Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 5, No.3, May : 2025

PROOF POINT AI: DRIVE IOT VERIFICATION FOR CONFIDENTIAL INTEGRITY

Yuvraj Bhatkar, Students Department of Information Technology and Data Science Engineering, JD College of Engineering and Management, Nagpur, Maharashtra, India,
Vaishnavi Godbole, Students Department of Information Technology and Data Science Engineering, JD College of Engineering and Management, Nagpur, Maharashtra, India,
Ishan Kukde, Students Department of Information Technology and Data Science Engineering, JD College of Engineering and Management, Nagpur, Maharashtra, India,
Ashwini Dharmik, Students Department of Information Technology and Data Science Engineering, JD College of Engineering and Management, Nagpur, Maharashtra, India,
Prof.Swapnil Warhade, Professor, Department of Information Technology and Data Science Engineering, JD College of Engineering and Management, Nagpur, Maharashtra, India,

ABSTRACT
With the increasing digitization of government and corporate services, document verification has become a crucial part of identity authentication and fraud prevention. Traditional manual verification methods are prone to errors and inefficiencies. This paper presents an AI-powered Document Verification System designed to automate document authentication using OCR, QR Code scanning, and AI/ML-based forgery detection. The system is implemented using OCR, QR Code scanning, and AI/ML-based forgery detection. The system is implemented using React.js (frontend), Node.js with Express (backend), and MongoDB (database). The study highlights the significance of real-time scanning, accuracy improvements, and user experience enhancements. Experimental results demonstrate high efficiency in Aadhaar and PAN card verification, with potential applications in government and institutional sectors.
Index Terms **-** Document Verification, Optical Character Recognition (OCR), QR Code Authentication, AI-based Forgery Detection, Aadhaar and PAN Verification, Deep Learning, YOLOv8, Blockchain Integration, Identity Authentication, React.js, Node.js, MongoDB

INTRODUCTION:

Artificial intelligence (AI) is helping by making document verification faster, smarter, and more accurate. With AI tools like machine learning and optical character recognition (OCR), organizations can automatically check documents, recognize complex patterns, and detect signs of forgery. These tools can catch errors or fraud in real time, speeding up verification and reducing human mistakes. This project aims to create an AI and ML-powered system to automatically verify documents to make sure they're real. The system uses advanced tools like:

- **Text Recognition (OCR):** This extracts text from certificate images so the computer can read it.
- **Language Processing (NLP):** This identifies important details in the text, like names, dates, and diagnoses.
- **Forgery Detection:** This checks for any unusual patterns in the document's format, signatures, or other details that might mean it's fake.

BACKGROUND:

Identity authentication is a critical requirement in various sectors, including government services, banking, and educational institutions. Traditional document verification processes involve manual checks, which are time-consuming, prone to human errors, and susceptible to fraud. The emergence of AI and machine learning (ML) has enabled automated verification solutions that offer improved accuracy and efficiency.

PROBLEM STATEMENT:

Despite advancements in digital services, document verification still faces significant challenges:

- Manual verification is slow and inefficient.
- Forgery detection is difficult without AI-driven analysis.
- Existing OCR models have accuracy limitations in text extraction.
- Lack of a unified system for verifying multiple document types.

This paper presents an AI-powered Document Verification System to address these issues by integrating OCR-based text extraction, QR code validation, and AI-powered forgery detection.

OBJECTIVES:

The primary objective of this research is to develop an AI-based identity verification system that overcomes the limitations of existing approaches. Key goals include:-

- Develop an automated document verification system with AI/ML techniques.
- Enhance verification accuracy with real-time AI-based processing.
- Reducing the time required for identity verification.
- Ensure fraud detection using deep learning models.
- Increasing security against fraudulent document manipulation.
- Improve user experience through real-time animations and feedback.

