IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Fake Pancard Detection Using Machine Learning

[1]S Ramesh Babu, [2]P Tharun, [3]B Sandeep,[4]L Prabhu Kiran

[1]Assistant Professor, [2]Student, [3]Student,[4]Student

[1]Electronics And Communication Engineering (ECE),

[1]G. Pulla Reddy Engineering College, Kurnool, India

Abstract: The Permanent Account Number (PAN) Card is a vital identity document in India, primarily used for tax purposes, but also serving as a verification tool in banks, corporations, and various government services. However, with the rising demand for PAN Cards, there has been a significant increase in fraudulent activities involving counterfeit PAN Cards. To mitigate this issue, a system leveraging Convolutional Neural Networks (CNN) is proposed for the detection of fake PAN Cards. The proposed model is trained on a dataset comprising both genuine and forged PAN Card images, enabling it to accurately classify them as real or fake. By extracting key visual features through convolutional layers, the system learns to differentiate between authentic and counterfeit cards effectively. This approach offers a robust and scalable solution to combating PAN Card fraud, thereby enhancing the integrity of identity verification processes and strengthening the reliability of the tax infrastructure in India.

Keywords: Machine Learning, Convolution Neural Networks, Real, Fake, Image Processing.

I. INTRODUCTION

The Permanent Account Number (PAN) card is a vital document issued by the Government of India, serving as a unique identifier for individuals and entities engaging in financial transactions. It plays a crucial role in tax filing, banking operations, and identity verification. However, the increasing dependence on PAN cards has led to a rise in fraudulent activities, where counterfeit PAN cards are used for tax evasion, money laundering, and other financial crimes. Detecting such fraud is essential to maintaining financial security and preventing economic losses.

With the advancements in artificial intelligence and machine learning, deep learning techniques have shown promising results in automating fraud detection. In particular, **Convolutional Neural Networks (CNNs)**—a specialized form of deep learning—have proven highly effective in image classification and pattern recognition. CNNs can analyze the visual features of PAN cards, distinguishing between genuine and fraudulent documents based on learned patterns. By leveraging a dataset containing both real and fake PAN card images, a CNN-based model can be trained to accurately classify PAN cards as authentic or counterfeit.

Machine learning, a subset of artificial intelligence, has revolutionized various industries by enabling computers to learn from data without explicit programming. Deep learning, an advanced branch of machine learning, utilizes artificial neural networks to process complex patterns and relationships. Several deep learning techniques, including **autoencoders, restricted Boltzmann machines, deep belief networks, and recurrent neural networks**, have been explored for fraud detection. Among these, CNNs stand out for their superior ability to analyze image-based data. The increasing cases of PAN card fraud necessitate the development of an **efficient and automated detection system**.

The proposed system employs a CNN model to detect fake PAN cards by analyzing key visual features. The system is designed to:

1. **Collect and preprocess a dataset** of real and fake PAN card images.
2. **Train a CNN model** to recognize unique patterns distinguishing genuine from counterfeit cards.
3. **Evaluate the model's accuracy** in identifying fake PAN cards using various performance metrics.

By integrating machine learning into fraud detection, the proposed system enhances security measures in financial transactions and prevents the misuse of PAN cards. This paper provides a detailed overview of the **dataset, model architecture, and performance evaluation** of the CNN-based PAN card detection system, demonstrating its effectiveness in mitigating fraud.

II. REVIEW OF LITERATURE

Paper Name: Credit card fraud detection using artificial neural network
Authors: Asha RB, Suresh Kumar K
This study highlights the increasing threat of credit card fraud across various industries. Traditional fraud detection techniques, including data mining and algorithmic methods, have shown limited success. The authors propose artificial neural networks (ANN) as an alternative and compare their effectiveness with other machine learning techniques. The ANN model, using multiple hidden layers and the ReLU activation function, demonstrated superior accuracy in distinguishing fraudulent transactions.

