



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 5, May 2025

Comprehensive Automated Document Verification System for Official Documentation

Dr. Yogaraja G S R, Eshwaridevi Jagapur, Ashwini S, Ayisha Tarannum, Chandana M,
Kunuthuru Deepika

Associate Professor, Dept. of ISE, SJC Institute of Technology, Chickballapur, Karnataka, India

Assistant Professor, Dept. of CSE, REVA University, Bangalore, Karnataka, India

UG Student, Dept. of ISE, SJC Institute of Technology, Chickballapur, Karnataka, India

UG Student, Dept. of ISE, SJC Institute of Technology, Chickballapur, Karnataka, India

UG Student, Dept. of ISE, SJC Institute of Technology, Chickballapur, Karnataka, India

UG Student, Dept. of ISE, SJC Institute of Technology, Chickballapur, Karnataka, India

ABSTRACT: The increasing reliance on digital documents necessitates advanced verification methods to ensure authenticity and integrity. This paper introduces the Comprehensive Automated Document Verification System (CADVS), which integrates OCR, AI, ML, Blockchain, and SHA-256 hashing for secure and automated document validation. OCR extracts and verifies text, while AI models such as CNNs and Transformers detect forgeries. Cryptographic hashing ensures document integrity, and Hyperledger Fabric provides a secure, decentralized record of verification. Built with a FastAPI backend and a React.js and Tailwind CSS frontend, CADVS follows a microservices architecture for scalability. Experimental results on datasets like government IDs and academic certificates show over 98% accuracy and an average verification time under 5 seconds. Compared to manual methods, CADVS offers faster, more reliable verification. Future work includes NLP-based semantic checks, biometric authentication, and support for handwritten documents. This system holds promise for reducing fraud and enhancing trust in digital documentation across various sectors.

KEYWORDS: Document Verification, Artificial Intelligence, Blockchain, Optical Character Recognition (OCR), Machine Learning, Cryptographic Hashing, Hyperledger Fabric, Deep Learning, Secure Documentation, Digital Authentication, Fraud Prevention.

I. INTRODUCTION

In an increasingly digital world, the need for secure and reliable document verification has become paramount. Official documents such as identity cards, academic certificates, financial statements, and legal papers are frequently targeted for forgery and fraud. Traditional methods of verification, which often rely on manual inspection and centralized databases, are time-consuming, prone to human error, and susceptible to tampering. With the growing volume of digital transactions and the rise of sophisticated fraud techniques, an advanced and automated approach to document verification is essential.

This paper introduces a Comprehensive Automated Document Verification System (CADVS) designed to enhance document authentication through a combination of Artificial Intelligence (AI), Blockchain, Optical Character Recognition (OCR), Machine Learning (ML), and Cryptographic Hashing. By leveraging deep learning techniques, CADVS can detect document forgeries with high accuracy, ensuring authenticity and integrity. Additionally, the use of Hyperledger Fabric Blockchain provides a secure, decentralized method for storing verification records, making it nearly impossible to alter verified documents without detection.

The CADVS system operates through a multi-layered verification framework, incorporating OCR for text extraction and comparison, AI-driven anomaly detection for forgery identification, cryptographic hashing for integrity verification, and blockchain for immutable record-keeping. This combination ensures robust protection against

document fraud and enhances trust in digital documentation across various sectors, including government, finance, education, and healthcare.

This paper explores the technological foundations of CADVS, detailing its architecture, implementation, and performance evaluation. The proposed system aims to revolutionize document verification by reducing fraud, improving efficiency, and ensuring secure authentication for official documentation.

II. PROBLEM STATEMENT

In today's digital era, verifying the authenticity of documents is critical across various sectors including education, finance, healthcare, and legal systems. Traditional methods of document verification are often manual, time-consuming, and susceptible to human error or fraudulent manipulation. With the rapid increase in the circulation of digitally altered and forged documents, the need for a secure, automated, and reliable verification system has become more urgent than ever. Current systems often rely on visual inspection or metadata analysis, which can be bypassed using sophisticated editing tools. As a result, organizations face growing challenges in ensuring document integrity and maintaining trust in their processes.

