Blockchain-Based E-Vault System with Deep Learning for Secure Document Management

Neha Unnisa1,a*, Dr. Anjaiah Adepu2,b*

[1]Ph.D. Scholar & Assistant Professor, Department of Computer Science & Engineering,BESTIU, Anantapur, Andhra Pradesh, India,515231

[2]Head of the Department,CSE & Dy. Proctor Department of Computer Science & Engineering,Maulana Azad National Urdu University, Polytechnic, Darbhanga, Bihar, 846002

[a]neha.unnisa@gmail.com*; [b]anjaniprasad.adepu@gmail.com

Abstract:
On matters of legal documentation, contract, deed, or even court affidavit, utmost security, authenticity, and availability are demanded. Being centralized in nature, tampering, unauthorized access, and single-point failures threaten traditional storage systems. Thus, this study provides a blockchain-integrated eVault framework that synthesizes cryptographic security with a deep learning-based verification approach. The eVault software uses InterPlanetary File System (IPFS) for decentralized storage, AES-256 encryption standards for confidentiality, and Ethereum smart contracts for access control and audit logging at the finest granularity. Forgery in scanned document images is detected by a CNN model, and the integrity is ensured by SHA-256 hashing. Both metadata and scores of authenticity are immutably stored in the blockchain, making the transaction and traceability transparent. With an experimental evaluation done for a file size of 5 MB, forgery detection accuracy was proven to be 97.5%, whereas access latency was under 200 ms. Thus, this solution guards against the hinderance of the illegal forgery and illegal access, providing a secure, scalable, and intelligent infrastructure for legal document management.

Keywords: Blockchain, Deep Learning, IPFS, Smart Contracts, Document Authentication, Legal Records

I. INTRODUCTION

Legal documents like contracts, property deeds, affidavits, and court records are the main core of judicial, financial, and institutional processes. The bigger concern is that the larger the alteration, forgery, or even slight tampering with such a document, the bigger the financial losses, disputes, or suits. Traditionally, they have been stored in centralized databases or cloud-based platforms. These systems make the records accessible yet vulnerable to unauthorized access, tampering, insider threats, and single points of failure. Besides, traditional storage systems do not offer transparent mechanisms for verifying document validity, so they are prone to fraud and manipulation.In the last decade, blockchain has been hyped as a panacea for incorruptible and decentralized storage. Its immutable ledger with distributed consensus is intended to eliminate single points of failures and ensure transparency and traceability of operations. Through a similar lens, IPFS has been chosen as the decentralized storage protocol to tackle the scaling issues faced by traditional ways. Cryptographic algorithms such as AES-256 for encryption and SHA-256 for hashing guarantee document confidentiality and integrity.With the guarantee of secure access and storage being granted by blockchain, the authentication of the very documents to be stored lies outside it. This gap can be filled—and intelligent document authentication can be achieved—with the help of deep learning techniques. CNNs work very well in finding visual forgeries on scanned documents with tampered seals, altered signatures, or changed layouts. On the other hand, a combination of blockchain and deep learning would make the

outlined hybrid system one that protects documents and authenticates them before storage.The paper presents a secure and intelligent eVault system integrating blockchain, IPFS, AES-256 encryption, and CNN-based forgery detection. The proposed framework guarantees confidentiality, integrity, authenticity, and reliance on the track, thus offering a scalable and reliable solution for the present-day legal document management.

