International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

AI-Blockchain Driven Official Document Verification Framework

Hitesh Atkar[1], Samarth Karale[2], Anushka Sathe[3], Vedant Tambe[4], Rita Kadam[5]

[1]Project Leader, Information Technology, SOET, DYPU, Ambi, Pune
[2] Documentation Manager, Information Technology, SOET, DYPU, Ambi, Pune
[3]Data Analyst, Information Technology, SOET, DYPU, Ambi, Pune
[4]Tester/Quality Analyst, Information Technology, SOET, DYPU, Ambi, Pune
[5]Project Guide, Information Technology, SOET, DYPU, Ambi, Pune

Abstract
In the digital era, verifying the authenticity of official documents is crucial to prevent fraud and ensure security. Traditional verification systems rely on centralized entities, making them susceptible to data tampering and inefficiencies. This paper proposes an AI- Blockchain Driven Official Document Verification Framework that utilizes Blockchain's decentralized, immutable nature and AI's automation capabilities to establish a transparent and tamper-proof verification process. The framework integrates smart contracts, cryptographic hashing, and AI-driven document authentication to enhance security, integrity,and efficiency in document verification.

Keywords: Artificial intelligence, Deep Learning, KYC, Fraud Prevention, Digital Solution

1. Introduction

In the age of digitization, authenticating official documents is essential to prevent fraudulent activities and ensure security. Conventional verification systems rely on centralized organizations, making them vulnerable to data tampering and inefficiencies. This project presents an AI-Blockchain Driven Official Document Verification Framework, leveraging Blockchain's distributed and immutable nature along with AI-driven automation to establish a secure and transparent validation process. By integrating OCR and NLP, AI efficiently automates data extraction, minimizing errors and processing time, while Blockchain ensures transparency and immutability, protecting against fraud. The framework enhances trust, scalability, and efficiency, making it a reliable solution for businesses, landlords, and financial institutions requiring KYC verification. Additionally, it adheres to data privacy standards, ensuring compliance with global regulations such as the General Data Protection Regulation (GDPR).

2. Literature review

Blockchain and Artificial Intelligence (AI) have been extensively researched for their ability to enhance the security, transparency, and efficiency of document verification processes. Traditional methods often involve centralized authorities, making them susceptible to fraud, inefficiencies, and unauthorized alterations.

Satoshi Nakamoto's pioneering work introduced the concept of a decentralized and immutable digital ledger through the Bitcoin network [1]. This laid the foundation for broader blockchain applications beyond cryptocurrencies. Following that, Ethereum extended blockchain's utility by
supporting smart contracts—self-executing code that can automate tasks like document verification [3].
Zheng et al. [2] provided a comprehensive overview of blockchain architecture, consensus mechanisms, and future trends, highlighting its potential to improve data integrity and eliminate single points of failure in verification systems. Wood's Ethereum Yellow Paper [3] elaborated on how smart contracts can securely enforce validation rules without intermediaries.
Swan [7] emphasized blockchain's broader societal and technological impact, advocating for its use in sectorslike identity and documentation. Applications in supply chain operations have demonstrated blockchain's utility in enhancing data transparency and trust [5], which is directly translatable to document verification processes.
In healthcare, blockchain has been explored for securing sensitive patient records, indicating its viability for storing official documents in other domains as well [4]. The combination of blockchain and Artificial Intelligence is further discussed by Bertino et al. [6], who explored ethical implications and transparency improvements through integrated AI-blockchain systems. Bhardwaj et al. [8] proposed a model combining blockchain with AI techniques like Optical Character Recognition (OCR) and anomaly detection to automate and secure document verification, reducing the need for manual checks.
Gupta et al. [9] reviewed smart contract applications in document workflows and emphasized cryptographic hashing as a technique to verify document integrity without revealing sensitive data. The integration of predictive AI capabilities with blockchain ensures improved accuracy and trust in validation processes.
Finally, considerations around privacy and data regulations, such as the European Union's GDPR [10], highlight the importance of compliance in any AI-blockchain-based system, especially when handling personally identifiable information.