Paper Name: Credit Card Fraud Detection Using Random Forest Algorithm
Authors: M. Suresh Kumar, V. Soundarya, S. Kavitha, E.S. Keerthika, E. Aswini
This research focuses on both online and offline credit card fraud detection using machine learning techniques. Fraudsters often steal sensitive user data for unauthorized transactions. Various classification methods, such as Support Vector Machines (SVM) and Naïve Bayes, have been explored in fraud detection. The study employs the Random Forest Algorithm, demonstrating its ability to classify fraudulent transactions efficiently by analyzing transaction behavior patterns.

Paper Name: Fraud Detection Using Machine Learning and Deep Learning
Authors: Pradheepan Raghavan, Neamat El Gayar
This research provides a comparative analysis of various machine learning and deep learning techniques for fraud detection across multiple datasets (European, Australian, and German). The study incorporates deep learning models such as convolutional neural networks (CNN), deep belief networks (DBN), and autoencoders. The findings indicate that an ensemble approach using the top-performing models yields improved fraud detection accuracy.

Paper Name: Detecting Credit Card Fraud Using ANN and Logistic Regression
Authors: Y. Sahin, E. Duman
This paper proposes a fraud detection system utilizing artificial neural networks (ANN) and logistic regression (LR). Each account is monitored individually, and transactions are classified as fraudulent or legitimate based on a suspicion score. The study also incorporates Merchant Category Codes (MCC) to group transactions based on risk levels. The results demonstrate that certain MCC categories are more prone to fraudulent activities, aiding in better fraud prediction.

III. METHODOLOGY

3.1 Overview of the Proposed Approach

The proposed system for PAN Card Fraud Detection utilizes machine learning and deep learning techniques to analyze and detect fraudulent activities related to PAN card forgery. The system is designed to identify discrepancies in PAN card details by leveraging image processing, feature extraction, and Convolutional Neural Networks (CNNs).

Key Components of the Proposed System

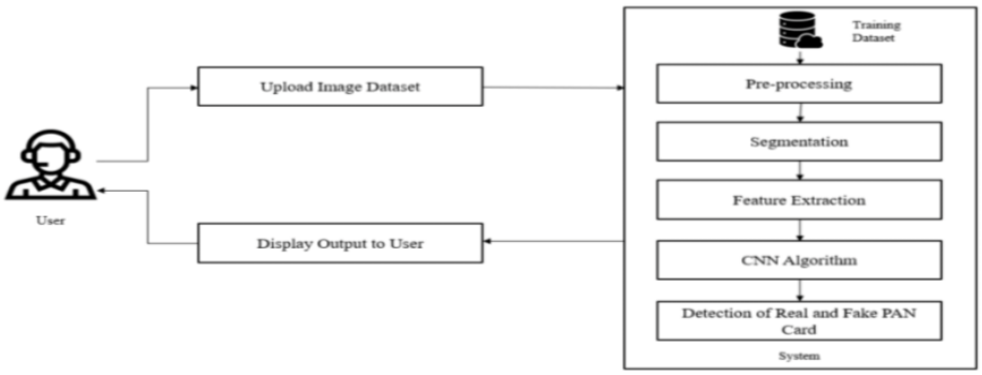

Fig 1. BLOCK DIAGRAM OF PROPOSED SYSTEM

3.2.1 Data Preprocessing

To enhance the accuracy of fraud detection, the system performs data preprocessing, which includes:

- Removing redundant or irrelevant data to ensure dataset efficiency**.**
- Handling missing values to maintain dataset completeness**.**
- Normalizing image and textual data for consistent processing.

3.2.2 Image Dataset and Processing

Once the image is uploaded, it is stored in a dataset for further processing. The dataset is used for training and testing a deep learning model to distinguish between real and fake PAN cards.

3.2.3 Pre-Processing Techniques

Pre-processing enhances image quality and prepares it for analysis using a Convolutional Neural Network (CNN). The key steps include:

- **Resizing**: Standardizing image dimensions to ensure uniformity and improve model efficiency.
- **Cropping**: Removing unnecessary background elements, irrelevant text, or borders to focus on key features.