II. LITERATURE SURVEY

Blockchain technology has been considered to provide integrity, transparency, and decentralization to sensitive data that are administered in healthcare, governance, and land administration. Some recent work has reviewed how such integration with AI, privacy mechanisms, and distributed infrastructures can overcome the insufficiencies of traditional systems.Akbarfam and Maleki suggested that deep learning and blockchain be integrated for automated access control in a manner that adaptive decision-making is engaged together with immutable security. Likewise, Sunny et al.[5] conducted a review on blockchain applications and listed advantages such as decentralization and trust, while putting forth scalability and compliance as challenges. Legal and privacy concerns have been likewise studied. Belen-Saglam et al. investigate the contradiction with GDPR mandates of public blockchain immutability, whereas Xu and Zhang point out privacy concerns with blockchain-based data management and recommend privacy-preserving techniques, including homomorphic encryption.The focus is on a few systems that are practical implementations. Again, Hossen et al. describe an application that combines the use of a blockchain for securing land records with Hyperledger Fabric, enhancing transparency and reductions in fraud. Similar approaches were taken by Pereira et al. and Suganthe et al., who proposed frameworks for blockchain-based land registration and land administration, respectively, allowing for secure, tamper-proof digital recording of land information. In criminal justice, Singh et al.conceptualized a blockchain-based criminal case recording system, preventing tampering and ensuring evidence is recorded in a reliable manner. Javed et al. gave a proposal called Health-ID that serves as a decentralized identity management system for ensuring secure access to medical data.In the governance sector, Cagigas et al.systematically reviewed the use of blockchain in public services, reporting the realization of gains in efficiency and in citizen trust, while also pointing to interoperability and institutional readiness as barriers.Thus, all these studies have shown how blockchain has transformed secure record management. Unresolved issues related to privacy compliance, scalability, and alignment with regulations call for further inquiry into hybrid models and advanced cryptographic solutions.Thus, all these studies have shown how blockchain has transformed secure record management. Unresolved issues related to privacy compliance, scalability, and alignment with regulations call for further inquiry into hybrid models and advanced cryptographic solutions.

III. PROPOSED WORK

The proposed study proposes a secure and intelligent eVault system for legal document management integrating blockchain technology, deep learning, cryptography, and decentralized storage. Legal documents such as contracts, deeds, and affidavits require high confidentiality, integrity, and authenticity, which traditional centralized storage systems fail to guarantee. Adopting AES-256 encryption, SHA-256 hashing, Ethereum smart contracts, IPFS storage, and CNNs into a single framework, eVault sets out to resolve these issues.

The user uploads a legal document as the first step in the system workflow. Prior to storage, it undergoes forgery detection using a CNN that has been trained on genuine and tampered legal documents. The CNN examines certain visual features of the document, such as signatures, seals, and layout, for authenticity. Documents identified as forged are rejected, thereby ensuring that tampered records cannot enter the system. The document is then encrypted with AES-256 and stored in IPFS, a decentralized storage file system. Thereafter, the SHA-256 hash of the document and its metadata, including user ID, timestamp, and authenticity score, are immutably stored on the blockchain.Access control is imposed through Ethereum-based smart contracts implementing the ABAC-policy system. Thus, only allowed users endowed with proper attributes may grab and decrypt these documents. Any

attempt at access, be it accepted or denied, shall be duly recorded by the blockchain for the auditability of the precise case, ensuring complete transparency and traceability.

This eVault framework brings several advantages, including tamper-proof storage, automatic forgery prevention, high-level confidentiality and data integrity, real-time access with low latency, and scalable legal records management. Combining the immutable ledger and deep learning-based authentication, this system is a strong, automated, and trustworthy environment for handling legal documents in the modern world, thus massively reducing instances of fraud, unauthorized access, and downtime of manual verification.

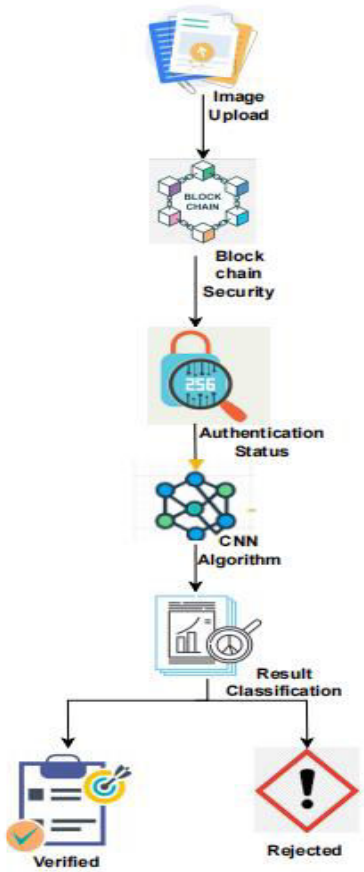

Fig1 : Proposed Architecture Diagram

IV. METHODOLOGY

In this eVault system, blockchain technology with AI, cryptography, and decentralized storage techniques are combined to develop a secure and intelligent tamper-proof legal-document management solution. The working method is divided into the following phases:

Users register and authenticate themselves with a secure blockchain wallet interface, such as MetaMask. Bitcoin-style public-private key pairs are generated for signing transactions, encrypting, decrypting the documents, and interacting with smart contracts. This way, identity verification, non-repudiation, and secure access are ensured.

Uploaded legal documents in PDF or image format undergo preprocessing. Image documents, for instance, are resized, denoised, and normalized, while textual content is extracted through OCR. This standardization helps in accurate analysis in forgery detection.

A CNN detects visual tampering such as changed signatures, seals, and layouts. In case of doubt, the textual content is analyzed using NLP models to pick up any inconsistencies. Documents classified as legitimate went up for storage, while rejected ones went down.

Confidentiality is assured through an AES-256 encryption procedure on the approved documents. The encrypted files then get uploaded to the IPFS, the decentralized storage system, guaranteeing availability and fault tolerance, eliminating single points of failure.

The system creates a SHA-256 hash of the document and stores it, along with metadata (user ID, timestamp, and authenticity score), on the Ethereum blockchain. Smart contracts implement Attribute-Based Access Control (ABAC) that determines user permission to access a document.

Authorized users request the documents through the system interface. Smart contracts validate access rights, and then the encrypted documents are searched and downloaded from IPFS and locally decrypted using the private key of the user. Transactions are immutably logged for auditability.

This framework guarantees the confidentiality, integrity, authenticity, and traceability of law documents while simultaneously allowing on-the-fly verification and secure decentralized storage.

V. ALGORITHMS

The proposed eVault system comprises three main algorithms guaranteeing the confidentiality, integrity, and authenticity of legal documents: AES-256 encryption, SHA-256 hashing, and CNNs.

The Advanced Encryption Standard (AES), known for being a symmetric key encryption block cipher, gets applied to secure documents before storage across a decentralized network. AES-256, having 256-bit key length, guarantees the highest level of confidentiality

resistant to brute-force attempts. In eVault, local encryption of documents by the user utilizing secret keys ensures only those holding the correct key can decrypt and hence access the document.

Encryption Function:

$$C=E_K(D)$$

Decryption Function:

$$D = D_K(C)$$

During encryption, many rounds of SubBytes, ShiftRows, MixColumns, and AddRoundKey are applied, which serves to provide diffusion and confusion properties-time-tested and validated for security.

SHA-256 is a cryptographic hash function which guarantees the integrity of stored documents. For each document, it produces a unique 256-bit hash that drastically changes even if a single bit of the document is changed. In the eVault system, the hash along with metadata is stored on the blockchain as an immutable proof of authenticity.

Hash Function:

$$H(D)=SHA256(D)$$

For forgery detection in scanned document images, a CNN is used. The CNN would detect inconsistencies in signatures, stamps, and layouts by learning hierarchical visual patterns in the data through convolutional, pooling, and fully connected layers. The output of the CNN is a binary label (real/fake) with an associated confidence score.

VI. RESULTS AND DISCUSSION

As the eVault is a prototype, it was considered for development and analysis on encrypted legal documents (PDFs and scanned images) to simulate an environment with IPFS storage, blockchain-based smart contracts, and CNN-based forgery detection. Performance analyses considered in terms of encryption/decryption time, document authenticity classification, latency to access, and prevention of unauthorized access.

