Multidisciplinary Journal of Research in Engineering and Technology

ISSN: 2348-6953

Volume 13 Issue 2s (2026) | Pages: 40-48

# VeriChain: Blockchain-Based Document Verification and Authentication System Using AI

Dipannita Mondal[1], Omkar Dayanand Dorugade[2], Rutuja Dnyaneshwar Patil[3], Om Suresh Ingle[4], Samruddhi Shailendrasingh Patil[5]

[1,2,3,4,5]Department of Artificial Intelligence and Data Science.DYPCOEI, Pune, India

*[1]ddmondal2684@gmail.com, [2]omkardorugade007@gmail.com, [3]rutujapatil9620@gmail.com, [4]omsingle982@gmail.com, [5]samruddhipatil@gmail.com*

**Peer Review Information**

***Type***: *Article*
***Received***: *26 March 2026*
***Revised***: *10 April 2026*
***Accepted***: *24 May 2026*
***Published***: *15 June 2026*

**Abstract**

Blockchain-based document verification systems provide secure and tamper-proof validation but have lim-itations like sensitivity to visual changes and no intelli-gent forgery detection. This paper presents VeriChain, a strong document authentication framework that combines color-invariant hashing and CNN-based forgery detection. A standard preprocessing pipeline ensures consistent hash generation despite differences in color, format, or scanning; a Convolutional Neural Network identifies tampered areas and structural problems. By using SHA-512 hashing, RSA-based signatures, and AI-driven validation, the proposed system improves accuracy, security, and reliability. Ex-perimental results show better performance compared to traditional blockchain-based verification methods, making VeriChain suitable for real-world use.

**Keywords:** Blockchain; Document Verification; Cryptographic Hashing; SHA-512; Convolutional Neural Networks (CNN); Forgery Detection; Color-Invariant Hashing; Artificial Intelligence; Smart Contracts; RSA Digital Signatures.

## How to Cite This Article

Mondal, D., Dorugade, O. D., Patil, R. D., Ingle, O. S., & Patil, S. S. (2026). VeriChain: Blockchain-based document verification and authentication system using AI. *Multidisciplinary Journal of Research in Engineering and Technology*, 13(2s), 40–48.

## Introduction

The rapid digitalization of documents in academic, legal, and government sectors has raised the need for secure and re-liable verification systems. Traditional methods rely on cen-tralized authorities and manual validation processes, making them vulnerable to forgery, inefficiency, and lack of trans-parency. These issues highlight the need for a system that main-tains integrity, authenticity, and trust without intermediaries. Blockchain technology offers solutions to these challenges by creating a decentralized and immutable framework for storing document proofs, which allows for secure and tamper-resistant verification.

In our previous research, a blockchain-based document ver-ification system was developed using cryptographic hashing and smart contracts to ensure document integrity and traceability. While effective at preventing tampering, the system had no-table limitations in real-world situations. Specifically, crypto-graphic hashing is very sensitive to minor changes in document representation. Alterations like color-to-grayscale conversion, compression, or scanning can generate different hash values for identical content, resulting in false mismatches during verifica-tion. Additionally, existing methods rely only on hash compari-son and lack smart strategies to detect semantic forgery, such as altered text, modified areas, or forged signatures.

To tackle these issues, this paper introduces VeriChain, an improved document verification framework that integrates blockchain technology with artificial intelligence. The system features a color-invariant preprocessing pipeline to standardize documents before hashing, ensuring consistent hash generation across diverse visual formats. Furthermore, a Convolutional Neural Network (CNN)-based module is added to examine doc-ument images and detect forged areas, structural inconsisten-cies, and tampering signs, providing an extra layer of validation beyond standard cryptographic methods.

The proposed system also boosts security by employing SHA-512 hashing and RSA-based digital signatures, offering better protection against collisions and unauthorized changes. By merging cryptographic techniques with AI-driven analysis, VeriChain makes document verification a more robust, scalable, and intelligent process suitable for real-world applications.