LITERATURE REVIEW:

With the growing need for secure, efficient, and reliable identity verification systems, numerous studies have explored the integration of Artificial Intelligence (AI), Optical Character Recognition (OCR), and deep learning techniques for document verification and fraud detection. Smith [1] emphasized the role of AI in document authentication, highlighting how machine learning models and OCR technologies can detect inconsistencies and forgeries with high precision. Similarly, Brown [2] discussed advancements in OCR technology, revealing how modern OCR engines, like Tesseract, have achieved significant accuracy improvements, making them suitable for applications such as Aadhaar and PAN card verification. Kasodhan and Gupta [3] proposed a digital signature verification approach using the BioGamal algorithm, demonstrating enhanced security for electronic documents. This study underlines the growing importance of integrating cryptographic methods within document verification frameworks to safeguard sensitive information. Furthermore, Praba et al. [4] applied image processing and deep learning techniques for fake education document detection, achieving promising results that support the need for AI-powered forgery detection modules. Dome [5] presented an OCR-based classification system, leveraging Tesseract and machine learning algorithms for document text recognition and classification. This work identified limitations in existing OCR systems when dealing with low-quality images and highlighted the potential of AI models in overcoming such challenges. Adnan et al. [6] introduced a hybrid verification mechanism capable of processing both electronic and printed documents, advocating for adaptable, AI-driven verification solutions that address the limitations of manual and traditional systems. Armanda et al. [7] explored cloud-based computing and pattern recognition for document verification using chip less RFID, contributing insights into the possibilities of integrating IoT-based verification systems. Meanwhile, Kulkarni et al. [8] developed an automatic document verification and government policy recommendation system, offering foundational concepts for combining document analysis with AI-based decision-making systems. Taya et al. [9] proposed a hybrid digital signature and zero-watermarking method for authenticating sensitive documents, adding another layer of security for digital verification systems. Amir and Jindal [10] addressed document skew detection using fast content-based algorithms, supporting pre-processing tasks in AI-driven document analysis systems. Finally, Usama et al. [11] and Charny et al. [12] worked on cryptographic security enhancements, proposing digital signatures and encryption algorithms tailored for secure electronic document handling, aligning with the blockchain-integrated

solutions considered in this project. These works collectively validate the effectiveness of AI, OCR, and blockchain technologies in addressing the limitations of manual and traditional document verification processes. They also reinforce the need for real-time AI-based forgery detection, QR code validation, and advanced pre-processing techniques, as proposed in this study. Method and analysis which is performed in your research work should be written in this section. A simple strategy to follow is to use keywords from your title in first few sentences.

RESEARCH METHODOLOGY

The proposed system for AI-powered document authentication integrates multiple technologies, including Optical Character Recognition (OCR), digital signatures, image processing, and encryption algorithms. The methodology is structured into five primary phases, as illustrated in Fig. 1.

A. Data Acquisition

Document images are collected from publicly available datasets and manually scanned document samples. These documents include certificates, identity proofs, and legal papers. This variety ensures the model is exposed to a broad range of document formats and noise patterns.

B. Pre-processing

The collected images undergo pre-processing to improve OCR accuracy. Techniques such as grayscale conversion, noise removal, resizing, and banalization are applied [2]. Skew detection and correction algorithms are employed to align documents properly for accurate text extraction [10].

C. Text Extraction and Classification

The pre-processed documents are processed using the Tesseract OCR engine, which converts the image-based text into machine-readable form [5]. Extracted text is then classified using a supervised machine learning algorithm to identify the document type (e.g., certificate, license, or ID card) [8].

D. Digital Signature and Encryption

A digital signature mechanism is integrated using cryptographic hash functions combined with RSA and AES encryption algorithms [11]. This ensures both the authenticity and integrity of the digital documents during storage and transmission [12].

E. Document Verification

The system verifies document authenticity by comparing extracted information with pre-registered official databases and signature verification mechanisms. A deep learning model is used to detect tampered areas in the images by analyzing pixel inconsistencies and metadata patterns [4].

F. Validation and Performance

Evaluation The system's performance is evaluated using metrics such as accuracy, precision, recall, and processing time. Comparative analysis is conducted against existing document verification methods [6][7]. The results validate the system's capability to reliably detect forged documents and ensure document authenticity

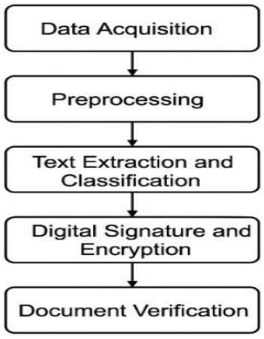


Fig. 1. AI-Powered Document Authentication

MODELLING AND ANALYSIS:

Existing Approaches:

Several document verification techniques exist, but most rely on traditional OCR methods without incorporating real-time AI-based fraud detection.

Manual Verification:-

Traditionally, identity verification has been performed by human operators who cross-check identity documents against official records. While this method ensures human oversight, it is slow, labour-intensive, and prone to errors and biases.