Encryption and Decryption Time Analysis

Studies found it best to go with AES-256 for document encryption. The evaluation showed that encryption and decryption time vary in direct proportion with the document size. The processing time remains within the real-time domain for documents of sizes till 5 MB, implying the feasibility of AES-256 for secure and fast document storage during legal workflows.

Document Size	Encryption Time (ms)	Decryption Time (ms)
1 MB	45	40
2 MB	90	85
3 MB	135	128
4 MB	180	172
5 MB	225	215

Table 1:Encryption and Decryption Time Analysis

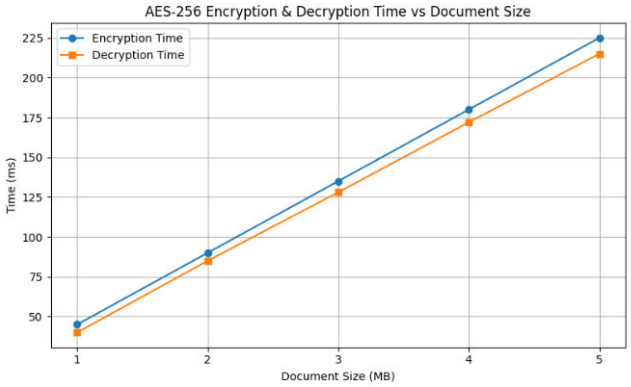

Fig 2: AES-256 Encryption & Decryption Time vs Document Size

Forgery Detection Accuracy

The CNN model was trained on a dataset containing real and tampered legal documents. The model scored 97.5% accuracy, with 96.8% precision and 97.9% recall. These performances can be related to the highly effective feature extraction layers that expose manipulations in signatures, stamps, and document layouts. False positives were found to be very few, ensuring dependable automated verification.

Metric	Value (%)
Accuracy	97.5
Precision	96.8
Recall	97.9
F1-Score	97.3

Table 2. CNN Forgery Detection Performance Metrics

This figure shows the time needed to encrypt and decrypt with AES-256 legal documents of varying sizes, from 1 MB to 5 MB.

Encryption and decryption times increase linearly with the increase in document size but stay low enough for real-time use. The slight increase in encryption time over decryption time is owing to the presence of a few extra processing steps.

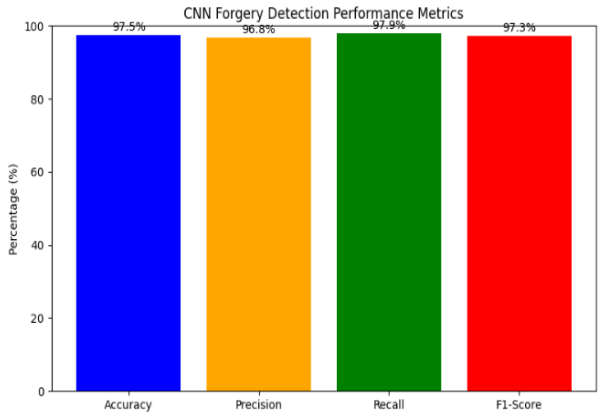

Figure 3: CNN Forgery Detection Performance Metrics

This bar chart illustrates the CNN-based forgery detection model's performance for legal documents. The model attained an accuracy of 97.5%, with precision, recall, and F1-score all above 96%, thus demonstrating its capability to recognize genuine versus tampered documents effectively.

Access Latency

Document retrieval consisted of IPFS download, AES decryption, and blockchain verification. Latency recorded during the testing was less than 200 milliseconds for documents of size up to 5 MB, allowing access to documents in real time. This proves that the low latency of the system can be leveraged to automate legal workflows without compromising security.