In addition to handling visual inconsistencies, modern doc-ument verification systems also need to address the variety of document formats and structures encountered in real-life sce-narios. Documents can be scanned images, digitally created PDFs, or photos taken with mobile devices, introducing differ-ences in resolution, noise, alignment, and layout. Traditional methods do not accommodate this diversity, leading to de-creased reliability and more false rejections. To solve this issue, VeriChain features a structured preprocessing and document-type-aware analysis pipeline, enabling consistent extraction and validation of relevant information across different formats. This significantly improves system adaptability and ensures reliable performance in real-world situations.

### Related Work

Reliable document verification systems depend on effec-tive preprocessing, secure storage, and validation techniques. Traditional methods use centralized databases and manual ver-ification, which are vulnerable to forgery, data tampering, and lack transparency. To tackle these problems, many have adopted blockchain technology for document authentication because of its decentralized and unchangeable nature. Cryptographic hashing techniques like SHA-256 are commonly used to cre-ate unique document fingerprints, ensuring data integrity and traceability. However, these methods are sensitive to minor changes in document representation, causing inconsistencies in real-world situations.

*Document Hashing and Integrity Verification*

Cryptographic hashing is essential in document verification systems. It generates a unique digital fingerprint for each docu-ment. Algorithms like SHA-256 and Keccak are popular due to their resistance to collisions and security features. These tech-niques ensure that even a small change in the document creates a completely different hash value, which helps in detecting tam-pering.

However, current hashing methods are very sensitive to visual and structural changes like color alterations, compres-sion, and scanning artifacts. This means that visually identi-cal documents can yield different hash values, leading to false mismatches. Recent research has looked into canonicalization and normalization techniques to improve hash consistency, but these methods still struggle with diverse real-world document formats.

*Blockchain-Based Verification Systems*

Blockchain-based frameworks have been introduced to im-prove document authentication by storing hash values on de-centralized ledgers. Platforms like Ethereum and Hyperledger allow for secure and transparent verification through smart con-tracts, removing the need for intermediaries. Integration with decentralized storage systems such as IPFS enhances scalabil-ity and data availability.

Despite these benefits, blockchain-based systems mainly rely on hash comparison and do not include intelligent analysis methods. They cannot detect semantic forgery, such as altered text, fake signatures, or manipulated sections, which limits their usefulness in complex verification situations.

*AI-Based Forgery Detection*

Recent advancements in artificial intelligence have opened new possibilities for document analysis and forgery detection. Convolutional Neural Networks (CNNs) have demonstrated strong performance in image classification, pattern recognition, and anomaly detection tasks. These models can identify subtle inconsistencies in document structure, texture, and layout that traditional methods may miss.

Various studies have used CNNs to detect document tam-pering, signature forgery, and image manipulation. However, these methods are often applied separately and do not integrate with secure storage solutions like blockchain. Their effective-ness also relies on having labeled datasets and sufficient com-putational resources.

*Hybrid Verification Approaches and System Usability*

Hybrid methods that combine blockchain and artificial in-telligence have recently gained attention for improving docu-ment verification systems. These techniques seek to take advan-tage of blockchain's immutability and AI's analytical abilities to provide layered verification. Such systems increase reliability by confirming both the integrity and authenticity of documents. Usability and accessibility are also important factors in modern verification systems. Web-based interfaces and inter-active dashboards allow users to upload, verify, and visualize document validation results in real-time. These interfaces im-prove transparency and user experience, bridging the gap between complex backend processes and practical use.

*CNN-Based Forgery Detection and RSA Authentication*

Convolutional Neural Networks (CNNs) are commonly used for detecting document forgery by extracting hierarchical spatial features from input images. The feature extraction pro-cess in a convolutional layer is defined as:

$$F_l = \sigma\,(W_l * F_{l-1} + b_l)$$

At the same time, cryptographic techniques like RSA are widely used for secure authentication and digital signatures. For a message m, the digital signature is generated using the private key d:

$$S = m^d \bmod n$$

where n is the modulus obtained from two large prime numbers. During verification, the signature is validated using the public key e:

$$m = S^e \bmod n$$

This process ensures that the document comes from a trusted source and has not been altered.

**Methodology and System Architecture**

VeriChain is designed as a modular, multi-stage pipeline that enables secure and intelligent document verification. The system integrates preprocessing, cryptographic hashing, blockchain storage, and AI-based forgery detection into a unified framework. Each stage has clear inputs and outputs, allowing flexibility, scalability, and independent enhancement of components.