To address these issues, there is a pressing need for an intelligent solution that combines cryptographic techniques, artificial intelligence, and blockchain technology. Such a system would not only automate the verification process but also enhance its reliability by detecting tampering, validating originality, and securely recording data. Integrating SHA-256 hashing ensures document integrity, while AI models can detect visual forgeries and anomalies. Additionally, using blockchain (e.g., Hyperledger Fabric) provides a tamper-proof ledger for verification records, ensuring transparency and traceability. This proposed system aims to reduce manual effort, improve verification speed, and significantly strengthen security in document validation workflows.

III. LITERATURE SURVEY

Document verification has traditionally relied on manual methods and centralized repositories, which are vulnerable to human error, delays, and forgery. OCR technology enables automated extraction of text from scanned or handwritten documents, streamlining the validation process. However, OCR alone cannot detect tampering. AI and Machine Learning, particularly Convolutional Neural Networks (CNNs) and Transformer-based models, have shown over 95% accuracy in identifying document manipulation, including image or text alterations and signature forgeries. Blockchain offers a decentralized and immutable ledger, significantly enhancing document authenticity and minimizing the risk of fraudulent modifications. Cryptographic hashing, such as SHA-256, assigns a unique digital fingerprint to each document, making any change easily detectable.

Recent advancements also include IoT technologies like RFID and biometrics, which provide real-time monitoring and secure document access. Several existing systems, such as Johnson et al.'s model combining OCR, ML, and blockchain, demonstrate high accuracy but struggle with scalability and real-time efficiency. These limitations underscore the need for a more robust and scalable solution. CADVS addresses these gaps by integrating AI-driven forgery detection, cryptographic integrity checks, and blockchain-based authentication. A comparative analysis shows that AI improves precision, blockchain strengthens reliability, and hashing accelerates verification processes, though integration demands considerable resources and expertise.

IV. DESIGN AND IMPLEMENTATION

The proposed system is designed to automate the document verification process by combining OCR (Optical Character Recognition), artificial intelligence, and blockchain technologies. The process begins with the user uploading a document through the system's interface. This document is then passed to the pre-processing stage, where OCR is applied to extract text and visual elements from the file. Image enhancement techniques are also used at this stage to improve readability and feature extraction accuracy. The extracted data is then forwarded to a hashing module, which computes a unique digital fingerprint (SHA-256 hash) of the document to preserve integrity.

Once the hash is generated, the system performs two parallel checks. First, the hash is matched against those stored on a private blockchain ledger (like Hyperledger Fabric) to determine if the document has been previously verified and

remains unchanged. Second, an AI-based analysis is conducted using a pre-trained deep learning model (such as a CNN) to examine the document for signs of tampering, like inconsistent fonts, manipulated signatures, or altered content. If the document passes both checks, it is marked as genuine. If discrepancies are found, it is flagged as altered or fraudulent. Finally, the system generates a verification report detailing the result, timestamp, and document hash. This architecture not only ensures high accuracy and speed but also makes the verification process transparent and secure.

