International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 14, Issue 2, 2026

Electronic Know Your Customer (E-Kyc) System

Tasneem Rahath[1], Kallepalli Bhavyasri[2], Thota Pragna Bharathi[3]

[1]Assistant Professor, Department of Information Technology, Bhoj Reddy Engineering College for Women, Hyderabad, India.

[2,3]B.tech Students, Department of Information Technology, Bhoj Reddy Engineering College for Women, Hyderabad, India.

kallepallibhavyasri@gmail.com, pragnathota2004@gmail.com

Abstract: *The automation of identity verification using digital documents and facial recognition has become important due to the increasing need for secure and efficient electronic Know Your Customer (eKYC) systems. This project focuses on developing an AI-based eKYC verification system that integrates Optical Character Recognition (OCR) and deep learning-based face verification to automate identity authentication using government-issued ID cards such as Aadhaar and PAN cards. The main objective of the system is to extract textual information from ID cards and verify user identity by comparing the face on the ID card with the uploaded face image. The system uses EasyOCR for text extraction and DeepFace with the FaceNet model for face verification. The uploaded ID card image undergoes preprocessing, contour detection, and text extraction to obtain user details such as name, ID number, date of birth, and gender. Face detection and embedding generation are performed on both the ID card image and the uploaded face image, and the similarity between the two faces is calculated to verify identity. The system also performs duplicate verification by checking existing records in the database before storing new user data. The extracted details, hashed ID number, and face embeddings are stored securely in a MySQL database and JSON storage.The proposed system provides an automated, secure, and efficient identity verification solution that reduces manual verification effort, improves accuracy, and prevents duplicate registrations. The modular architecture allows future enhancements such as live face capture and multi-factor authentication, making the system suitable for real-world eKYC applications in banking, finance, and digital identity verification systems.*

Keywords: *Face Verification, OCR, EasyOCR, DeepFace, FaceNet, Identity Verification, Image Processing.*

Introduction

In today's rapidly digitalizing world, organizations across sectors such as banking, finance, telecom, and government services increasingly rely on online platforms for customer onboarding. While this shift enables convenience and scalability, it also introduces challenges related to user verification, document authenticity, and fraud prevention. Traditional Know Your Customer (KYC) processes remain largely manual, requiring physical document submission, in-person verification, and significant human involvement. These conventional methods are time-consuming, prone to human error, and susceptible to document forgery, identity misuse, and impersonation attempts. As digital transactions rise, the demand for a secure, automated, and efficient verification mechanism has become essential.

Electronic KYC (E-KYC) has emerged as a modern solution, yet many existing systems rely on basic verification steps with limited automation. They often lack robust face matching, deepfake resistance, duplicate detection, and reliable document validation, making them vulnerable to fraudulent submissions. Additionally, inconsistencies in image quality, lighting, and distortions in uploaded ID cards further affect verification accuracy. These challenges highlight the need for an intelligent system capable of performing end-to-end verification using advanced computer vision and deep learning techniques.

To address these limitations, this project introduces an E-KYC Verification System that automates identity verification using face detection, face matching, OCR-based data extraction, deepfake/spoof detection, and document authenticity analysis. The system utilizes Haar Cascade for face detection, EasyOCR for text extraction, and DeepFace/FaceNet for generating face embeddings. It further incorporates document tampering detection and static-image deepfake analysis to ensure that the photograph and ID card submitted are genuine, untampered, and consistent. With database-based duplicate checks, secure storage, and a streamlined web interface, the system provides a fast, accurate, and user-friendly method for performing digital KYC.

Literature Review

Electronic Know Your Customer (eKYC) systems are widely used for remote identity verification in banking, finance, and digital onboarding. However, with the rapid growth of face-swapping technologies and AI-generated video manipulation, deepfake attacks have become a major threat to eKYC platforms. Recent research highlights that deepfakes can bypass traditional verification methods such as face

matching, simple liveness prompts, and document-image comparison, making enhanced security measures essential [1].