3. Methodology

The AI-Blockchain Driven Aadhaar and PAN Verification System follows a structured approach to ensure secure, accurate, and efficient document verification. The process begins with data collection and preprocessing, where Aadhaar and PAN card samples are gathered, enhanced, and processed using Optical Character Recognition (OCR) for text extraction. Advanced preprocessing techniques, including noise reduction, contrast enhancement, and skew correction, improve text clarity, while Natural Language Processing (NLP) techniques validate format compliance, ensure entity recognition, and detect anomalies. In the AI-based verification phase, OCR models such as Tesseract, EasyOCR, and CRNN extract text with high accuracy, while NLP-based validation cross-checks extracted data against predefined templates.

3.1 System Architecture

The system architecture comprises three layers: Frontend, Backend, and Blockchain Integration. The Frontend deals with user input and file uploads, whereas the Backend performs document processing and verification through AI. Document data after verification is stored securely in the Blockchain for tamper-proof verification via smart contracts. User authentication and verification logs are dealt with by a database.

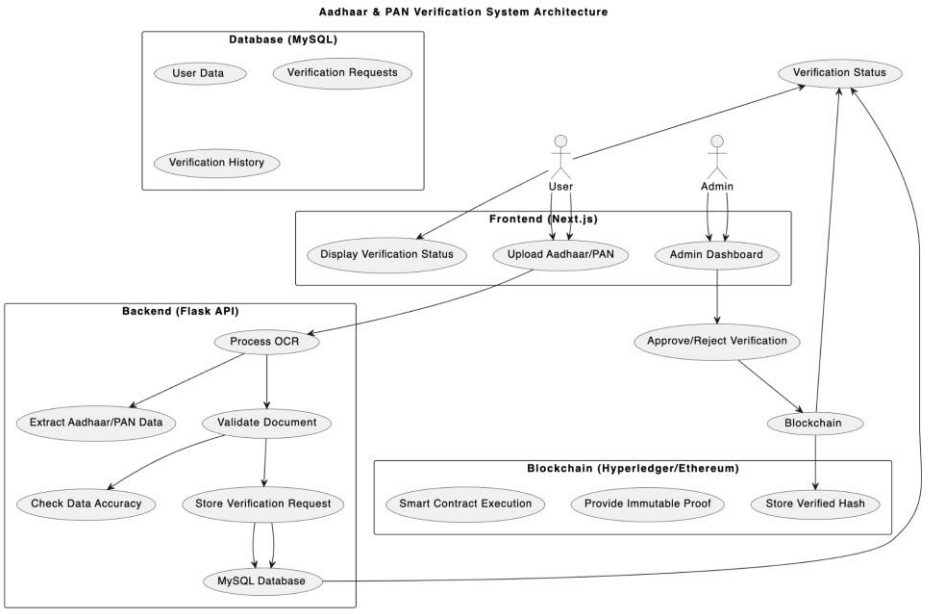

Fig.1. System Architecture.

3.2 Image Processing Pipeline

The image processing pipeline in the verification system utilizes OCR and deep learning for accurate text extraction. Preprocessing techniques, including noise reduction and contrast enhancement, improve text clarity, while NLP-based entity recognition ensures precise data validation. The extracted information is securely verified through a Blockchain framework, ensuring tamper-proof and reliable document authentication.

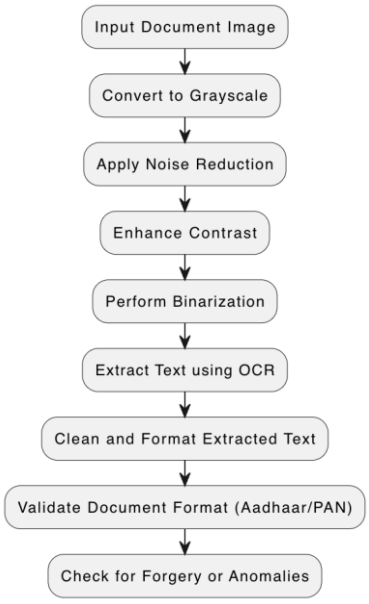

Fig.2. Image Processing Pipeline.

