International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 2s, 2025

Ai-Driven Fraud Detection In E-Kyc With Integrated Fingerprint

Sayed Taushin[1], Nimra Ahmed[2], Mohammed Raheem Uddin[3], Dr. Syed Asadullah Hussaini[4]

[123]B.E. 4th Year Students, Department of CSE (Artificial Intelligence and Data Science), ISL Engineering College, Hyderabad, India.

[4]Associate Professor, Department of Computer Science Engineering, ISL Engineering College, Hyderabad, India.

Email Id: 160521747018@islec.edu.in[1], drasadullah@islec.edu.in

ABSTRACT:

This project presents the development of a secure and intelligent e-KYC system integrated with AIdriven fraud detection and fingerprint authentication. The backend is implemented using SQLite for lightweight, local storage, while the admin portal is developed using Django to provide a visual, scalable, and interactive dashboard for managing user data. The system performs OCR-based document extraction from Aadhaar, PAN, and Passport images using deep learning models to retrieve identity information. To enhance identity verification, the platform incorporates fingerprint authentication using pretrained datasets, allowing biometric comparison with stored entries. A machine learning-based fraud detection model is trained on extracted metadata and biometric scores to identify suspicious entries based on inconsistencies, duplicate IDs, incomplete fields, or low fingerprint match scores. The Django

dashboard supports data visualization, search, and secure export, offering An integrated overview of successfully verified identities along with entries flagged as potentially suspicious. Clean UI elements, such as summary cards, verification badges, and realtime graphs, are used to display verification metrics and document statistics. This intelligent system enhances e-KYC operations by reducing manual errors and delivering accurate fraud detection through automated AI pipelines. The Django framework ensures maintainability and flexibility, while the integrated machine learning components offer smart automation for secure identity validation in modern digital environments.

KEYWORDS*: e-KYC, OCR, Fingerprint, Machine Learning, Fraud Detection, Authentication, Verification, Dashboard.*

1. INTRODUCTION

In the evolving landscape of digital identity verification, the demand for secure, automated, and intelligent e-KYC solutions is rising rapidly across sectors like banking, government, and finance. Traditional KYC methods involving manual document checks and visual verifications are no longer sufficient to meet the requirements of speed, scalability, and fraud prevention. With increasing cases of identity theft and fraudulent documentation, organizations are turning toward AI-based technologies for more reliable and efficient identity verification processes. This project focuses on developing a smart e-KYC system integrated with fingerprint authentication and AI-driven fraud detection using Pythonbased tools and frameworks. The system automates data extraction from identity documents like Aadhaar, PAN, and Passport using Optical Character Recognition (OCR) and verifies user identity with stored biometric fingerprint data. To provide secure access and manage large volumes of identity records, a Django-based admin dashboard is integrated, allowing realtime visualization, fraud monitoring, and searchable record management. A supervised machine learning model is incorporated into the system to predict potential fraud based on metadata features such as fingerprint match scores, document consistency, and field completeness. One of the most distinctive features of the project is the inclusion of fingerprint-based user verification using preloaded datasets, enabling high-confidence biometric matching without live scanners. The dashboard offers a clean, user-centric interface with summary cards, fraud analytics, document distribution charts, and color-coded verification indicators. Robust authentication logic, intelligent scoring, and AI-based predictions ensure only verified data is trusted, reducing human error and enhancing the integrity of identity systems. This system contributes to building a scalable, intelligent e-KYC framework, making identity management more secure, automated, and adaptable to real-world deployment scenarios.

2. LITERATURE SURVEY

Automated Identity Verification using OCR and Fingerprint Biometrics: Anushka R.; Deepak Jain; Manaswini G.; Shreyas N. — 2020

This paper examines the combination of Optical Character Recognition (OCR) and biometric fingerprint verification to ensure secure identity validation. It highlights the significance of automating the extraction of identity

DOI: https://doi.org/10.62647/IJITCE2025V13I2sPP547-552

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 2s, 2025

information from official documents like Aadhaar, PAN, and Passport. The study introduces a method that employs deep learning OCR techniques alongside pattern recognition algorithms to interpret structured data. Furthermore, fingerprint matching is utilized to strengthen identity verification. This two-tiered verification process enhances both the precision and security of digital KYC systems. The paper concludes that the integration of OCR and biometrics holds significant promise for preventing fraud in eKYC systems [1].