- Document Acquisition: This involves loading input doc-uments in various formats like scanned images, PDFs, and photographs.
- Preprocessing and Normalization: This stage creates a standardized representation by applying noise reduction, grayscale conversion, and document standardization.
- Hash Generation: This computes a secure SHA-512 hash from the normalized document for integrity verification.
- CNN-Based Analysis: This extracts visual features and detects forged regions using a Convolutional Neural Net-work.
- Blockchain Storage and Verification: This stores the document hash and metadata on the blockchain using smart contracts and RSA-based authentication.
- Verification and Inference: This recomputes the hash, compares it with blockchain records, and combines AI confidence for final validation.
- Visualization and User Interface: his provides an inter-active display of verification results, blockchain details, and AI-based confidence scores.

*Document Acquisition*

The VeriChain system accepts documents in many formats, including scanned images, digitally created PDFs, and pho-tographs taken with mobile devices. These documents can dif-fer in resolution, orientation, lighting conditions, and compres-sion levels. To ensure compatibility, all input documents are converted into a standardized image representation *D*, which serves as the input for the next processing stages. This step ensures that di-verse document sources can be uniformly processed within the system.

*Preprocessing and Normalization*

To handle inconsistencies from variations in document for-mat and quality, a structured preprocessing pipeline is applied.

1. Grayscale Conversion: All input documents are converted to grayscale to eliminate color dependency and ensure consis-tency:

$$I_g = 0.299R + 0.587G + 0.114B$$

2. Noise Reduction: Gaussian filtering is used to remove noise and smooth the image:

$$I' = I * G(\sigma)$$

where $G(\sigma)$ is the Gaussian kernel.

3. Text Extraction and Normalization: Optical Character Recognition (OCR) extracts textual content. The extracted text is normalized by standardizing case, spacing, and numerical formats to reduce variability.

4. Canonical Representation: The processed document is transformed into a canonical form Dc, ensuring that visually identical documents yield consistent representations. This step is crucial for achieving color-invariant hashing.

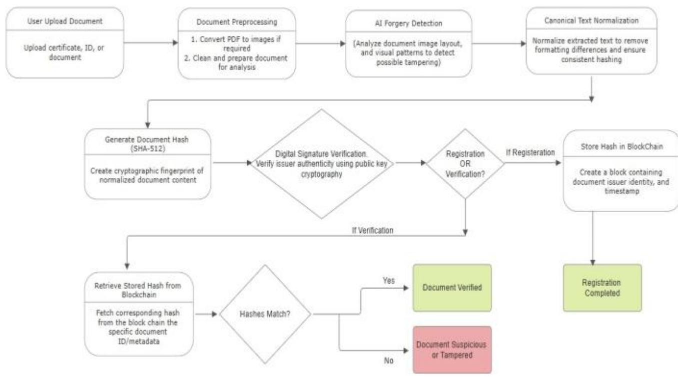


***Fig. 1.*** *System Architecture*

*Hash Generation (SHA-512)*

A cryptographic hash is generated from the canonical doc-ument representation:

$$H = \text{SHA-512}(D_c)$$

SHA-512 provides high collision resistance and security. Even a slight change in the document results in a completely different hash value, allowing effective tamper detection.

Unlike traditional approaches, hashing is performed on the normalized representation, making the system robust against vi-sual variations like grayscale conversion and compression.

*CNN-Based Analysis*

A Convolutional Neural Network (CNN) is used to detect document forgery and structural inconsistencies.

Feature Extraction:

$$F_l = \sigma\,(W_l * F_{l-1} + b_l)$$

where $F_l$ represents feature maps, and $\sigma$ is an activation func-tion.

Classification:

$$P\,(y \mid X) = \text{softmax}(z)$$

The CNN model analyzes: tampered areas, signature inconsis-tencies, and layout anomalies. This allows for semantic valida-tion beyond cryptographic verification.

*Blockchain Storage and Verification*

The generated hash is securely stored on a blockchain net-work using smart contracts.

