Automated KYC System

Streamlining Identity Verification with OCR and Face Recognition

Toma Radu & Ivan Adrian

Problem Statement

Financial institutions and service providers need a secure, efficient way to verify customers' identities (KYC).

Objective: Develop a proof-of-concept web application that automates ID scanning, text extraction, and face verification.

Scope:

1. OCR of Romanian ID cards to retrieve user data.
2. Face detection on the ID card to isolate the printed face.
3. 3. Face verification between the ID’s face and a live/selfie image.

Implementation Architecture and Workflow

User uploads ID image:	• PaddleOCR extracts textual information from the ID. • Labels to Exclude and MRZ Pattern Matching parse critical fields.
User uploads selfie image:	• MTCNN detects and crops faces from the ID and selfie.
Face Verification:	• DeepFace compares the ID face and selfie face.
Results & Visualization:	• Extracted ID text fields and face verification result are displayed.

Key Components

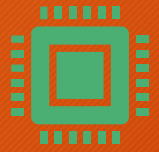
- Streamlit for user interface.
- PaddleOCR for Optical Character Recognition (OCR).
- MTCNN for face detection and cropping.
- DeepFace for face embedding and similarity comparison.

Key Performance Indicators (KPIs)

OCR Accuracy:

Metrics: Character Error Rate (CER), Word Error Rate (WER).

Target: High accuracy in extracting ID fields.



Face Detection Reliability:

Metrics: Precision, Recall of face detection.

Target: Correct bounding box detection.

Face Verification Accuracy:

Metrics: FAR, FRR, and overall accuracy.

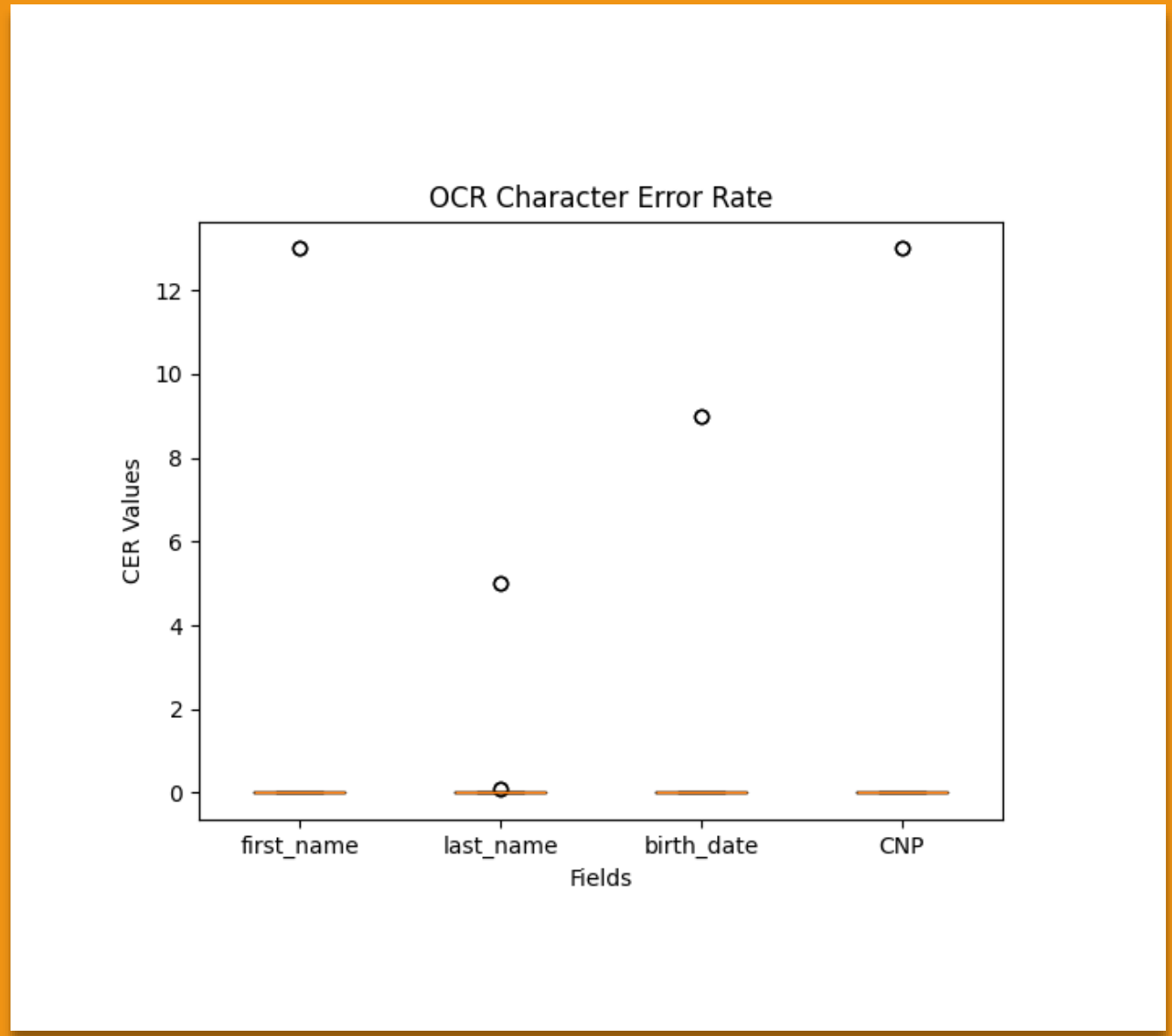
Target: Minimal false positives or negatives.