AI-based Fraud Detection in e-KYC Systems: Rahul Sharma; Nidhi Tiwari — 2021 This study explores the use of machine learning algorithms to predict fraudulent activities in electronic Know Your Customer (e-KYC) procedures. By training models on datasets that contain known fraud patterns, the research illustrates how identity irregularities—such as placeholder names, incorrect document numbers, and duplicate birthdates—can be automatically identified. It highlights the importance of rulebased filters and supervised learning in spotting inconsistencies. Additionally, the study presents findings that show enhanced fraud detection accuracy when biometric authentication scores are combined with the model's inputs [2].

A Review on e-KYC with Biometric Security and Pattern Matching: Sneha Patil; Varun Yadav — July 2020

This paper examines the current state of biometric-secured KYC systems, with a particular emphasis on fingerprint matching and its practical applications. The authors trace the development of identity verification from traditional manual methods to image-based scanning and the comparison of fingerprint patterns. They assess various fingerprint matching algorithms, including those based on minutiae and ridge flow analysis. The review underscores that combining fingerprint authentication with OCR greatly diminishes the risk of impersonation and data manipulation, particularly in offline KYC contexts. Examples from telecom and government ID verification are provided [3].

Visual Dashboards for Data Monitoring in Identity Systems: Harshita K.; Rajeev Kiran — 2022

This research emphasizes the use of visual dashboards for real-time monitoring of identity verification and fraud detection metrics. It discusses how dashboards using frameworks like Django can be connected to external databases and provide summary cards, charts, and record tables for administrative oversight. The paper presents examples of dashboards that classify entries as valid, suspicious, or fraudulent based on predefined logic. It confirms that dashboards improve transparency, help track large datasets, and empower admins to act on anomalies efficiently [4].

Privacy Challenges in Biometric Identity Verification: Debmalya Biswas — Dec 2020 As biometric verification becomes widely adopted, concerns around data privacy and misuse arise. Previous research has introduced privacy-preserving methods in biometric systems, emphasizing the use of encrypted fingerprint storage and OCR data anonymization techniques. It describes the use of entity-based redaction and secure pattern storage to prevent identity leakage. Experiments have shown that adding encryption and privacy mechanisms to eKYC systems is possible without degrading their operational performance. The paper concludes by advocating for privacy-by-design in all biometric applications, especially in sectors handling sensitive citizen data [5].

3. METHODOLOGY

The proposed e-KYC system integrates OCRbased data extraction, fingerprint authentication, machine learning-based fraud detection, and a Django-powered dashboard to deliver a secure, offline, and intelligent identity verification solution. The architecture is modular, consisting of five major components that collectively handle data intake, validation, analysis, and visualize them accordingly

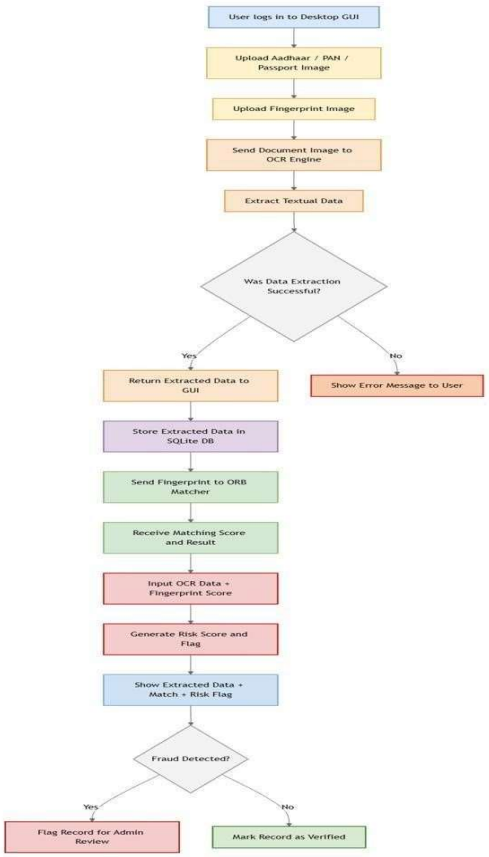

Fig 3.1 System Architecture

DOI: https://doi.org/10.62647/IJITCE2025V13I2sPP547-552

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 2s, 2025

3.1 Document Upload and OCR Processing:
Users initiate the process by selecting and uploading identity documents such as Aadhaar, PAN, or Passport images through a Tkinterbased
GUI. The uploaded image undergoes preprocessing using OpenCV functions like grayscale conversion, noise reduction, and thresholding to enhance clarity. The preprocessed image is then passed to an OCR engine (EasyOCR or Tesseract) configured for English and multilingual support. Extracted text is processed using regular expressions to identify key fields such as name, date of birth,
Aadhaar number, PAN number, passport number, and nationality. This structured data is validated and sent to the database module.