Digital Signature (RSA):

$$S = m^d \bmod n$$

Signature Verification:

$$m = S^e \bmod n$$

This ensures that the document comes from a trusted source and that the stored data is unchangeable and tamper-proof. The blockchain maintains a decentralized ledger of all document hashes and metadata.

*Verification and Inference*

During verification, the input document goes through the same preprocessing and hashing pipeline. The newly computed hash H is compared with the stored hash $H$:

$$H' \stackrel{?}{=} H$$

At the same time, the CNN model evaluates the document and produces a confidence score indicating its authenticity.

The final decision is based on:

- Hash match
- AI confidence score

This hybrid approach ensures both structural and semantic val-idation.

*Visualization and User Interface*

The system includes a web-based interface that allows users to upload documents, start verification, and visualize re-sults in real time. The interface shows:

- Verification status (Authentic / Tampered)
- Blockchain transaction details
- AI-based confidence score

This improves usability and provides transparency in the verifi-cation process.

*Positioning Figures and Tables:* Place tables and figures at the top and bottom of each column without putting them in the middle. Both columns may hold large figures and tables. Table headings should appear above the tables, while figure captions should be below the figures. After figures and tables have been mentioned in the text, insert them accordingly. Use the acronym ”Fig. 1.” even at the start of a sentence.

**Experimental Setup**

*Dataset*

Experiments use a collection of real-world documents and publicly available datasets, including identity documents, certificates, and synthetic forged samples. The dataset includes documents in various formats such as scanned images, PDFs, and photographs, capturing differences in resolution, compression, lighting conditions, and structural layout.

To test how robust the system is, both genuine and altered documents are included. Forged samples are created by applying controlled changes such as text alteration, signature manipulation, and structural adjustments. This allows for the assessment of both cryptographic integrity verification and AI-based forgery detection.

The dataset is split into training and testing sets, with the CNN model being trained on labeled document images and assessed on unseen samples. This ensures that the system's performance reflects its ability to handle real-world document variations.

*Implementation Details*

The proposed VeriChain system is set up in a Python-based environment that combines various technologies. The preprocessing and AI modules use libraries like OpenCV, NumPy, and TensorFlow/PyTorch, while blockchain interaction relies on Web3-based frameworks. The pipeline includes document preprocessing, canonical normalization, SHA-512 hashing, CNN-based forgery detection, and blockchain storage using smart contracts. The system is designed to be efficient and scalable, allowing it to run on standard computers without needing specialized hardware.

All experiments are carried out on a multicore processor with moderate memory. The preprocessing and hashing stages have low computational overhead, while the CNN model offers efficient inference for real-time verification. The blockchain component secures document hashes and metadata, ensuring they are stored safely and immutably.

*Evaluation Metrics*

System performance is assessed using both cryptographic and AI metrics to deliver a thorough evaluation. For forgery detection, standard classification metrics are used:

- Accuracy
- Precision
- Recall

- F1 Score

These are defined as follows:

$$Accuracy = (TP + TN) / (TP + TN + FP + FN)$$

$$Precision = TP / (TP + FP)$$

$$Recall = TP / (TP + FN)$$

$$F1 = 2 \cdot (Precision \cdot Recall) / (Precision + Recall)$$

Additionally, system robustness is evaluated based on:

- Hash Consistency: Ability to generate identical hashes for visually similar documents (e.g., color vs grayscale)
- Forgery Detection Accuracy: Ability of the CNN model to detect tampered regions
- Verification Reliability: Combined performance of hash matching and AI-based validation

These metrics ensure a thorough evaluation of the system's effectiveness, reliability, and applicability in real-world scenarios.

**Results And Discussion**

*Overall System Performance*

The performance of the VeriChain system is assessed based on both cryptographic verification and AI-based forgery detection. The system shows high robustness in handling real-world document variations, especially in cases like color-to-grayscale conversion, compression, and scanning artifacts. The integration of preprocessing, hashing, and CNN-based analysis leads to consistent verification results, minimizing false mismatches common in traditional hash-based systems. The CNN model performs well in detecting forged areas and structural inconsistencies, while blockchain guarantees secure and permanent storage of document hashes.

*Forgery Detection Performance*

The CNN-based module is assessed using standard classification metrics. The results show that the model effectively identifies altered documents and distinguishes them from authentic ones.