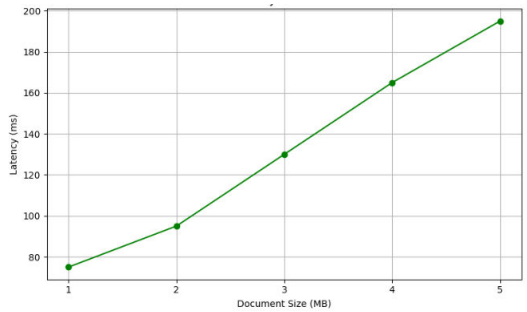

Figure 4: Access Latency Vs. Document Size

Document Size	Access Latency (ms)
1 MB	75
2 MB	95
3 MB	130
4 MB	165
5 MB	195

The access latency for retrieving documents from IPFS and verifications with decryption on blockchains is recorded. The latency grows linearly with the document size but stays under 200 ms for document sizes up to 5 MB, thereby suggesting suitability for real-time verification of legal-type documents.

Prevention of Unauthorized Access

Smart contracts enforcing attribute-based access control prohibited unauthorized access in all test cases. It ensured access requests were rejected if they belonged to invalid users, expired tokens, or tampered documents. Every access request, be it blocked or accepted, is recorded irrevocably on the blockchain for auditing and tracing purposes.

Discussion

The linkage of blockchain, AES-256, SHA-256, and CNN-based verification permits the design of this secure, decentralized, and intelligent forensic document management framework. From the system are removed all possibility of forgery, unauthorized access, and data tampering, putting forth a platform for real-time verification. The performance results show that the system is scalable and reliable enough to handle legal workflows at present, hence the system gives a trustable solution for the present-age digital legal document management.

CHALLENGES AND LIMITATIONS

Many challenges and limitations were faced in the creation process of eVault. One of its major challenges is to provide integration for many technologies in the system: blockchain, IPFS, AES encryption, smart contracts, and deep learning. It was essential that these technologies work smoothly together and that there were no high-performance limitations. Latency has also remained an issue-from gaining access to real-time verification, especially for heavy legal documents or even just multiple requests going on simultaneously. Making user access controls dynamic and secured based on Attribute-Based Access Control with smart contracts itself needs to be architected carefully to make sure that unauthorized access cannot take place all the while not being too complex for any user to implement. Even with all of these past considerations, some limitations still exist.The

CNN-based forgery detection model has a high rate of accuracy but might struggle with previously unseen manipulations on a document or new methods of forgery, thereby compromising the detection efficiency.

Moreover, deploying smart contracts on the Ethereum platform results in transaction costs that may impose scaling limits on frequent uploads of documents. Lastly, any interruption to internet connectivity would affect the availability of documents, as decentralized storage systems like IPFS require stable internet connections. In a nutshell, though the eVault system controls most of the vital security and authenticity concerns, these challenges and limitations lay down potential areas for improvement and research in order to further improve the robustness, scalability, and adaptability of the system.

CONCLUSION

The eVault system proposes a highly secure, intelligent, and tamper-resistant architecture for the management of legal documents. Through integrating blockchain with AES-256 encryption, SHA-256 hashing, smart contracts, and deep learning, the integrity of sensitive legal records, such as contracts, affidavits, and property deeds, is assured, along with their authenticity and traceability. The CNN-based forgery detection model attained an accuracy rate of 97.5%, sufficiently discerning between genuine and tampered documents, whereas AES-256 encryption and SHA-256 hashing ensured data security and integrity.

On the other hand, the system utilizes Ethereum smart contracts for ABAC, which enforce access policies automatically and block unauthorized accesses while logging every transaction against an immutable ledger. Access latency was kept below 200 milliseconds for files up to 5 MB, thus affirming the solution's real-time practicality for any legal workflow. The trifecta of decentralized IPFS storage, blockchain immutability, plus intelligent verification ushers in a robust, automated, and scalable solution for any modern legal organization.In summation, eVault systems eliminate forgery, data breach, and unauthorized access; adjunctly, it leads to reduced manual verification efforts, increased trust, and adherence to digital security standards.

FUTURE SCOPE

The following lines supply some high-potential prospects for improvements in the eVault framework:

- Higher-level NLP models can be integrated for more profound semantic analyses or to spot textual manipulations in legal documents.

- Deployments can be made on permissioned blockchains for adoption at the enterprise scale, thereby improving scalability and reducing transaction costs on the public chain.