Existing studies show that earlier eKYC systems were primarily designed to detect basic spoofing attempts such as printed photos, masks, or replay videos. These systems fail to detect advanced AI-driven manipulations generated through identity-preserving face-swap models like SimSwap [2], FaceDancer [3], and SberSwap [4]. These tools can create highly realistic face-swapped videos that maintain natural expressions, smooth facial motion, and identity consistency—making them difficult for humans and machines to distinguish.

To analyze the impact of deepfake attacks on identity verification, the eKYC-DF dataset was introduced. It contains 228,760 real and manipulated videos created using multiple face-swapping techniques and reenactment models, including First-Order Motion Model-based animation [5]. This dataset is designed specifically for eKYC environments with realistic head-turn actions, blinking, lip movements, and mobile-camera conditions, making it highly relevant for training and testing robust deepfake detectors [1].

The eKYC-DF analysis also shows that face recognition systems experience a significant drop in accuracy when subjected to deepfake attacks, indicating that standard recognition models alone are insufficient for secure identity verification [7]. Additionally, the dataset highlights the need for liveness-driven protocols and controlled action prompts (e.g., turning head, blinking) to strengthen identity validation, as these are commonly used in practical KYC workflows [8].

According to the project PPT included in this work, the proposed eKYC system integrates face recognition with deepfake detection and follows a structured workflow involving image capture, preprocessing, verification checks, and decision generation [9]. This aligns with current research recommendations that emphasize combining recognition models, liveness tasks, and deepfake detection for improved system resilience.

Overall, the reviewed literature establishes that modern eKYC systems must incorporate deepfake detection alongside traditional biometric verification. The eKYC-DF dataset provides essential insights, tools, and evaluation protocols for developing systems that can withstand advanced face-swapping attacks and ensure secure, reliable digital identity verification.

Architecture:

System Architecture:

System architecture defines the structure and behavior of a system or solution, illustrating how various components interact to fulfill specific objectives. It provides a high-level representation of the application's flow, often depicted through diagrams.The primary goal of system architecture is to design a comprehensive and logically consistent solution based on established principles and concepts. It ensures that all system elements work cohesively to meet both functional and non-functional requirements. Architecture activities are focused on aligning the system with business goals and stakeholder expectations, while also addressing scalability, reliability, and performance. By identifying potential risks early, the architecture supports better planning and execution. Serving as a blueprint, it guides developers and engineers throughout the implementation process.

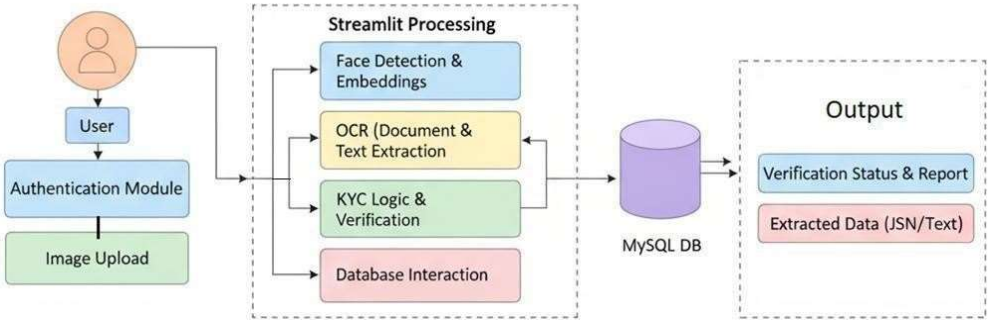


Fig.1 System Architecture

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 14, Issue 2, 2026

Technical Architecture:

The technical architecture of the eKYC Face Verification system is designed as a secure, modular, and scalable platform for automated identity verification. The system allows users to upload ID card images and face images through the front-end interface built using Streamlit. The Python-based backend handles image processing, OCR text extraction, face detection, face verification, database operations, and system logging. The AI pipeline integrates EasyOCR for text extraction and DeepFace with the FaceNet model for face verification. The uploaded ID card image undergoes preprocessing, contour detection, and text extraction to obtain user details such as name, ID number, date of birth, and gender. Simultaneously, face detection algorithms such as Haarcascade or RetinaFace detect and extract faces from both the ID card and uploaded face image. These faces are then converted into embeddings using deep learning models and compared to verify identity. After successful face verification, the system performs duplicate verification by checking the database for existing records. The extracted details, hashed ID number, and face embeddings are stored in a MySQL database and JSON storage for future reference and verification. The system also includes logging and configuration modules to manage system settings and record system activities. The modular architecture allows integration of additional features such as live face detection, duplicate face detection, and enhanced security mechanisms, making the system scalable, secure, and efficient for automated eKYC verification.

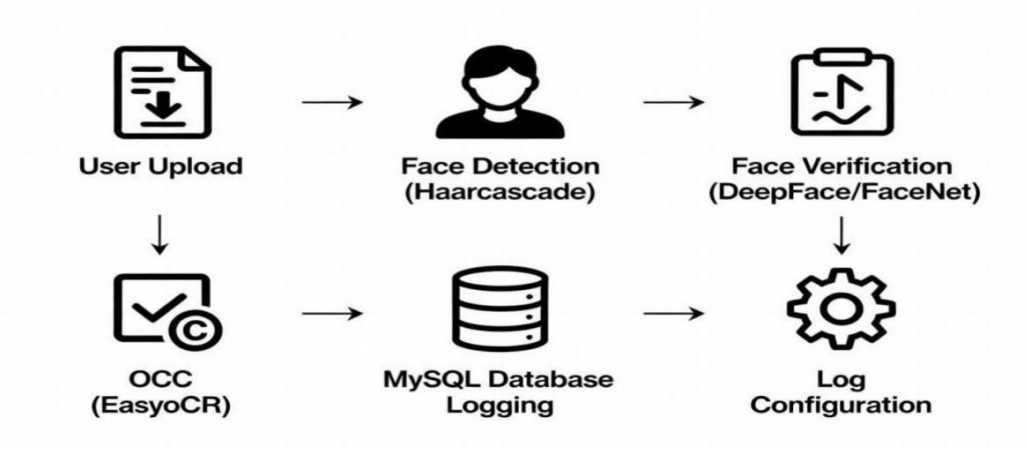


Fig.4.1.2Technical Architecture

Methodology:

The proposed eKYC system automates the identity verification process by integrating Optical Character Recognition (OCR), face detection, face verification, and database management techniques. The system verifies user identity using government-issued ID cards such as Aadhaar and PAN cards. The workflow involves uploading the ID card, extracting text information, detecting and comparing faces, checking duplicate records, and storing verified user data in a database.The system uses EasyOCR for text extraction and DeepFace with the FaceNet512 model for face verification. EasyOCR extracts textual information such as ID number, name, date of birth, and gender from the uploaded ID card image. The DeepFace model generates face embeddings and compares the face extracted from the ID card with the user-uploaded face image to verify identity. If the faces match, the extracted details are stored in the database after duplicate verification.The system consists of multiple modules including image preprocessing, OCR processing, text parsing, face detection, face embedding generation, face matching, duplicate checking, and database storage. Each module plays an important role in ensuring accurate and secure identity verification.

1. Image Preprocessing

Image preprocessing is performed on the uploaded ID card image to improve image quality and enhance OCR and face detection accuracy. The preprocessing stage includes grayscale conversion, resizing, noise reduction, thresholding, and contour detection to isolate the ID card region from the background.

Image preprocessing steps include:

- Grayscale conversion
- Image resizing
- Gaussian blur for noise reduction

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 14, Issue 2, 2026

- Adaptive thresholding
- Contour detection for ID card extraction
- Cropping the ID card region

This step improves the clarity of text and facial features, which increases OCR accuracy and face detects text regions and converts image text into machine-readable text. The extracted text includes important details such as name, ID number, date of birth, and gender.

OCR process includes:

- Preprocessed image input
- Text detection
- Character recognition
- Confidence filtering
- Text extraction output

The extracted text is then passed to the text parsing module to obtain structured information.