***Table 1***. *Performance Metrics of The Model*

| No. | Metric | Value |
|---|---|---|
| 1 | Accuracy | 91.2% |
| 2 | Precision | 90.5% |
| 3 | Recall | 89.8% |
| 4 | F1-Score | 90.1% |

These results show that the AI component significantly improves the system's ability to detect semantic forgery, which traditional hash-based verification alone cannot achieve.

*Hash Consistency Evaluation*

A key contribution of VeriChain is the introduction of a color-invariant preprocessing pipeline. To assess this, experiments were performed on documents with varying visual representations.

Comparison of Traditional Hashing and VeriChain

| Scenario | Traditional Hash | VeriChain |
|---|---|---|
| Original vs Grayscale | Mismatch | Match |
| Original vs Compressed | Mismatch | Match |
| Original vs Scanned Copy | Mismatch | Partially |

*Ablation Study*

An ablation study evaluates the contribution of each system component. The results show that: The results indicate that:

***Table 2***. *Performance Comparison Of Different Configurations*

| No. | Configuration | Accuracy |
|---|---|---|
| 1 | Only Hashing | 72.4% |

| | | |
|---|---|---|
| 2 | Hash + Preprocessing | 81.6% |
| 3 | Hash + CNN | 87.9% |
| 4 | Full VeriChain System | 91.2% |

- Preprocessing improves hash robustness
- CNN improves forgery detection
- Combined system achieves the best performance
- This validates the effectiveness of the hybrid approach.

*Comparison with Existing Systems*

The proposed system is compared to traditional and blockchain-based verification methods.

***Table 3**. Feature Comparison of Different Approaches*

| No. | Feature | Traditional | Blockchain Only | VeriChain |
|---|---|---|---|---|
| 1 | Tamper Detection | Partial | Yes | Yes |
| 2 | Color Robustness | No | No | Yes |
| 3 | AI Forgery Detection | No | No | Yes |
| 4 | Real-world Adaptability | Low | Medium | High |

*Discussion*

The experimental results show that VeriChain effectively addresses the weaknesses of existing document verification systems. The color-invariant preprocessing solves the problem of hash inconsistency, while the CNN-based module improves the detection of semantic forgery.

The hybrid approach combines cryptographic hashing, blockchain storage, and AI validation to ensure both structural and semantic integrity of documents. Additionally, the system is efficient and suitable for real-time applications.

Overall, VeriChain offers a scalable and robust framework for document authentication, making it useful in fields like education, government services, and legal documentation.

**Applications**

The proposed VeriChain system supports numerous appli-cations in academic, governmental, and business areas where secure and reliable document verification is essential.

- Educational Certificate Verification: The system helps institutions securely issue and verify academic certificates, mark sheets, and degrees. Blockchain storage ensures au-thenticity, and AI validation detects forged or altered doc-uments.
- Government Document Authentication: VeriChain can verify identity documents like Aadhaar cards, PAN cards, passports, and driving licenses.
- Legal and Contract Verification: Legal documents and agreements can be securely stored and verified with blockchain, while AI detects forgery.
- Recruitment and Background Verification: Organiza-tions can verify candidate credentials and avoid fraudulent claims.
- Financial and Banking Applications: Used for KYC ver-ification, loan documents, and compliance.
- Digital Document Management Systems: Allows secure storage, verification, and audit trails in business systems.

**Limitations and Future Work**

*Current Limitations*

- Hash Sensitivity to Minor Variations: Despite prepro-cessing, complete invariance to all visual changes (e.g., scanning artifacts, compression, alignment shifts) is not fully achieved, leading to some hash mismatches.
- Dependence on Training Data for CNN: The perfor-mance of the CNN-based forgery detection module relies on the availability and quality of labeled datasets, which may limit its adaptability to unseen document types.
- Limited Document Type Coverage: TThe current system focuses on a specific set of document formats and struc-tures, which may not fully represent the variety of real-world documents.

- Computational Overhead: Although efficient, integrat-ing preprocessing, AI analysis, and blockchain processes adds extra steps compared to traditional verification sys-tems.