3.2.4 Image Segmentation

Segmentation techniques such as thresholding and edge detection isolate important sections of the PAN card. Key regions, like the PAN number, name, and photograph, are extracted and processed for further validation. Optical Character Recognition (OCR) is used to identify and verify textual information.

3.2.5 Feature Extraction

Feature extraction focuses on identifying key attributes that indicate fraudulent modifications in PAN card images. Some of the critical features include**:**

- Alterations in personal details (e.g., name, date of birth, signature).
- Changes in PAN card photograph (facial mismatches, photo tampering).
- Text inconsistencies (modification in PAN number or address details).

3.2.6 Machine Learning Model – CNN Approach

The system employs Convolutional Neural Networks (CNNs) for fraud detection, consisting of the following layers:

- **Convolutional layer:** Extracts image features for pattern recognition.
- **Pooling Layer:** Reduces dimensionality while preserving essential information.
- **Fully Connected layer:** Classifies the image as genuine or fraudulent.

ALGORITHM

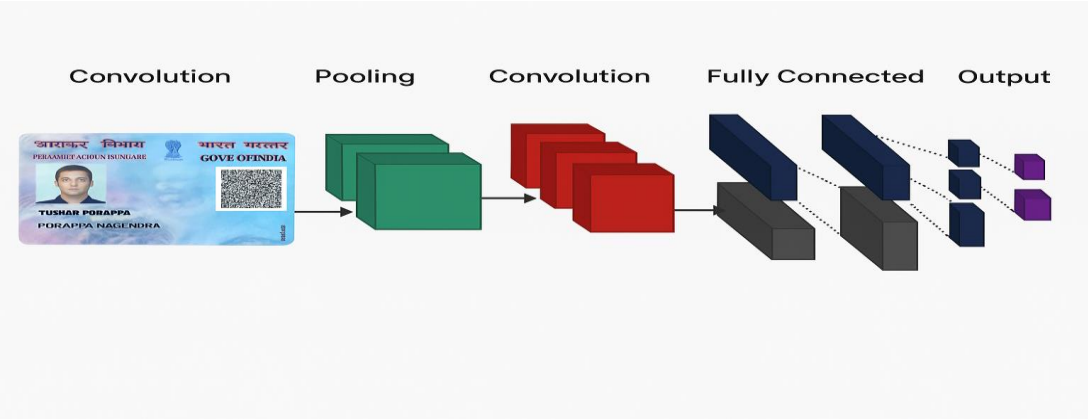

Fig 2. Layers in CNN Algorithm

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 to it.

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.

3.2.7 Datasets

The 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 some real images and fake images of PAN card.