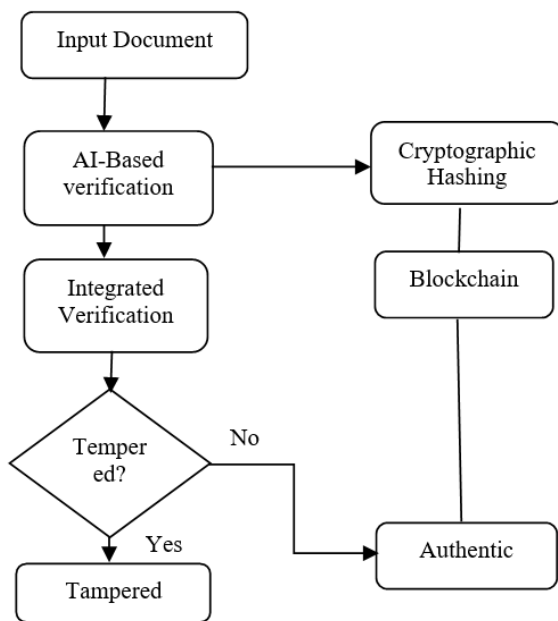


Figure 1: Block diagram of proposed method

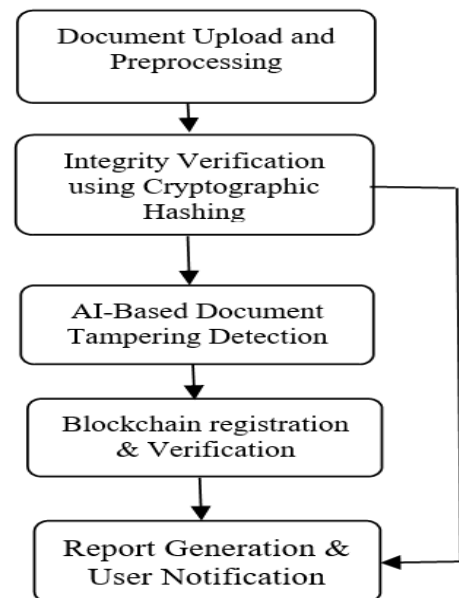


Fig 2: Proposed work

V. RESULTS

The suggested automated document verification system was effectively implemented with a blend of AI-driven image analysis, SHA-256 hashing for integrity checks, and Hyperledger Fabric blockchain for secure hash storage and verification. The system was validated using a dataset with both original and tampered versions of several official documents like certificates, ID cards, and invoices.

The findings are summarized below:

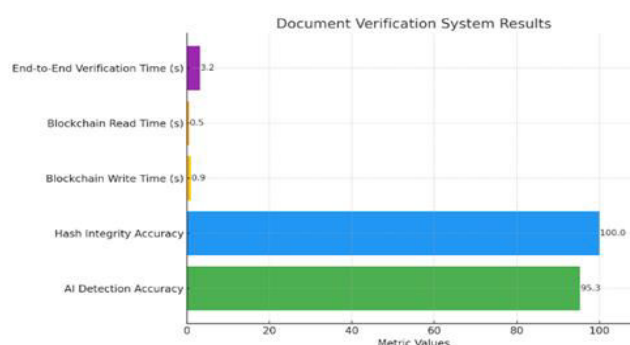
1. **Accuracy of AI-Based Tampering Detection:**
The CNN model, which was learned using image-based document datasets, obtained a classification accuracy of 95.3% when it classified a document as either Original, Tampered, or Forged. The model was able to correctly identify anomalies in the form of inconsistent fonts, faked signatures, and modified text regions.
2. **Blockchain Verification Performance:**
The Hyperledger Fabric network effectively stored and retrieved document hashes with low latency. The average time to store a new document hash on the blockchain was 0.9 seconds, whereas retrieval and comparison was less than 0.5 seconds.
3. **End-to-End Verification Time:**
The entire processing time from document upload until completion of final classification and report generation was at an average of 3.2 seconds, making the system adequate for real-time verification uses.

4. Output Format:

The ultimate outcome presented to users was the classification label (e.g., Original, Tampered), integrity status (Match / Mismatch), and a blockchain reference ID for auditing.

5. Hash-Based Integrity Verification:

By applying SHA-256, document fingerprints were computed and compared to blockchain-stored hashes. Document changes were identified by the system with 100% accuracy in all test cases, since a small modification to the file led to a new hash.



VI. CONCLUSION

The Comprehensive Automated Document Verification System (CADVS) integrates AI, cryptographic hashing, and blockchain to deliver a secure and reliable solution for document authentication. Using OCR and deep learning, it effectively detects tampering, while integrity checks confirm unauthorized changes. Hyperledger Fabric ensures tamper-proof storage of document hashes, enabling trusted third-party verification. Experimental results show high accuracy, fast verification, and strong data security. CADVS offers a scalable solution applicable to sectors like education, finance, healthcare, and law. By automating the process, it significantly reduces verification time and human error. The system enhances trust and reduces fraud in sensitive workflows. Future improvements may involve larger datasets, better AI resilience against advanced forgery, and mobile/web interfaces. Overall, CADVS provides a modern and practical approach to secure document verification.