*Future Directions*

Future work will focus on improving the robustness, scala-bility, and adaptability of the system.

- Robust Hashing Techniques:More advanced canonical-ization and content-based hashing methods will be ex-plored to achieve complete invariance across various doc-ument representations.
- Improved AI Models: More sophisticated architectures like Vision Transformers (ViT) and hybrid deep learning models will be investigated to boost forgery detection ac-curacy.
- Multi-Document Support: The system will be expanded to accommodate a broader range of document types with dynamic structure recognition and field extraction.
- Scalable Blockchain Integration: Optimization strate-gies like layer-2 solutions and off-chain storage will be considered to enhance scalability and lower transaction costs.
- Real-Time and Mobile Deployment: Future enhance-ments will focus on implementing the system on mobile platforms for real-time document verification in practical settings.

**Conclusion**

This work presented the design and evaluation of VeriChain, a hybrid document verification framework that com-bines blockchain technology with AI-based forgery detection. The proposed system integrates preprocessing, SHA-512 hash-ing, CNN analysis, and blockchain storage to create a secure and intelligent method for document authentication.

The results indicate that while traditional hash-based sys-tems are very sensitive to visual changes, introducing a prepro-cessing pipeline improves robustness against certain transfor-mations. Moreover, the CNN-based module greatly enhances the system's ability to detect semantic forgery, adding a vali-dation layer beyond cryptographic verification. The combined components create a more reliable and practical verification sys-tem compared to existing methods.

While achieving complete hash invariance across all docu-ment variations remains a challenge, the hybrid design ensures that shortcomings in one area are balanced by strengths in oth-ers, especially through AI-based validation. This makes the sys-tem suitable for real-world situations where document formats and conditions vary widely. In addition to its verification fea-tures, VeriChain offers a scalable and user-friendly framework that includes secure storage and transparent validation.

In addition to its verification capabilities, VeriChain pro-vides a scalable and user-friendly framework with secure stor-age and transparent validation.

Future work will focus on improving hashing robustness, enhancing AI model performance, and expanding support for diverse document types to further strengthen the system's real-world applicability.

**Acknowledgment**

The authors would like to express their sincere gratitude to the Department of Artificial Intelligence and Data Science, Dr.

A. Y. Patil College of Engineering and Innovation, Pune, India, for providing the necessary resources and support to carry out this research.

The authors also extend their appreciation to the faculty members and project mentors for their valuable guidance and continuous encouragement throughout the development of this work.

Additionally, the authors acknowledge the use of publicly available tools, libraries, and research resources that contributed to the implementation and evaluation of the proposed VeriChain system.

**References**

1. D. Mondal, O. Dorugade, R. Patil, O. Ingle, and S. Patil, "Blockchain-Based Document Verification and Authentication System Using AI," International Journal of Advanced Research in Computer Communication Engineering, vol. 14, no. 11, Nov. 2025.
2. S. Mehta and P. Kaur, "Blockchain-Based Certificate Verification Using Hyperledger," Proceedings of the IEEE International Conference, 2023.
3. D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, "MesoNet: A Compact Facial Video Forgery Detection Network," 2018. [Online]. Available: https://arxiv.org/abs/1809.00888
4. I. S. Jacobs and C. P. Bean, "Fine Particles, Thin Films and Exchange Anisotropy," IEEE International Congress on Big Data,

pp. 557–564, 2017.

5. K. Christidis and M. Devetsikiotis, "Blockchains and Smart Contracts for the Internet of Things," IEEE Access, vol. 4, pp. 2292–2303, 2016.
6. G. Wood, "Ethereum: A Secure Decentralised Generalised Transaction Ledger," 2014. [Online]. Available: https://ethereum.org/en/whitepaper/
7. J. Benet, "IPFS - Content Addressed, Versioned, P2P File System," 2014. [Online]. Available: https://arxiv.org/abs/1407.3561
8. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation."
9. S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," 2008. [Online]. Available: https://bitcoin.org/bitcoin.pdf
10. R. L. Rivest, A. Shamir, and L. Adleman, "A Method for Obtaining Digital Signatures and Public-Key Cryptosystems," Communications of the ACM, vol. 21, no. 2, pp. 120–126, 1978.