR code. It can detect irregularities that point to fraud by training on a dataset of authentic and fraudulent Aadhaar cards. High accuracy, resilience, and real-time categorization for authentication purposes are guaranteed by their capacity to generalize across various Aadhaar card variations.

Random Forest: The submitted image is classified as authentic or fake using the Random Forest algorithm on the labeled dataset. In order to increase accuracy and resilience, it builds several decision trees during training and combines their outputs. It can deal with high-dimensional data, which lessens the likelihood of overfitting. It can also Can also classify both seen and unseen Aadhaar cards correctly.

- Output:

The system displays the final classification result that the uploaded image of Aadhar Card is Real Or fake based on the machine learning model's prediction.

IV. CONCLUSION:

The proposed system enhances Aadhaar card verification by combining techniques like CNN, OCR and Random Forest. It effectively identifies manipulated images by analyzing textural changes, altered text fields, and layout inconsistencies.

The integration of OCR ensures accurate text recognition and cross-verification with QR code data, improving fraud detection. Overall, it offers a reliable and efficient solution for safeguarding Aadhaar-based identity verification processes.

V. ACKNOWLEDGMENT:

The authors sincerely thank Dr. Sandeep Chaware for his invaluable guidance, encouragement, and support during this work. We also appreciate the Department of Information Technology at JSPM's Rajarshi Shahu College of Engineering for providing the necessary infrastructure and academic support. Finally, we thank all researchers whose studies have helped shape this review.

VI. REFERENCES:

[1] Dr. G. Simi Margarat, Dr. K. Ravikumar, Dr. S. Brittoraj, Dr. Bsant "An Aadhaar Authentication Application Using a Face Recognition System and Verification for Identifying" UGC Care Journal (2023)

[2] Rohini Hanchate, Shreyas Yerole, Nazir Lalloti, Piyush Mahajan: PAN Card Fraud Detection Using Machine Learning. International Journal of Science and Healthcare Research Vol. 8; Issue: 2;(April-June 2023)

[3] Aziz Makandar, Syeda Bibi Javeriya: Enhancing Aadhar Card Image Security with Machine Learning-Based Face Morphing Detection. Conference Paper · May 2024

[4] K. Ramya, A. Sumathi "Big Data Applications in Aadhar Card Fraud Detection" International Journal of Research Publication and Reviews (2019)

[5] Kshitij Pareek, Dr. Srikanth V Pan Guard: Fake PAN CARD Spotting using deep learning. International Journal of Research Publication and Reviews, Vol 5, no 3, pp 1342-1344 (March 2024)

[6]S P Maniraj, Aditya Saini, Swarna Deep Sarkar, Shadab Ahmed, "Credit Card Fraud Detection using Machine Learning and data Sscience".DOI:10.17577/IJERTV8IS09003 (2019)

[7] Emmanuel Ileberi, Yanxia Sun and Zenghui Wang "A machine learning based credit card fraud detection using the GA algorithm for feature selection" Journal of Big Data (2022) doi.org/10.1186/s40537-022-00573-8 (2022)

[8] (2024). Aadhaar Card Details Extraction Dataset. Available: https://universe.roboflow.com/ akash-k-p-gs9iu/aadhaar-card-details-extraction/dataset/6

[9] R. Smith, "An overview of the tesseract OCR engine," in Proc. 9th Int. Conf. Document Anal. Recognit. (ICDAR), vol. 2, Sep. 2007, pp. 629–633.

[10] Z. Yang, H. Huangfu, L. Leng, B. Zhang, A. B. J. Teoh, and Y. Zhang, "Comprehensive competition mechanism in palmprint recognition," IEEE Trans. Inf. Forensics Security, vol. 18, pp. 5160–5170, 2023.