Implementation

Libraries:

Streamlit

Streamlit is used to develop the web interface for the eKYC verification system. It allows users to upload ID card images and face images through a user-friendly interface. Streamlit provides built-in components such as file uploaders, buttons, text inputs, and image display functions, which makes frontend development simple and fast. It also supports session state management, which is used for login authentication, registration, and page navigation within the application. Streamlit connects the frontend and backend directly using Python, eliminating the need for separate frontend frameworks. Overall, Streamlit acts as the main interface between the user and the eKYC system.

OpenCV (cv2)

OpenCV is used for image processing and preprocessing operations in the eKYC system. It helps in reading images, converting images to grayscale, resizing images, applying filters, detecting contours, and cropping the ID card region from the uploaded image. OpenCV is also used for face image enhancement and image saving operations. These preprocessing steps improve OCR accuracy and face detection performance. OpenCV plays an important role in preparing images before they are passed to OCR and face verification modules.

EasyOCR

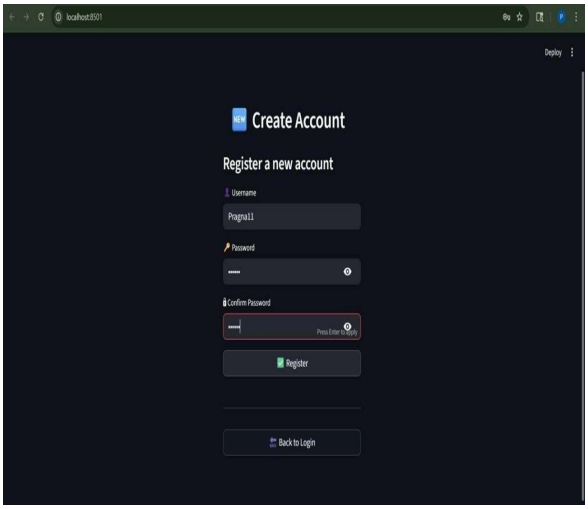
EasyOCR is used for Optical Character Recognition (OCR) to extract text from ID card images such as Aadhaar and PAN cards. It detects text regions in the detection performance.

2. Optical Character Recognition (OCR)

The system uses EasyOCR to extract text from the ID card image. The OCR engine

image and converts them into machine-readable text. The extracted text includes important information such as ID number, name, date of birth, and gender. EasyOCR supports multiple languages and works well with low-quality images, making it suitable for document text extraction. In this project, EasyOCR is used as the main text extraction engine.