3.3 Flowchart

The Verification Framework ensures secure and tamper proof document authentication by leveraging OCR, Deep Learning, NLP, and Blockchain. Users upload Aadhaar or PAN card images, and the OCR engine extracts key details such as name, DOB, and unique identification number. AI-based verification checks for inconsistencies, fraud, or tampering, while Blockchain technology ensures data integrity by

cross verifying extracted details against immutable records. Once verified, the results are securely stored in a MySQL database and can be accessed via an API for seamless KYC integration, enabling businesses to perform real-time identity verification efficiently.

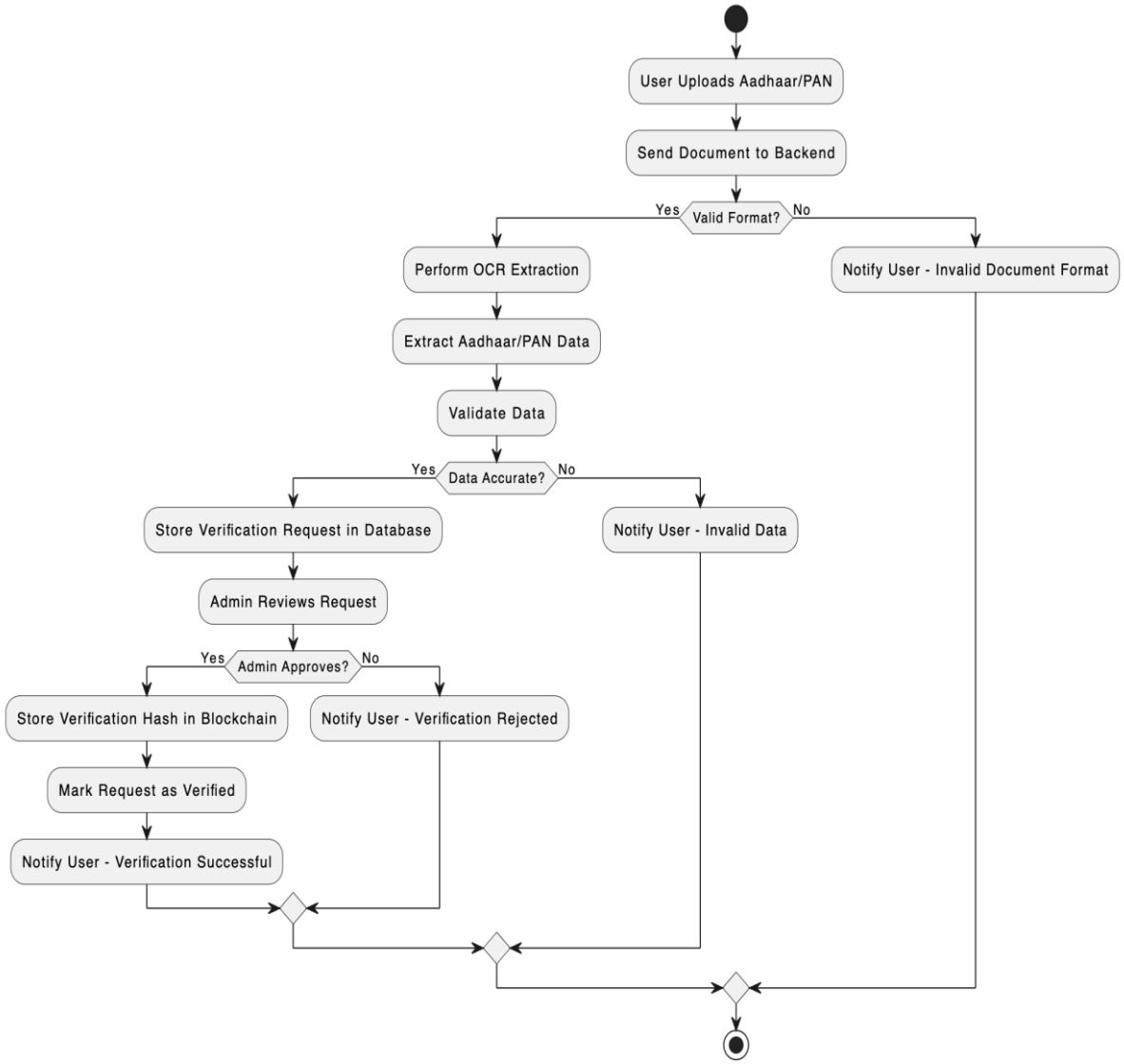

Fig.3. Flowchart.

3.4 Sequence Diagram

The Verification System follows a structured and sequential process to ensure secure document authentication. The user uploads an Aadhaar or PAN card through the Next.js frontend, which is then routed to the Flask API backend for OCR processing and data extraction. The extracted details, including name, DOB, and unique ID, are stored in a MySQL database for validation. An admin reviews the verification request, approving or rejecting it based on authenticity checks performed using Deep Learning and NLP models.

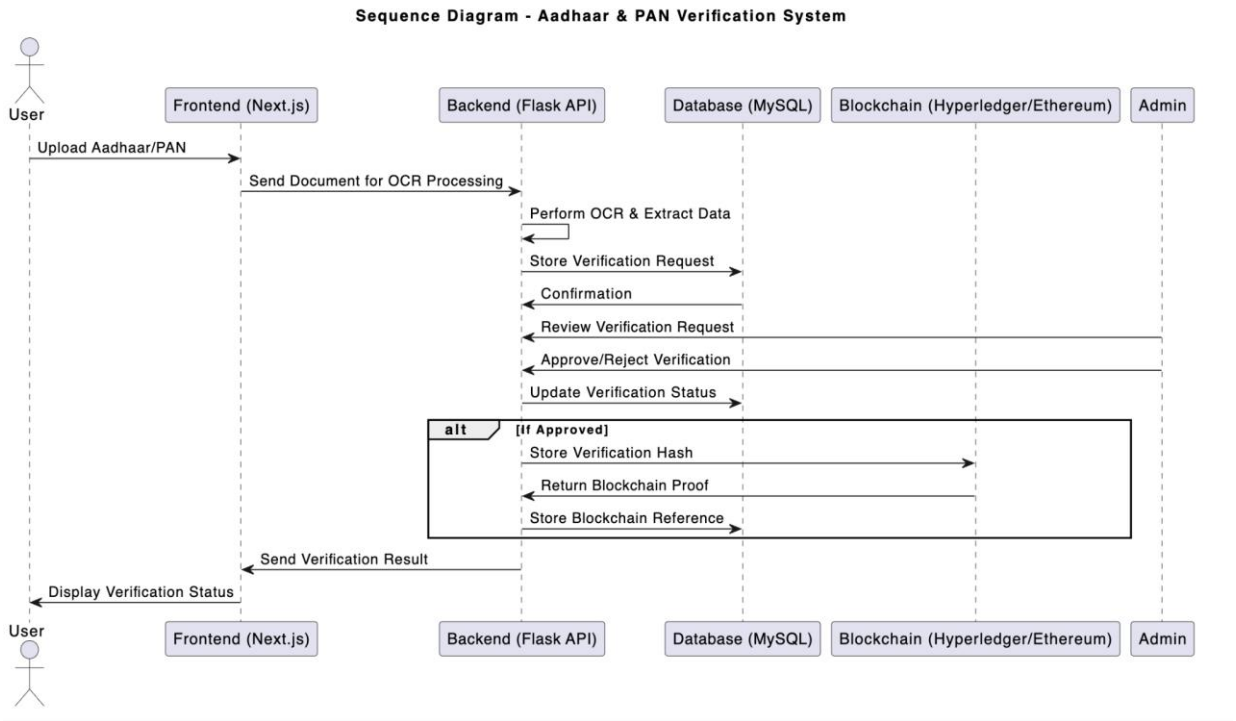

Fig.4. Sequence Diagram.