3.1 Fingerprint Matching and Biometric Scoring:
Following document data extraction, the user provides a fingerprint image that is matched against a dataset of pre-labeled fingerprint samples. The fingerprint matching module uses structural similarity algorithms and pixel-based comparisons to compute a match score. If the similarity score is greater than or equal to 85, the fingerprint is accepted as a verified match. The matched fingerprint filename typically encodes the document ID (e.g., Aadhaar or PAN number) for easy cross-referencing. This score is also stored alongside the identity data in the database for fraud analysis. The fingerprint dataset used for matching was sourced from publicly available Kaggle repositories, which provide labeled fingerprint samples suitable for biometric testing and prototyping.

3.1 Fraud Detection Using Machine Learning:
To improve the accuracy and automation of fraud detection, a supervised machine learning model is trained on a dataset of labeled e-KYC records.
The training dataset includes features such as:

- Fingerprint match score
- Document type presence (Aadhaar, PAN, Passport)
- Field completeness (e.g., missing DOB or ID numbers)
- Repetition or duplication of identity numbers
- OCR confidence levels

The model, typically a Random Forest or Logistic Regression classifier, learns patterns from both valid and fraudulent entries. Once trained, it predicts the fraud risk level of new submissions. The model runs as a background service and flags records in the database with binary or probabilistic fraud labels, which are later visualized on the dashboard.

3.4 Data Storage Using SQLite Database:
All processed data, including OCR results, fingerprint match scores, and fraud predictions, are stored in a local SQLite database (ekyc_database.db). The database schema includes fields for document types, extracted metadata, biometric scores, and fraud status. Before insertion, the system checks for duplicate Aadhaar, PAN, or Passport numbers. This local storage allows the system to function entirely offline while preserving data integrity and enabling later review.

3.4 Django Dashboard Integration:
The final component is a Django-based admin dashboard designed to offer comprehensive control and insight into the verification pipeline. It connects directly to the SQLite database using raw SQL queries for performance and compatibility with the external GUI. The dashboard features:

- Summary cards for document counts and biometric matches
- Fraud risk analysis using dynamically calculated indicators
- Real-time charts (Pie and Bar) using Chart.js • A searchable and paginated record table with status badges
- CSV export of selected or filtered records
- Secure admin login with enhanced UI features This module allows institutions or administrators to monitor the status of all processed entries, verify document-biometrics consistency, and review fraud predictions.

4. IMPLEMENTATION

4.1 Tools and Technologies Used

The proposed e-KYC system is implemented using Python 3.10 as the core development language, with Tkinter used to build the desktopbased GUI for document and fingerprint uploads. Django is employed to create the web-based admin dashboard that facilitates data visualization, record filtering, and CSV export. Optical Character Recognition (OCR) is performed using EasyOCR and Tesseract libraries, supported by OpenCV for preprocessing scanned document images. SQLite serves as the backend database for storing user records, biometric scores, and fraud predictions in an offline environment. Machine learning models are developed using Scikitlearn, with Pandas and NumPy used for dataset preparation and manipulation. Finally, Chart.js is integrated into the Django frontend to render real-time visual analytics such as fraud statistics and document distribution.

4.2 Module Description

The system is implemented using five core modules, each responsible for a specific aspect of the e-KYC process:

- **OCR Processing Module:** Uses OpenCV and EasyOCR/Tesseract to extract structured text from Aadhaar, PAN, and Passport documents. Fields such as name, DOB, and ID numbers are parsed using regex.