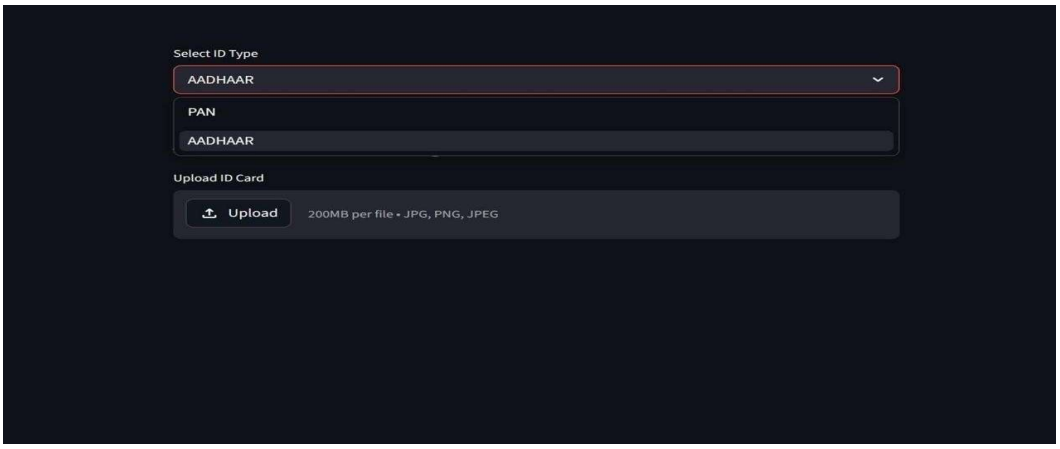
SCREENSHOTS



Screenshot 6.1 Register/Login Page

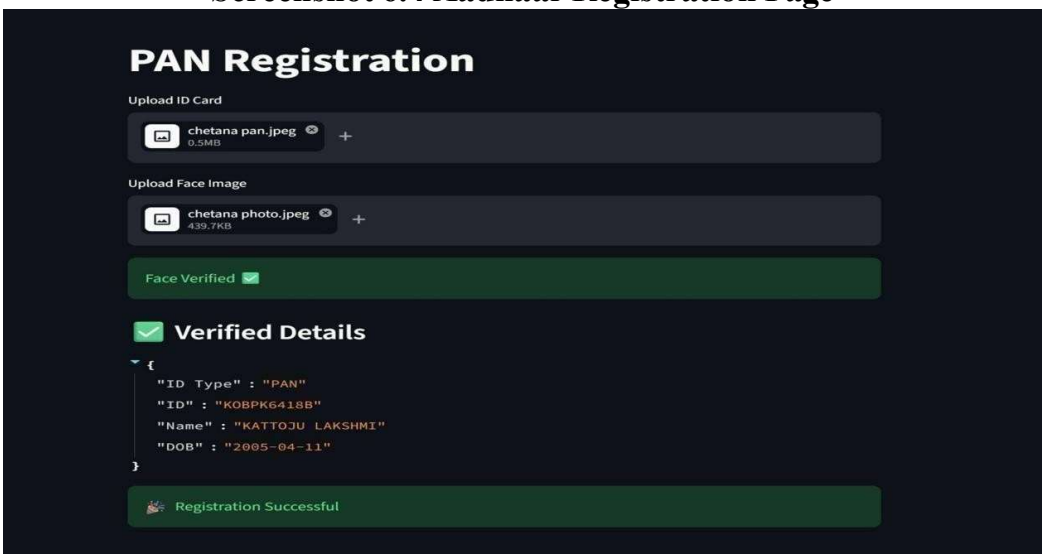
Screenshot 6.2 eKYC Verification Page

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 14, Issue 2, 2026