Fig 3. Sample Pan Cards

IV. RESULTS AND DISCUSSION

4.1 Model Accuracy

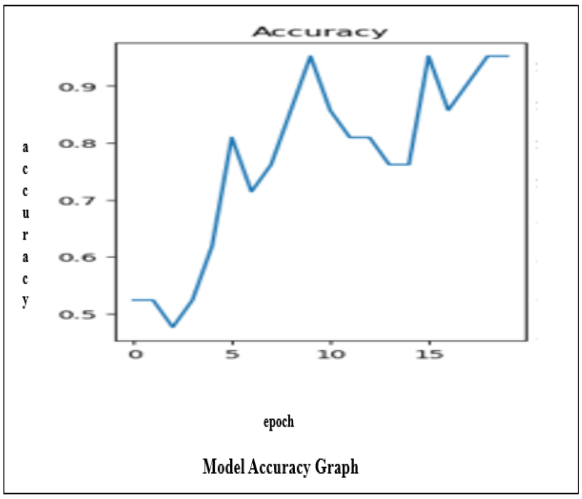

Fig 4.1 Model Accuracy Graph

4.2 Model Loss

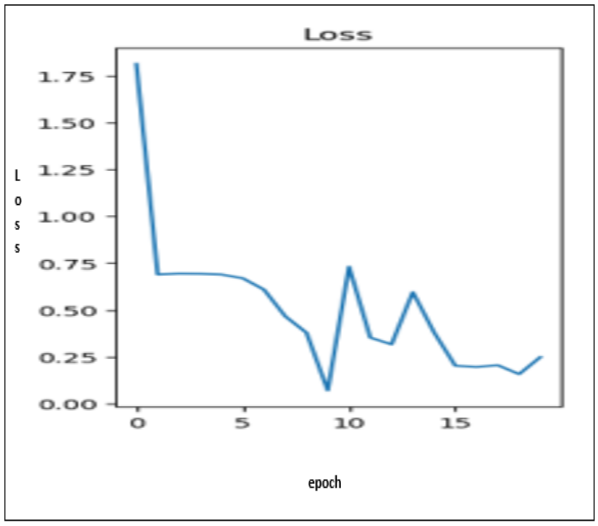

Fig 4.2 Model Loss Graph

Based on the above information from the graph, the CNN model used for PAN Card fraud detection has a training loss of 1.2 and a testing loss of 0.2 after 15 epochs. The precision of the model is 0.88, indicating that out of all the fraud predictions made by the model, 88% of them are fraud cases. The recall of the model is 0.94, indicating that out of all the actual fraud cases, the model correctly identifies 94% of them. These results suggest that the model has a high accuracy in detecting PAN card fraud cases with relatively low false positives.