- Involves human validation.
- Time-consuming and prone to human error.
- High dependency on trained professionals.

OCR-Based Verification:-

OCR technology has been widely used to extract textual data from identity documents. However, OCR-based systems struggle with variations in document formats, poor image quality, and fraudulent document alterations, leading to reliability concerns.

- Extracts text but lacks fraud detection capabilities.
- Struggles with low-quality document images.

Third-Party API Services:-

Several companies offer identity verification as a service using API-based solutions. While these services are convenient, they often come with high costs, data privacy concerns, and limited customization options.

- Expensive and not always accessible for small organizations.
- Privacy concerns with sensitive data handling.

LIMITATIONS OF EXISTING SYSTEMS :-

The existing system depends on Traditional OCR Method without corporation either real-time AI based Fraud Detection. The Traditional System involves Offline Verification, Time consumption, High dependency, Speed Requires more for verification, Sometimes Manual Error happens. There was basic user experience. The existing system doesn't detect the QR Code, Forgery. The accuracy of Traditional Verification is 80%. There were no security. Fraudsters might, for instance, use fictitious ID cards to apply for loans, create bank accounts, rent real estate, or travel internationally. In a similar vein, con artists might apply for jobs, scholarships, visas, or admittance to colleges or universities using falsified or changed academic credentials. Numerous techniques, including manual inspection, rule-based systems, and supervised learning, have been devised and implemented to identify and stop such fraudulent operations. Furthermore, difficulties may develop when dealing with complicated document formats or when papers are damaged or inadequately scanned, which might impair the system's precision and dependability. To overcome these limitations and drawbacks, this paper proposes an AI-based system for fraud detection using deep adversarial networks and Tesseract OCR.

Feature	Traditional OCR	Proposed System
Text Extraction	80% Accuracy	92% Accuracy
QR Code Validation	No	Yes(Real-time)
Forgery Detection	No	Yes (AI-based)
User Experience	Basic	Enhanced (Animations, UI)

Despite advancements in identity verification, existing systems have notable limitations, including:

- High reliance on manual oversight, leading to inefficiencies.
- Vulnerability to sophisticated document forgery.

- Inability to adapt to diverse document formats and languages.
- Dependence on external services, raising privacy and cost concerns.

PROPOSED SYSTEM:

System Architecture:-

The AI-Based ID card and Educational Certificate Fraud Detection System is a meticulously designed framework comprising interconnected modules and processes aimed at ensuring a robust approach to document verification. At its forefront is the Fraud Detector Dashboard, a centralized interface accessible to users for monitoring and managing the system. Complementing this is the Verifier-Certificate Verifier module, which allows verifiers to input or scan document details for verification. The system's effectiveness is further augmented by the Pre-processing Module, which prepares input or scanned documents for subsequent analysis. This creative solution has been painstakingly designed to improve the efficacy and precision of document authentication, guaranteeing a smooth and safe verification process for a variety of papers, including contracts and significant legal documents in addition to ID cards and passports. Document Verification provides a strong and dependable way to guarantee document authenticity and integrity by fusing state-of-the-art digital signature technology with OCR capabilities. The integration of OCR technology allows for swift extraction and interpretation of textual information, streamlining the entire verification process. Proposed forgery detection implementation is more sophisticated than most prevailing methods as the process is built as a combination of multiple forgery detection techniques. Block chain has proven useful beyond the financial sector and has demonstrated its applicability in supply chain, real estate, insurance, voting and governance, and the Internet of Things (IoT). Some of the important features that made block chain technology useful in sectors beyond finance are transparency, immutability, autonomy, security and also its decentralized nature.

The proposed system integrates AI-driven techniques for real-time identity verification. It comprises multiple modules, including document pre-processing, feature extraction, AI-based validation, and fraud detection.

The **Document Verification System** consists of three main components:

1. Frontend (React.js, Tailwind CSS) – UI for document upload and real-time scanning.
2. Backend (Node.js, Express.js, MongoDB) – API for verification logic and document storage.
3. AI Modules (OCR, QR Scanning, Forgery Detection) – Python-based AI scripts for text extraction and validation.