Processing Time:

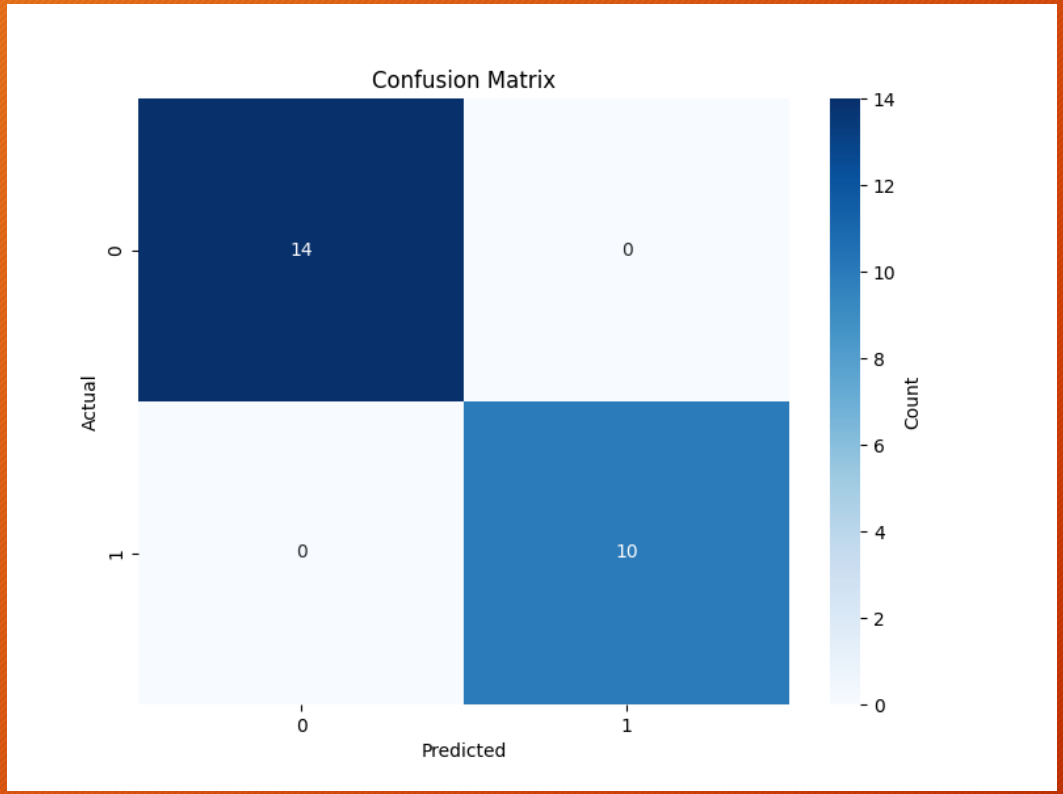
Metric: Average time for results.