IMPLEMENTATION OF GRAPHICAL USER INTERFACE

Fig 4.3 Sample Template Before Implementation

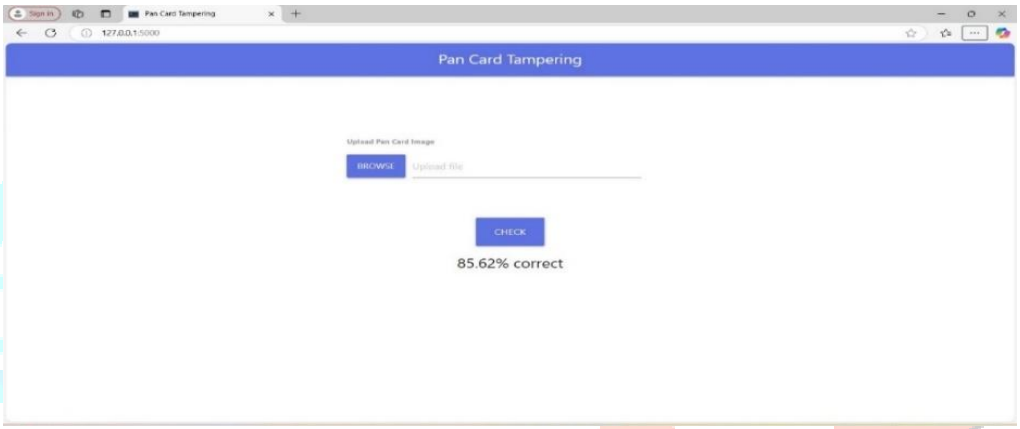

Fig 4.4 After Implementation

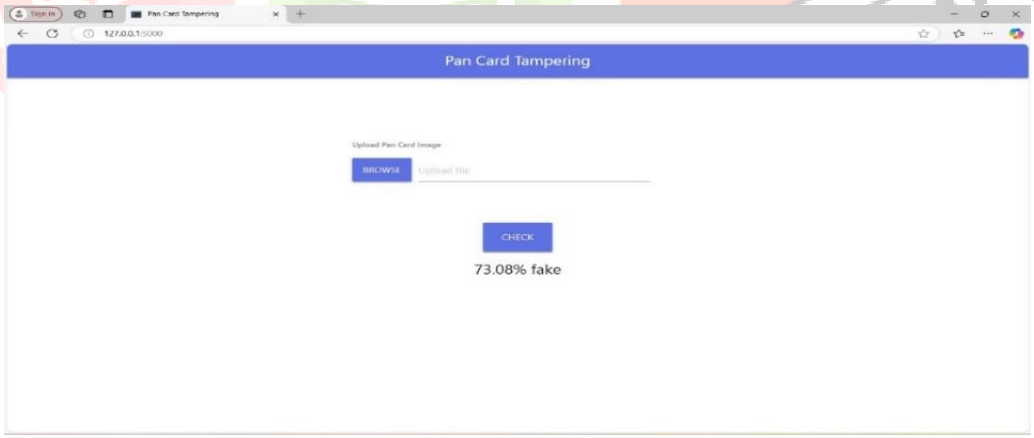

Fig 4.5 After Implementation

Web Interface

Flask API

- **Flask** is a lightweight WSGI web application framework. It's ideal for quickly developing APIs to expose your model as a service.
- **Guni corn** is a Python Web Server Gateway Interface (WSGI) HTTP server used for running Python Web

applications in production.

You can use either to:

- Accept image uploads.
- Run the image through ML models.
- Return a real/fake classification or detailed report.

V. Conclusion And Future Scope

The proposed system for **Fake PAN Card Detection using Machine Learning** provides modern and efficient solution to combat identity fraud and document forgery. By integrating powerful techniques like image preprocessing, feature extraction, and machine learning algorithms, the system ensures accurate and reliable identification of fake PAN cards. Through automated detection, the system minimizes manual verification errors and reduces the time and cost associated with traditional verification methods. It leverages structured analysis of visual and textual features, ensuring that any irregularities or deviations from standard formats are flagged effectively. This project not only showcases the practical application of machine learning in real-world problems but also opens new avenues for integrating artificial intelligence in digital document verification systems. As digital fraud continues to grow, such smart systems play a critical role in maintaining security, trust, and compliance in identity verification processes.

In conclusion, this system stands as a scalable and intelligent solution that can be further extended to detect forgery in other identity documents, making it a promising tool in the domain of AI-powered fraud detection. However, there is considerable scope for enhancing the system's functionality and applicability in the future. Some of the key future enhancements are:

1.Extension to Other Identity Documents

2.Mobile App Implementation

3.Cloud-Based and Scalable Architecture

4. Blockchain for Document Integrity.

References

- Kaur, P., & Sharma, M. (2019). Intelligent Document Verification System using Machine Learning Algorithms. International Journal of Engineering and Advanced Technology, 8(6), 1805–1810.
- Smith, R. (2007). An Overview of the Tesseract OCR Engine. Proceedings of the Ninth International Conference on Document Analysis and Recognition (ICDAR), Vol. 2, IEEE, 629–633.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.
- Bradski, G. (2000). The OpenCV Library. Dr. Dobb's Journal of Software Tools.
- Kaur, S., & Sharma, R. (2021). Document Forgery Detection Using Template Matching and Machine Learning Techniques. Journal of Advanced Research in Dynamical and Control Systems, 13(1), 145–151.
- Gupta, M., & Singh, V. (2019). Application of OCR in Document Analysis: A Review. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 5(2), 38–45.
- Pedregosa, F., Varoquaux, G., Gram fort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825–2830.

• Raj, A., & Patel, M. (2022). AI-Based Identity Verification for Preventing Document Fraud. International Journal of Emerging Technologies and Innovative Research, 9(2), 102–108.

• Patel, D., & Bhavsar, A. (2021). Image-Based Document Authentication using Deep Learning. International Journal of Computer Science and Network Security, 21(4), 32–39.

• Wang, J., & Luo, Z. (2020). A Comprehensive Survey on Document Image Analysis Techniques. Pattern Recognition Letters, 131, 1–12.

• Sharma, R., & Mehta, P. (2020). Fake Identity Detection System using OCR and Machine Learning. Journal of Computer Engineering & Intelligent Systems, 11(3), 55–63.

• Singh, T., & Yadav, N. (2018). A Review on Document Authentication using OCR and Machine Learning. International Journal of Innovative Research in Computer and Communication Engineering, 6(3), 2501–2507.

• Arora, A., & Verma, S. (2021). Document Verification System using AI and Image Processing Techniques. International Journal of Scientific Research in Engineering and Management, 5(12), 88–95