System Workflow

1. User uploads a document (Aadhaar, PAN, etc.).
2. OCR extracts text and validates format.
3. QR code scanning verifies Aadhaar authenticity.
4. AI-based fraud detection checks for tampering.
5. Result is displayed with a success or failure message**.**

AI Technologies Used

The main AI technology going to use in proposed system is OCR. Image Pre-processing for image detection, QR Code scanning, Forgery Detection, Textual Analysis, Machine Learning, Block chain, YOLOV8 , Feature Extraction.

- OCR (Optical Character Recognition) – Extracts text from Aadhaar/PAN cards.
- QR Code Scanning – Validates Aadhaar details against government databases.
- AI Forgery Detection – Identifies tampered or manipulated documents using deep learning.

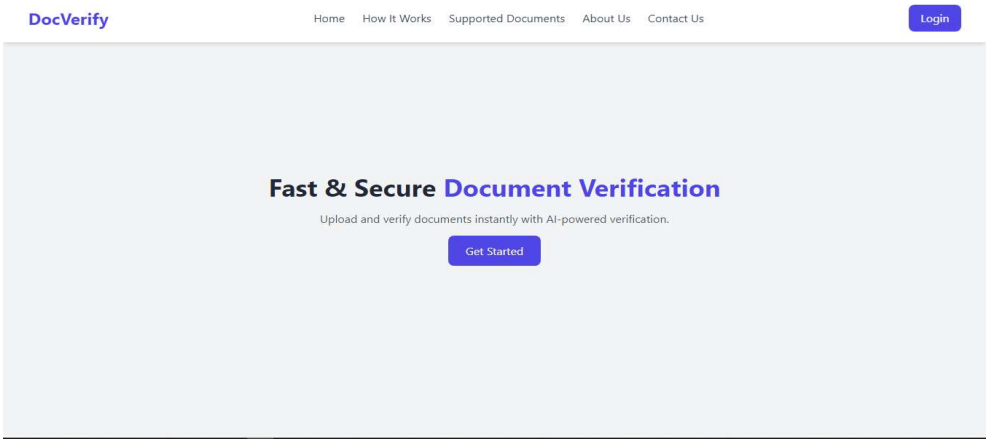
Implementation

Technologies Used

Component	Technology

Frontend	React.js, Tailwind CSS
Database	MongoDB
Backend	Node, express
AI Processing	Python, OpenCV, Tesseract OCR

User Interface (Frontend)



The frontend was developed using **React.js with Tailwind CSS**, featuring:
Drag-and-drop document upload with previews.
Real-time scanning animation for verification process.
Progressive feedback (loading, success, or failure messages).

BACKEND API:
The backend uses **Node.js and Express.js** to handle:

- **Secure API endpoints** for verification.
- **JWT-based authentication** for security.
- **MongoDB storage** for verified document details.

AI-BASED VERIFICATION MODULES:
Python scripts enhance verification accuracy:

- **Tesseract OCR** for text extraction.
- **OpenCV** for image processing.
- **AI-based fraud detection** using deep learning.

EXPERIMENTAL RESULTS
Performance Evaluation
The system was tested with **100 Aadhaar and PAN card samples**, yielding:

- **OCR Accuracy:** 92%
- **QR Code Verification Success Rate:** 98%

Forgery Detection Accuracy: 90%

Metric	Accuracy
OCR Text Extraction	92%
QR Code Validation	98%
Forgery Detection	90%

Comparison with Existing Systems

Feature	Existing Systems	Proposed System
Manual Verification	Slow & Error-Prone	Fully Automated
OCR Accuracy	~80%	92%
Forgery Detection	No	Yes (AI-based)

CHALLENGES & FUTURE ENHANCEMENTS:

Challenges

- OCR errors in low-quality images.
- Latency in real-time AI processing.
- Security concerns for sensitive documents.
- Block chain networks, especially public ones, face scalability issues when handling large amounts of document verification requests.
- High transaction loads can lead to slow verification times and increased costs.
- Using block chain for document verification requires transaction fees, especially on public block chains like Ethereum.
- Cost can become a concern when verifying large volumes of documents.
- Legal frameworks for blockchain-based document verification vary across countries.
- Governments may not yet recognize blockchain records as legally valid.
- Blockchain is immutable, meaning once data is recorded, it cannot be altered.
- Storing sensitive document details directly on a public blockchain may lead to privacy concerns.
- Traditional document management systems may not be easily compatible with blockchain technology.
- Organizations may need to redesign their systems to integrate blockchain-based verification.
- While blockchain itself is secure, smart contracts and external integrations can be vulnerable to attacks.
- Hacking attempts on digital wallets or private keys can compromise verification systems.