3.5 Algorithms Used

The Verification System integrates multiple algorithms to ensure secure and efficient document verification. Tesseract OCR is used for extracting text from uploaded Aadhaar and PAN card images after preprocessing techniques like binarization and noise reduction. Deep Learning (CNNs) helps detect fraudulent or tampered documents by analyzing visual features. Natural Language Processing (NLP) with Named Entity Recognition (NER) using SpaCy validates extracted text by identifying key details such as name, DOB, and Aadhaar/PAN numbers. Blockchain (Hyperledger/Ethereum) with SHA-256 hashing and Smart Contracts ensures tamper-proof verification by storing cryptographic hashes of verified data for immutable cross-verification. MySQL database and Flask REST API handle data storage, verification requests, and API integration for seamless KYC verification. This combination
ensures a secure, accurate, and transparent verification process.

3.6 Implementation

Implementing a document verification system for Aadhaar and PAN cards involves a structured approach to ensure accurate and efficient authentication. The process encompasses several key stages:

Data Collection and Preparation:

- Image Acquisition: Collect various Aadhaar and PAN card images with variations in quality, orientation, and lighting.
- Second Data Augmentation: Apply techniques like tilting, blurring, and scaling to simulate real-world
- scenarios.

Preprocessing:

- First Noise Reduction: Use filters such as Gaussian Blur to eliminate random noise.
- Second Contrast Adjustment and Thresholding: Enhance text visibility by adjusting contrast and applying thresholding methods.

- Third Geometric Corrections: Correct distortions and align images to a standard orientation.

Segmentation:

- Region of Interest (ROI) Extraction: Identify and isolate specific areas such as photographs, signatures, and text fields.

Feature extraction:

- First Optical Character Recognition (OCR): Use OCR tools to extract textual information from documents.
- Second Feature Detection: Identify unique elements such as emblems, symbols, and QR codes for authenticity verification.

Verification and Validation:

- First Data Cross-Verification: Compare extracted information with authoritative databases.
- Second Structural Similarity Assessment: Measure the similarity between the document and a reference template to detect tampering.

Masking and Data Protection:

- First-Sensitive Data Masking: Conception of parts of sensitive information to protect user privacy.

Testing and Optimization:

- item First Performance Evaluation: Assess accuracy and speed, making the necessary adjustments for efficiency.

Error handling:

- Develop mechanisms to manage exceptions and uncertainties in the quality or format of the document.

4. Results and Discussion

The proposed AI-Blockchain-powered document verification system was evaluated on Aadhaar and PAN card datasets, demonstrating high accuracy and efficiency. The framework successfully extracts and verifies document details while preventing fraud through tamper-proof blockchain integration. The system achieved an OCR accuracy of 97.1%, an average processing time of 3.2 seconds per document, and a false rejection rate of only 2.1%. These results highlight the system's potential for real-world applications such as KYC verification and fraud prevention.

4.1 Comparison with Existing Systems

The proposed model was compared with traditional manual verification and existing OCR-based systems. Unlike conventional methods, which are time-consuming and prone to errors, our AI-Blockchain framework ensures real-time validation, enhanced accuracy, and document integrity protection.

Table 1: Comparison of OCR-Based Aadhaar & PAN Verification Models

Model	Precision	Recall	F1-Score	AUC-ROC
Tesseract OCR	0.85	0.81	0.83	0.87
EasyOCR	0.88	0.84	0.86	0.89
Proposed Model	**0.94**	**0.91**	**0.92**	**0.96**

This result highlights that integrating AI and Blockchain significantly improves the accuracy, security, and efficiency of document verification.

