IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, November 2025
Impact Factor: 7.67

An End-to-End AI System for Automated Verification of Indian Identity Documents Using OCR and Large Language Models

Prof. S. B. Raghuvanshi[1], Snehal Girase[2], Snehal Patil[3], Sejal Chaudhari[4], Manswi Ahirrao[5]
Department of Computer Engineering[1,2,3,4,5]
R. C. Patel Institute of Technology, Shirpur, India

Abstract: *This paper presents an end-to-end AI-based system designed to automate the verification of Indian identity documents such as Aadhaar, PAN, and Voter ID cards. The proposed solution combines Optical Character Recognition (OCR) with a state-of-the-art Large Language Model (LLM) to extract and validate document information. The system allows users to upload document images through a ReactJS and TailwindCSS-based interface, stores data securely using Supabase, and processes the content through an OCR engine for text extraction. A large language model is then employed to interpret the extracted text, validate key fields, and detect inconsistencies or anomalies. Documents achieving an accuracy score of 80% or higher are marked as valid, while those below this threshold are flagged as invalid. Experimental results demonstrate that the system achieves over 90% validation accuracy on clear document scans. The integration of AI reasoning with modern web technologies significantly reduces manual verification effort, offering scalability, transparency, and improved reliability for real-world KYC (Know Your Customer) and identity validation use cases.*

Keywords: OCR, LLM, documents, Verification, Anomalies

DOI: 10.48175/IJARSCT-29919