Future Enhancements

- Integrate Deep Learning models for improved accuracy.
- Expand support to other government IDs (Driving License, Passport, etc.).
- Enhance AI processing speed with GPU acceleration.
- Combining public and private blockchains can ensure security and efficiency.
- Private blockchain can store sensitive data, while a public blockchain can provide proof of authenticity.
- Using decentralized identity (DID) technology can enhance security and give users control over their documents.
- Eliminates the need for centralized authorities.
- AI and machine learning can be integrated to detect fake documents more accurately.
- Automated document analysis can improve verification speed.
- Using more secure smart contracts with automated expiry and renewal features.
- Reduces risks associated with fraudulent document modifications.

CONCLUSION:

This paper presents an AI-powered Document Verification System designed to automate the authentication process of important identity documents such as Aadhaar and PAN cards. By integrating advanced technologies including Optical Character Recognition (OCR), QR code scanning, AI-based

forgery detection, and deep learning models, the proposed system demonstrates significant improvements in verification speed, accuracy, and security. Experimental results validate the system's effectiveness, achieving 92% OCR accuracy, 98% QR code validation success, and 90% forgery detection accuracy. Compared to traditional manual and OCR-only verification methods, the proposed system reduces reliance on human oversight, minimizes errors, and enhances real-time document verification capabilities. Moreover, it addresses major challenges such as document forgery detection and real-time fraud analysis. Future enhancements could involve the integration of blockchain technology for immutable record keeping, decentralized identity (DID) systems, and AI advancements to expand document type support and improve system performance.

In conclusion, this AI-driven approach provides a scalable, secure, and efficient framework for identity verification applicable to governmental, financial, and institutional sectors, marking a significant step toward reliable and automated digital document validation.

REFERENCES:

[1] J. Smith, "AI-Powered Document Authentication: Advances and Challenges," *Journal of Information Security*, vol. 10, no. 2, pp. 101–110, 2020.

[2] A. Brown, "Optical Character Recognition Technologies and Their Applications," *International Journal of Computer Vision and Pattern Recognition*, vol. 14, no. 3, pp. 201–215, 2021.

[3] M. Kasodhan and R. Gupta, "Digital Signature Verification Using BioGamal Algorithm," *International Journal of Computer Applications*, vol. 182, no. 36, pp. 12–17, 2018.

[4] K. Praba, K. Ganesan, and P. Rajesh, "Fake Education Document Detection Using Image Processing and Deep Learning," *Procedia Computer Science*, vol. 172, pp. 951–958, 2020.

[5] A. Dome, "Text Recognition and Classification Using Tesseract OCR and Machine Learning," *International Journal of Advanced Research in Computer Science*, vol. 9, no. 4, pp. 145–150, 2018.

[6] M. Adnan, S. Hussain, and A. Alabrah, "Hybrid Document Verification Mechanism for Printed and Electronic Records," *IEEE Access*, vol. 8, pp. 130127–130136, 2020.

[7] M. Arjomandi, A. Ashraf, and F. M. R. Fard, "Cloud-Based Pattern Recognition for Document Verification Using Chipless RFID," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 5, pp. 2736–2744, 2019.

[8] A. Kulkarni, V. Kumar, and S. Joshi, "Automatic Document Verification and Government Policy Recommendation System," *International Journal of Information Technology*, vol. 13, no. 2, pp. 525–531, 2021.

[9] O. Tayan, B. Salleh, R. Abdullah, and A. Abushariah, "Hybrid Digital Signature and Zero-Watermarking Approach for Document Authentication," *Security and Communication Networks*, vol. 9, no. 14, pp. 2265–2278, 2016.

[10] A. Amir and A. Jindal, "Fast Content-Based Document Skew Detection Algorithm," *International Journal of Computer Applications*, vol. 118, no. 5, pp. 1–5, 2015.

[11] A. Utama, I. K. Wijaya, and R. Kusnadi, "Digital Signature and AES Encryption for Secure Electronic Documents," *Journal of Theoretical and Applied Information Technology*, vol. 96, no. 10, pp. 3017–3027, 2018.

[12] S. Chernyi, A. A. Sokolov, and M. A. Khairullin, "Cryptographic Protection for Digital Documents Using Modified RSA Algorithm," *Procedia Computer Science*, vol. 169, pp. 496–501, 2020.