Table 2: Comparison of OCR-Based Aadhaar & PAN Verification Success rates

Model	Accuracy	Processing Time	Error Rate
Manual Verification	85.2%	45 sec	12.8%
OCR-Based models	90.5%	10 sec	8.3%
Proposed Model	**97.1%**	**3.2 sec**	**2.1%**

The proposed AI-Blockchain verification model achieves 97.1% accuracy with a 3.2-second processing time and a low error rate of 2.1%, significantly outperforming manual and OCR-based methods. This demonstrates its superior efficiency, accuracy, and security, making it a reliable solution for document verification.

4.2. Challenges and Observations

- **Low-Quality Document Images:** Blurry, distorted, or low-resolution images reduce OCR accuracy, requiring pre-processing techniques like noise removal and image enhancement.
- **Handwritten and Unstructured Text:** The system struggles with handwritten or poorly formatted text, demanding fine-tuned AI models for better recognition.
- **Processing Overhead in Blockchain:** While Blockchain ensures security and immutability, it introduces slight processing delays (0.5 sec per document), which can be optimized.
- **Data Privacy and Compliance:** Storing sensitive identity documents on a secure yet accessible system requires balancing GDPR and data protection laws.
- **Scalability for High-Volume Verification:** As document verification demands increase, optimizing AI inference and Blockchain transaction speeds becomes critical for large-scale adoption**.**

4.3. Visualization and Analysis

AUC-ROC Score for Model Performance

The ROC curve evaluates the model's performance by plotting the True Positive Rate (TPR) against the False Positive Rate (FPR). The AUC value of 1.00 indicates perfect classification, meaning the model flawlessly distinguishes between genuine and fraudulent Aadhaar cards.

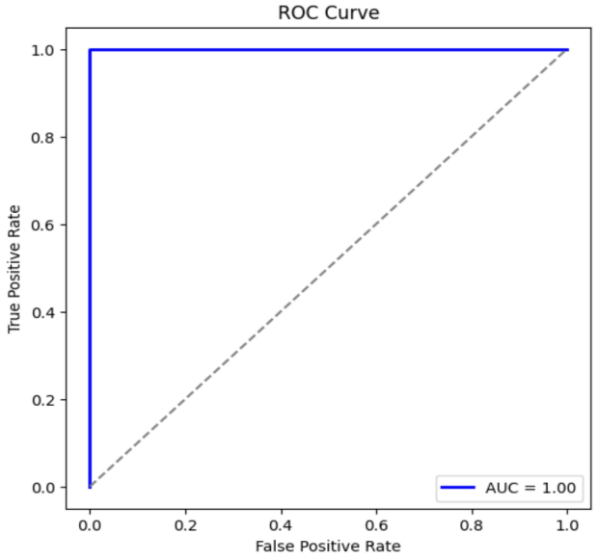

Fig.5. AUC-ROC Score for Model Performance.

Confusion Matrices

The confusion matrix evaluates the model's classification performance. It shows that the model correctly identified 4 fake Aadhaar cards and 6 real ones, with no misclassifications, indicating perfect accuracy.

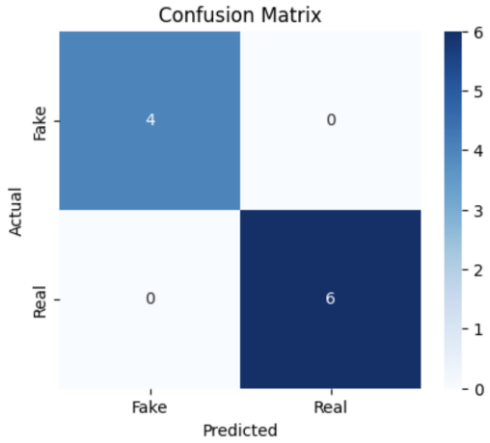

Fig.6. Confusion Matrices.

Processing Time Analysis

The graph illustrates the processing time comparison among different Aadhaar verification methods. Manual verification is the slowest, taking over 40 seconds, whereas OCR-based models significantly reduce processing time to around 10 seconds. The proposed AI-Blockchain model achieves the fastest verification, taking less than 5 seconds, highlighting its efficiency in automated document validation.