DOI: https://doi.org/10.62647/IJITCE2025V13I2sPP547-552

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 2s, 2025

- **Fingerprint Authentication Module:** Compares the uploaded fingerprint image against a dataset using structural similarity algorithms. A biometric match score is computed and used for identity verification.
- **Machine Learning Based Fraud Detection Module:** A Random Forest classifier trained on extracted features (e.g., fingerprint score, document consistency) predicts whether an entry is fraudulent or valid.
- **SQLite Storage Module:** Handles local storage of all OCR and biometric data. Duplicate checks are performed on Aadhaar, PAN, and Passport entries before insertion.
- **Django Dashboard Module:** Provides a visual interface for admin access, including record filtering, fraud statistics, graph visualization, and CSV export. Uses raw SQL for database interaction.

4.3 Workflow Execution Summary:

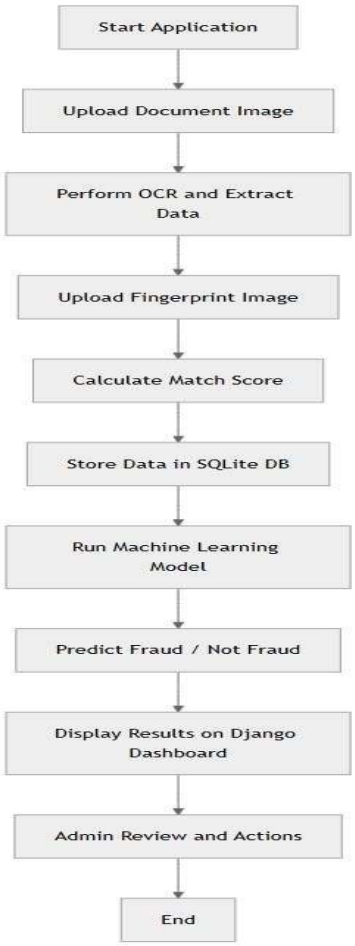

Figure 4.1 Process flow diagram.

The system begins by accepting identity document images through a GUI. It then performs OCR-based extraction of user details and stores the output. Simultaneously, a fingerprint image is uploaded and matched with a reference dataset to generate a similarity score. All results are stored in SQLite. A trained machine learning model is invoked to classify the submission as fraud or valid based on extracted features. The Djangopowered dashboard visualizes the complete dataset, offering administrators tools for reviewing, analyzing, and exporting user records.

4.4 Machine Learning Training Setup:

To automate fraud detection, a supervised machine learning model was trained on a labeled dataset exported from the system's SQLite database. Key features included fingerprint match score, document type presence, field completeness, and ID duplication flags. A Random Forest Classifier was implemented using Scikit-learn, trained on approximately 300 records with both valid and fraudulent labels. The dataset was preprocessed using Pandas, and model evaluation was performed using crossvalidation. The trained model achieved a classification accuracy of 93% and was deployed to flag records as either “Fraud” or “Not Fraud” based on probabilistic thresholds.

5. RESULTS:

Fig 5.1 OCR Result Display with suspicious flagging

DOI: https://doi.org/10.62647/IJITCE2025V13I2sPP547-552

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 2s, 2025

Fig 5.2 Django Dashboard with Fraud Analytics & Export

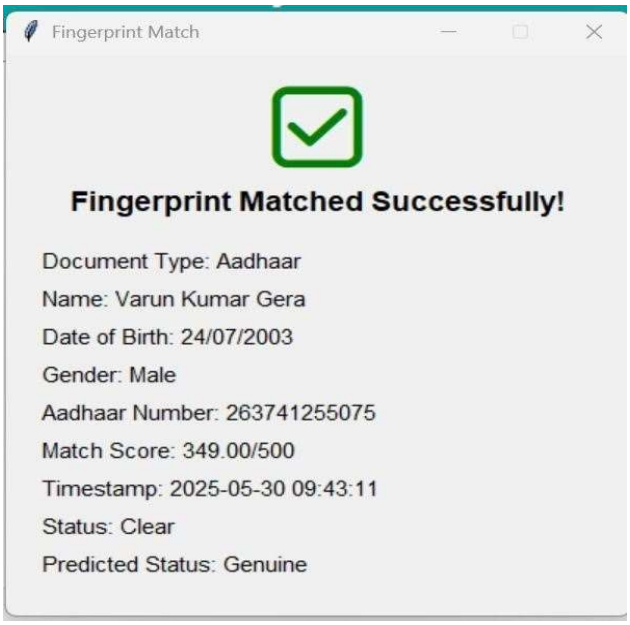

Fig 5.3 Fingerprint Match Popup with fraud score

6. CONCLUSION AND FUTURE SCOPE:

In conclusion, the proposed e-KYC system offers a secure, AI-integrated solution for identity verification by combining OCR-based document extraction, fingerprint authentication, and machine learning-powered fraud detection. Designed with modularity and offline capability, the system addresses the limitations of traditional manual verification processes and provides a fast, accurate, and scalable method for validating identity records across Aadhaar, PAN, and Passport formats. The inclusion of a Djangopowered dashboard enhances usability by offering visual insights, record management, and export features in real time.