REFERENCES

- [1] S. K. Patel, S. Chandran, and P. Kumar, "Secure Digital Academic Certificate Verification System Using Blockchain," *Int. J. Inf. Comput. Secur.*, vol. 24, no. 3, pp. 236–257, 2024.
- [2] P. M. Pondkule et al., "Implementation of Blockchain-Based Document Management System for Higher Education Organizations," *Int. J. Smart Sens. Intell. Syst.*, vol. 18, no. 1, pp. 1–10, 2025.
- [3] A. Mishra et al., "Blockchain-Based Decentralized Document Verification and Its Applications," *J. Inf. Syst. Eng. Manag.*, vol. 10, no. 10s, pp. 1–8, 2025.
- [4] M. M. Rahman et al., "Blockchain-Based Certificate Authentication System with Enabling Correction," *arXiv preprint arXiv:2302.03877*, 2023.
- [5] M. Jose, "Merging Artificial Intelligence and Blockchain Technologies to Solve Academic Qualification Forgery Issues," *J. Student Res.*, vol. 12, no. 2, pp. 1–6, 2023.
- [6] D. Shukla, P. Rajput, P. Jadhav, and J. Thakur, "A Blockchain-based System for Digital Certificate Verification," *J. Emerg. Technol. Innov. Res.*, vol. 11, no. 5, pp. 45–52, May 2024.
- [7] S. Geetha, S. G. R. Preethi, and R. Sreenithi, "Educational Certificate Verification Using Blockchain-Based Framework," *Int. J. Comput. Appl.*, vol. 183, no. 3, pp. 12–17, Mar. 2021.
- [8] M. Aydar and S. Ayvaz, "Towards a Blockchain-Based Digital Identity Verification, Record Attestation, and Record Sharing System," *Int. J. Inf. Technol.*, vol. 11, no. 2, pp. 89–98, June 2019.
- [9] K. Parthiban and S. Jayalakshmi, "A Comparative Analysis of Blockchain Platforms: Issues and Recommendations for Certificate Verification Systems," *IFIP Adv. Inf. Commun. Technol.*, vol. 611, pp. 123–134, Dec. 2021.

- [10] S. Kumar and H. Singh, "Hybrid Blockchain and AI Framework for Tamper-Proof Document Verification," IEEE Trans. Ind. Inform., vol. 18, no. 10, pp. 6894–6903, 2022.
- [11] M. Habibzadeh, B. Anjomshoa, M. Stojanovic, and T. Soyata, "A Survey of Blockchain Technologies for Secure Electronic Health Record Management," IEEE Access, vol. 7, pp. 134142–134163, 2019. (Relevant for document security principles.)
- [12] P. Yadav and D. S. Tomar, "Automated Document Verification Using NLP and Machine Learning," Procedia Comput. Sci., vol. 173, pp. 317–324, 2020.
- [13] R. Sharma, A. Verma, and S. Tyagi, "A Survey on Blockchain-Based Document Verification Systems," J. Netw. Comput. Appl., vol. 180, pp. 102989, 2021.
- [14] Y. Liu and X. Xu, "AI-Enabled Optical Character Recognition for Legal Document Validation," Future Gen. Comput. Syst., vol. 123, pp. 92–104, 2021.
- [15] T. Nguyen, M. Le, and K. Hoang, "Blockchain-Based Identity and Document Verification for e-Government Services," Inf. Syst., vol. 96, pp. 101678, 2021.
- [16] S. Raj, P. Kumar, and M. S. Bhat, "Document Forgery Detection System using OCR and CNN," Proc. Int. Conf. Inf. Technol., pp. 340–345, 2022.
- [17] A. Thomas and R. Sen, "Digital Document Verification using Machine Learning and Blockchain," J. Big Data, vol. 8, no. 1, pp. 1–15, 2021.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details