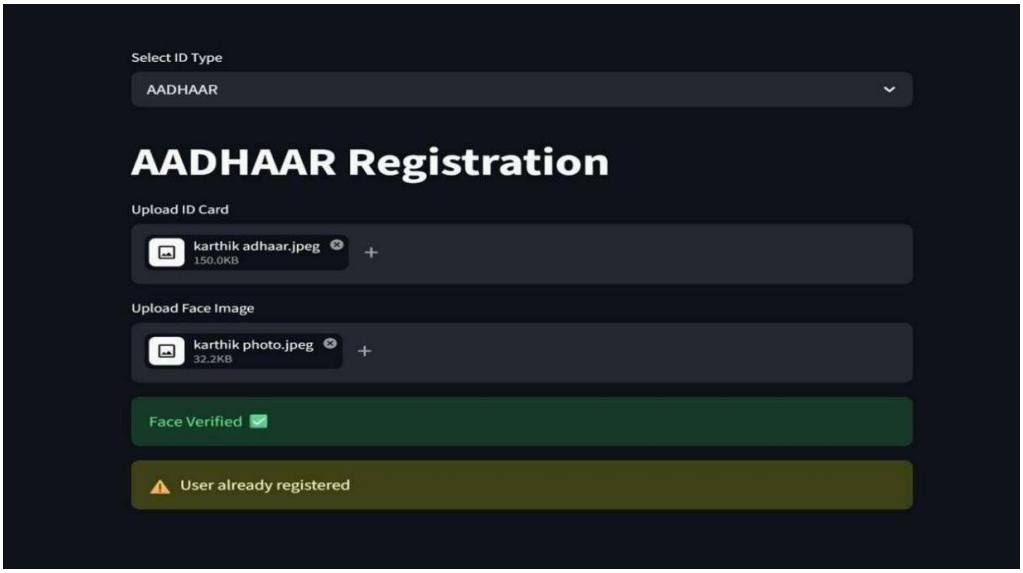
Screenshot 6.3 Documents Upload Page

Screenshot 6.4 Aadhaar Registration Page

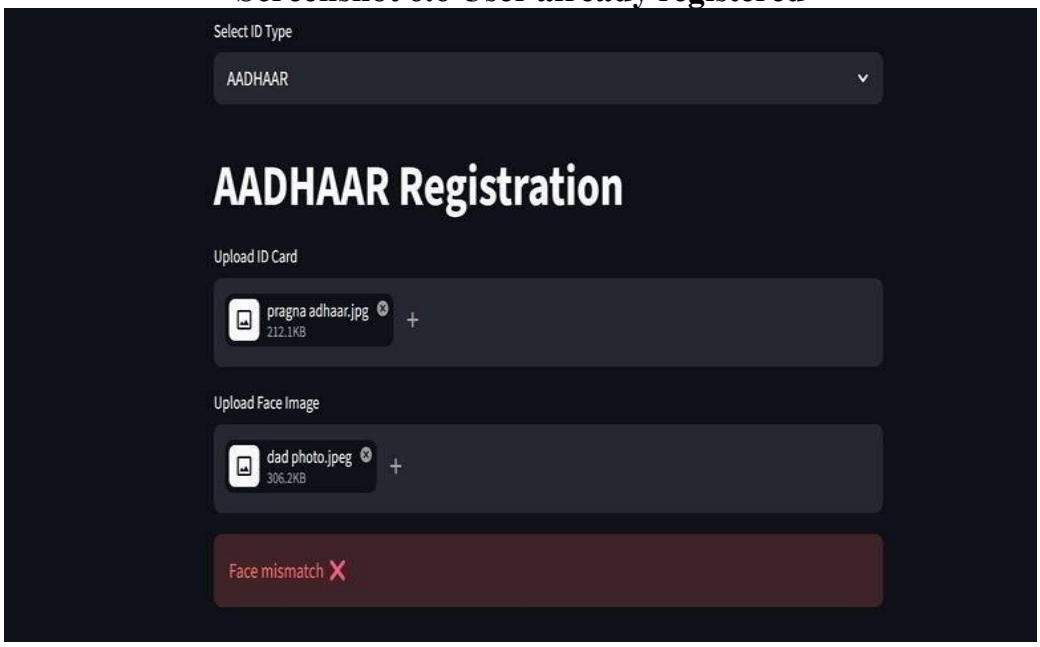


International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 14, Issue 2, 2026