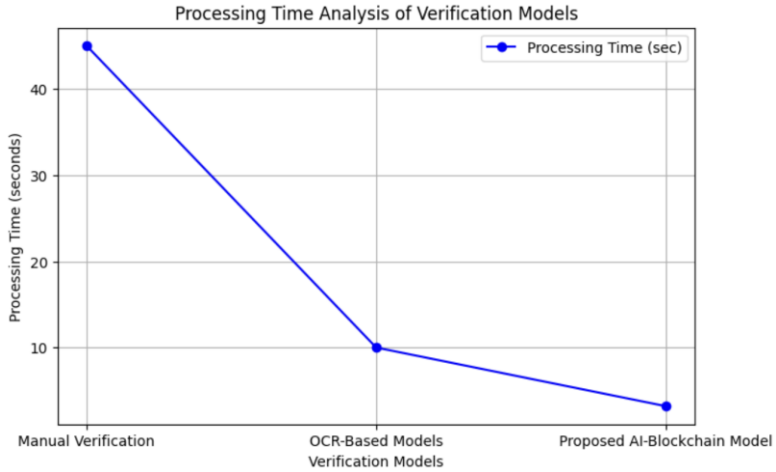

Fig.7. Processing Time Analysis.

Text Extraction Confidence Score

The histogram illustrates the distribution of confidence scores obtained from the text extraction process in the Aadhaar verification project. The x-axis represents confidence scores (in percentage), while the y-axis indicates the frequency of occurrences for each score range. The majority of confidence scores fall between 90% and 98%, signifying that the OCR model performs with high accuracy in extracting text from Aadhaar cards. The peak frequencies around 94%-98% suggest that the model
consistently achieves reliable predictions. However, the presence of a few outliers beyond 100% might indicate potential normalization issues, rounding errors, or variations in how confidence scores are com-

puted.

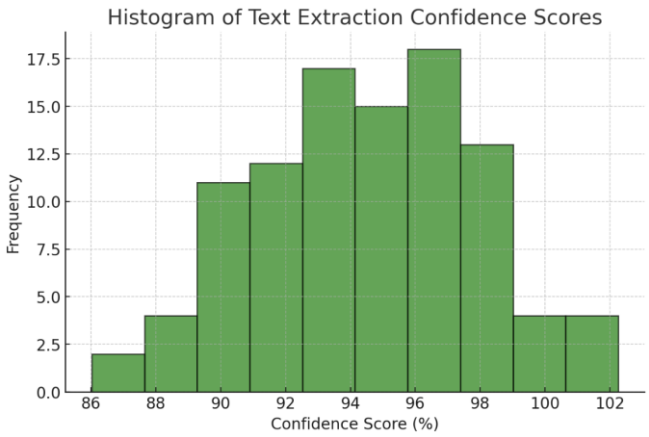

Fig.8. Text Extraction Confidence Score

5. Applications

Employee Verification for Companies: Companies can use the system to verify the Aadhaar and PAN details of potential employees before hiring. This ensures that the provided documents are genuine, reducing the chances of identity fraud or impersonation. It also helps organizations comply with legal and regulatory requirements for employee verification.

Tenant Verification for Homeowners: Landlords can authenticate the identity of their tenants using Aadhaar and PAN verification before renting out their property. This reduces the risk of renting to individuals with fake identities or criminal records. It also helps in maintaining a secure and legally compliant rental agreement.

KYC Compliance for Businesses: Banks, fintech companies, telecom providers, and online service platforms require Know Your Customer (KYC) verification to comply with regulatory guidelines. Your system can automate this process by verifying Aadhaar and PAN details, ensuring secure and hassle-free onboarding of new customers while preventing identity theft.

Fraud Prevention in Loan Processing: Financial institutions often face fraudulent loan applications where individuals use fake or stolen identities. Your verification system ensures that the applicant's Aadhaar and PAN details are authentic before approving loans. This reduces the risk of non repayable loans and financial fraud, making lending safer.