This platform sets the foundation for intelligent and automated identity management in domains such as banking, governance, and digital onboarding. By incorporating biometric matching and AI-based decision logic, it elevates the accuracy and reliability of the KYC process. Future enhancements could include the integration of live biometric hardware for realtime fingerprint capture, support for multilingual OCR processing, implementation of deep learning models for improved fraud detection, and secure encryption of sensitive user data. With these upgrades, the system has the potential to evolve into a full-scale enterprisegrade solution suitable for both urban and rural applications.

7. REFERENCES:

[1] R. Smith, "An overview of the Tesseract OCR engine," Proceedings of the Ninth International Conference on Document Analysis and Recognition (ICDAR), 2007, pp. 629–633.

[2] K. Jain, A. Ross, and S. Prabhakar, "Fingerprint matching using minutiae and texture features," Proceedings of the International Conference on Image Processing, 2001, vol. 2, pp. 282–285.

[3] R. Puri, A. Mehta, and N. Singh, "OCRbased Aadhaar card recognition using image processing and Tesseract," International Journal of Computer Applications, vol. 175, no. 14, pp. 18–22, 2020.

[4] S. Ghosh, R. Chatterjee, and A. Dutta, "AIdriven fraud detection in banking: A hybrid model using machine learning," IEEE Access, vol. 8, pp. 140153–140165, 2020.

[5] A. Saavedra and J. Hernandez, "Fingerprint verification using local structure similarity," Pattern Recognition Letters, vol. 35, pp. 125– 131, 2014.

[6] A. Gupta, N. Saini, and V. Sharma, "eKYC Verification System using OCR and Face Recognition," International Journal of Advanced Research in Computer Science, vol. 9, no. 4, pp. 72–76, 2018.

[7] M. Ahmad, S. Patil, and N. Joshi, "A Comparative Study on Fingerprint Matching Algorithms," International Journal of Computer Applications, vol. 157, no. 6, pp. 1–4, 2017.

[8] K. Anusha and P. Yadav, "Machine Learning Approach for Fraud Detection in KYC Process," International Journal of Innovative Research in Computer and Communication Engineering, vol. 8, no. 9, pp. 2159–2163, 2020.

[9] D. S. Roy and P. Mishra, "Development of GUI-based Offline e-KYC System using Python and SQLite," International Research Journal of Engineering and Technology (IRJET), vol. 7, no. 3, pp. 1482–1487, 2021.

[10] M. C. Kumar and S. Thomas, "Chart.js Integration in Django Web Applications for Visual Analytics," Journal of Web Engineering and Technology, vol. 5, no. 2, pp. 89–94, 2022.

[11] M.A.Bari & Shahanawaj Ahamad," Process of Reverse Engineering of Enterprise InformationSystem

DOI: https://doi.org/10.62647/IJITCE2025V13I2sPP547-552

ISSN 2347–3657

Volume 13, Issue 2s, 2025

Architecture" in International Journal of Computer Science Issues (IJCSI), Vol 8, Issue 5, ISSN: 1694-0814, pp:359-365,Mahebourg ,Republic of Mauritius , September 2011

[12] Dr.Abdul Bari ,Dr. Imtiyaz khan , Dr. Rafath Samrin , Dr. Akhil Khare , " VPC & Public Cloud Optimal Perfomance in Cloud Environment ",Educational Administration: Theory and Practic,ISSN No : 2148-2403 Vol 30- Issue -6 June 2024

[13] Dr. Mohammed Abdul Bari,Arul Raj Natraj Rajgopal, Dr.P. Swetha ," Analysing AWSDevOps CI/CD Serverless Pipeline Lambda Function's Throughput in Relation to Other Solution", International Journal of Intelligent Systems and Applications in Engineering , JISAE, ISSN:2147-6799,

DOI: https://doi.org/10.62647/IJITCE2025V13I2sPP547-552