Screenshot 6.5 Pan Registration Page

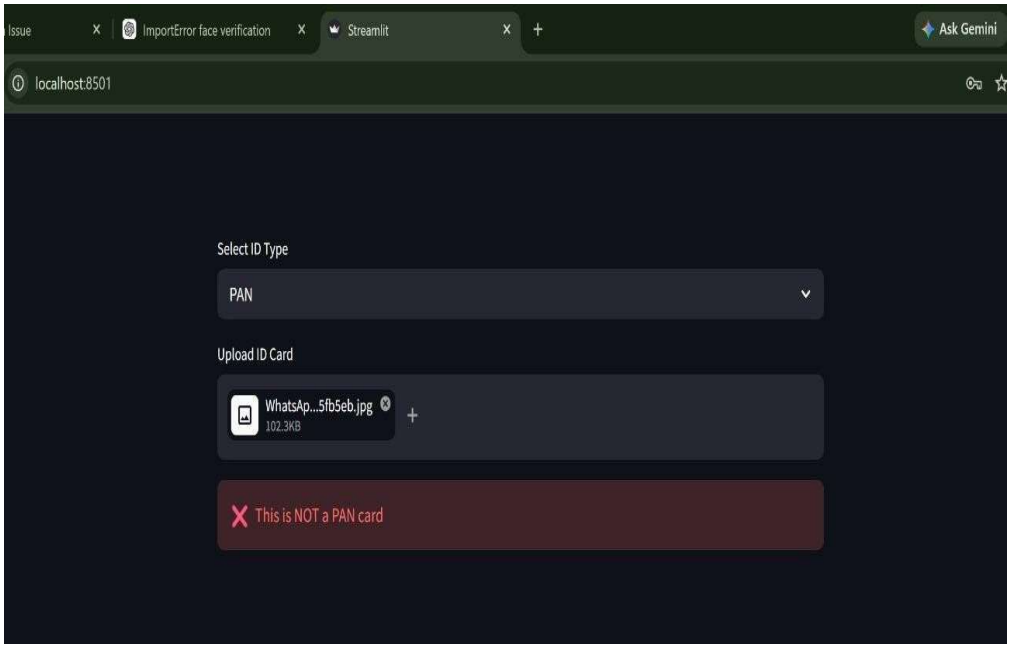


Screenshot 6.6 User already registered



Screenshot 6.7 Face Mismatch

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 14, Issue 2, 2026



Screenshot 6.8 Invalid Document

Conclusion

In this project, we developed an automated electronic Know Your Customer (eKYC) verification system using OCR and deep learning-based face verification techniques. The system demonstrates how computer vision can automate identity verification using government-issued ID cards like Aadhaar and PAN. It integrates image preprocessing, OCR-based text extraction, face detection, face embedding, face matching, duplicate verification, and database storage into a single workflow. EasyOCR is used to extract details such as name, ID number, date of birth, and gender. DeepFace with the FaceNet model is used for accurate face verification. The system compares the face from the ID card with the uploaded image to confirm identity. It also checks for duplicate records before storing new user data. This improves security and prevents multiple registrations. During development, challenges like OCR inaccuracies and face detection errors were addressed using preprocessing and enhancement techniques. Database integration issues were solved with proper handling methods. Python libraries such as OpenCV, EasyOCR, DeepFace, and MySQL Connector enabled smooth implementation. Streamlit was used to build a simple and user-friendly interface. The system performs well but may be affected by image quality and lighting conditions. Future improvements can include live face capture, liveness detection, and multi-factor authentication. Overall, the system provides a secure, automated, and efficient solution for real-world identity verification applications.

FutureScope

The proposed eKYC system can be improved by using live webcam capture instead of static images to enhance security and prevent misuse. Adding liveness detection techniques such as eye blink or head movement detection ensures that the user is physically present. The accuracy can be increased by using advanced face recognition models like ArcFace or FaceNet and by training custom OCR models for Aadhaar and PAN cards. The system can also be extended to support additional identity documents like passports and driving licenses. Converting it into a web or mobile application and deploying it on the cloud will improve accessibility and scalability. Furthermore, integrating features like OTP-based authentication, encryption, fraud detection, and duplicate face detection will strengthen security and make the system more suitable for real-world applications.

References

A. Rosebrock, “Optical Character Recognition (OCR) with Python and EasyOCR,” PyImageSearch, 2023.

ISSN 2347–3657
Volume 14, Issue 2, 2026

S. Serengil and A. Ozpinar, “DeepFace: A Lightweight Face Recognition and Facial Attribute Analysis Framework,” GitHub Repository, 2020.

F. Schroff, D. Kalenichenko, and J. Philbin, “FaceNet: A Unified Embedding for Face Recognition and Clustering,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

OpenCV Documentation, “Open Source Computer Vision Library,” Available: https://opencv.org/

J. Redmon and A. Farhadi, “YOLO Object Detection System,” IEEE Conference on Computer Vision and Pattern Recognition, 2018.

EasyOCR Documentation, Jaided AI, “EasyOCR Text Recognition Library,” GitHub Repository, 2023.

MySQL Documentation, “MySQL Database Management System,” Oracle Corporation, 2023.

Streamlit Documentation, “Streamlit Web Application Framework for Machine Learning,” 2023.

R. Gonzalez and R. Woods, “Digital Image Processing,” Pearson Education, 4th Edition.

A. Krizhevsky, I. Sutskever, and G. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” Neural Information Processing Systems, 2012.