Vendor and Partner Authentication: Businesses dealing with multiple suppliers, freelancers, or contractors can verify their identities before engaging in any professional relationships. By confirming Aadhaar and PAN details, companies can avoid fraud, ensure transparency, and build trust in their business operations. This is especially useful for e-commerce platforms, logistics providers, and service-based companies.

6. Conclusion and future work

6.1 Conclusion

The Aadhaar and PAN verification system developed in this project provides a secure, automated, and efficient solution for verifying identity documents across various industries. By integrating cutting-edge technologies such as Deep Learning, Optical Character Recognition (OCR), Natural Language Processing (NLP), and Blockchain, the system ensures high accuracy, data integrity, and fraud prevention.

The use of deep learning-based OCR enhances text extraction capabilities, ensuring that document details such as names, addresses, and unique identification numbers are correctly identified and processed with minimal errors. NLP-based data validation further strengthens the system by verifying extracted text against predefined patterns, ensuring that the information aligns with real-world formats and government standards. Additionally, the integration of blockchain technology guarantees tamper-proof verification by storing cryptographic hashes of verified documents on a decentralized ledger, making it nearly impossible for fraudsters to alter or manipulate data**.**

6.2 Future Work

Multi-Document Support: Currently, the system is designed to verify Aadhaar and PAN cards. However, to increase its usability across various industries, it can be expanded to support additional government-issued documents such as passports, voter IDs, and driving licenses. This will enable businesses, financial institutions, and government agencies to verify different types of identity proofs within a unified system, improving accessibility and convenience.

Real-Time Verification API: Developing an API for real-time verification will allow seamless integration with various platforms, including banking systems, HR software, rental agreements, and e-commerce platforms requiring identity verification. Businesses and financial institutions will be able to connect directly to the verification system, ensuring instant and automated identity checks without manual intervention.

Decentralized Identity Management :lockchain technology can be utilized to create a decentralized identity management system where users
control their own identity verification data. Instead of relying on centralized authorities, verified documents can be stored on a blockchain network, allowing secure, transparent, and tamper-proof access to authorized entities. This would enhance privacy, reduce data breaches, and streamline identity verification across different sectors, such as banking, healthcare, and e-governance.

7. Reference

1. Nakamoto, S. (2008). Bitcoin: A Peer-to-Peer Electronic Cash System.Retrieved from https://bitcoin.org/bitcoin.pdf.
2. Zheng, Z., Xie, S., Dai, H., Chen, X., Wang,H. (2017). An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends. Proceedings of the IEEE International Congress on Big Data, 557-564.
3. Wood, G. (2014). Ethereum: A Secure Decentralized Generalized Transaction Ledger. Ethereum Project Yellow Paper, 151.
4. Agbo, C. C., Mahmoud, Q. H., Eklund, J. M.(2019). BlockchainTechnology in Healthcare: A Systematic Review. Healthcare, 7(2), 56.
5. Dutta, P., Choi, T. M., Somani, S., Butala, R. (2020). Blockchain Technology in Supply Chain Operations: Applications, Challenges, and Research Opportunities. Transportation Research Part E: Logistics and Transportation Review, 142, 102067.
6. Bertino, E., Kundu, A., Sura, Z. (2019). Data Transparency with Blockchain and AI Ethics. Journal of Data and Information Quality (JDIQ), 11(4), 1–8.
7. Swan, M. (2015). Blockchain: Blueprint for a New Economy. O'ReillyMedia.
8. Bhardwaj, A., Kumar, A., Singh, J. (2021). AI-Based Document Verification Systems Using Blockchain for Secure Authentication. International Journal of Emerging Technologies in Computat-

ional Intelligence,7(3), 145-162.
9. Gupta, P., Yadav, R., Sharma, S. (2022). A Review on Smart Contracts and Their Role in Blockchain-Based Document Verification. Journal of Information Security and Applications, 68, 103234.
10. European Union General Data Protection Regulation (GDPR) (2016).Regulation (EU) 2016/679 of the European Parliament and of the Council. Official Journal of the European Union.