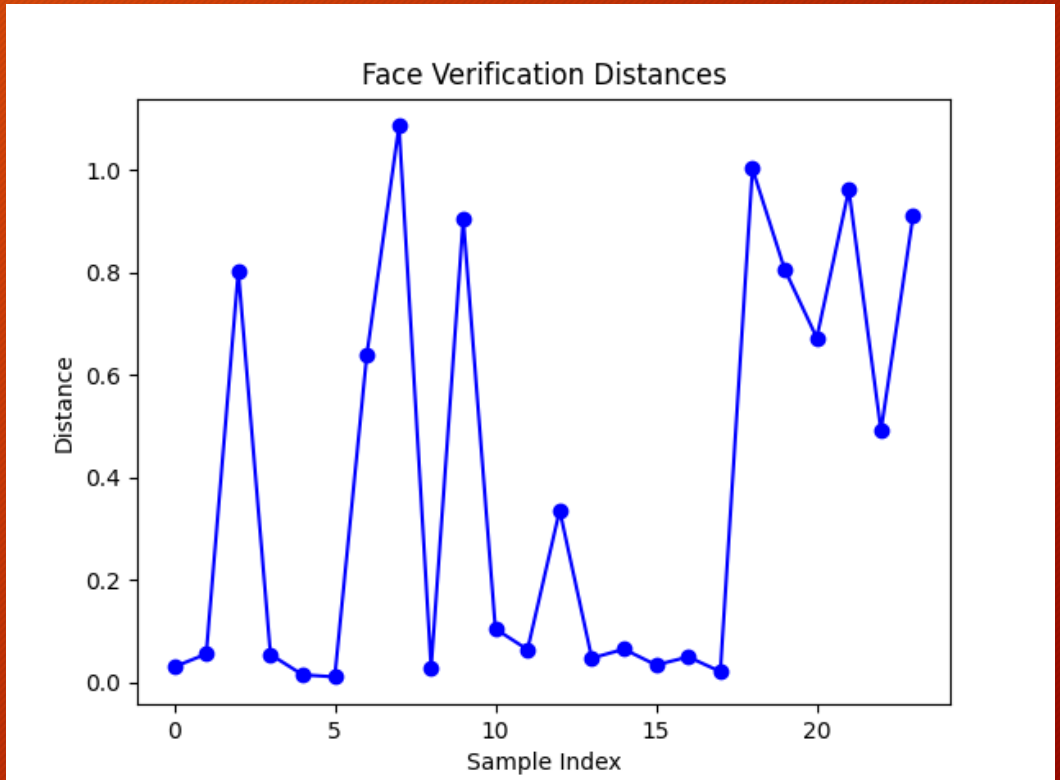
Target: Near real-time (few seconds).

OCR


OCR Character Error Rate
CER Values
0
2
4
6
8
10
12
first_name
last_name
birth_date
CNP
Fields

Face Recognition





Results

Successful OCR Data Extraction:

- Reliable extraction of CNP, name, and MRZ fields.

Face Cropping and Verification:

- High-confidence face detection and cropping using MTCNN.

Real-Time Demonstration:

- Deployed via Streamlit for quick testing.

Potential Improvements:

- Larger dataset for model fine-tuning.
- Enhanced preprocessing for low-quality images.
- Tuning DeepFace thresholds for specific risk tolerances.