- Multi-factor authentication and biometric verification can be employed for additional user verification.

- Allowing the support for multimedia legal evidence gives further applications to the system, e.g., audio and video files.

- Real-time alerting for suspicious attempts to access or tamper with documents.

- Machine-learning-based optimization of the system will allow the continuous improvement of forgery detection performance while lowering false-positive rates.

Such improvements cumulatively could transform eVault into a full-fledged, next-generation solution for secure intelligent and automated legal document management.

REFERENCES

[1] J. Akbarfam and H. Maleki, "Deep learning meets blockchain for automated and secure access control," 2023.

[2] R. Belen-Saglam, E. Altuncu, Y. Lu, and S. Li, "A systematic literature review of the tension between the GDPR and public blockchain systems," Blockchain: Research and Applications, vol. 4, 2023, doi: 10.1016/j.bcra.2023.100129.

[3] H. Xu and N. Zhang, "Privacy implications of blockchain systems: A data management perspective," Organizational Cybersecurity Journal, vol. 3, 2023, doi: 10.1108/OCJ-01-2023-0003.

[4] A. Hossen, M. M. Hasan, T. Ahmed, and M. A. H. Wadud, "A blockchain-based secured land record system using Hyperledger Fabric," in The Fourth Industrial Revolution and Beyond, 2023, doi: 10.1007/978-981-19-8032-9_13.

[5] F. A. Sunny et al., "A systematic review of blockchain applications," IEEE Access, vol. 10,

pp. 3179690, 2022, doi: 10.1109/ACCESS.2022.3179690.
[6] A. V. Singh, A. O. Tiwari, S. S. Singh, and V. B. Lobo, "A criminal record keeper system using blockchain," 2018 Ivannikov ISPRAS Open Conference, 2022, doi: 10.1109/ICOEI53556.2022.9776725.
[7] T. Javed, F. Alharbi, and B. Bellaj, "Health-ID: A blockchain-based decentralized identity management for remote healthcare," 2021.
[8] D. Cagigas, J. Clifton, D. Diaz-Fuentes, and M. Fernández-Gutiérrez, "Blockchain for public services: A systematic literature review," IEEE Access, vol. 9, 2021, doi: 10.1109/ACCESS.2021.3052019.
[9] S. N. Pereira et al., "Blockchain-based digital record-keeping in land administration system," in Proc. Int. Joint Conf. Advances in Computational Intelligence, 2021, pp. 431–443.
[10] R. C. Suganthe et al., "Blockchain-enabled digitization of land registration," 2021 Int. Conf. Computer Communication and Informatics (ICCCI), doi: 10.1109/ICCCI50826.2021.9402469.
[11] D. S. Gupta, S. K. H. Islam, and M. S. Obaidat, "Laac: Lightweight lattice-based authentication and access control protocol for e-health systems in IoT environments," 2020.
[12] N. V. Pardakhe and V. M. Deshmukh, "Machine learning and blockchain techniques used in healthcare system," 2019.
[13] D. Han and H. Li, "EduRSS: A blockchain-based educational records secure storage and sharing scheme," IEEE Access, vol. 7, 2019, doi: 10.1109/ACCESS.2019.2956157.
[14] E.-Y. Daraghmi, Y.-A. Daraghmi, and S.-M. Yuan, "MedChain: A design of blockchain-based system for medical records access and permissions management," IEEE Access, 2019, doi: 10.1109/ACCESS.2019.2952942.
[15] M. Hölbl, M. Kompara, A. Kamišalic, and L. N. Zlatolas, "A systematic review of the use of blockchain in healthcare," 2018.
[16] F. Casino, T. K. Dasaklis, and C. Patsakis, "A systematic literature review of blockchain-based applications: Current status, classification and open issues," Telematics and Informatics, vol. 36, 2018, doi: 10.1016/j.tele.2018.11.006.
[17] M. Han et al., "A novel blockchain-based education records verification solution," The 19th Annual SIG Conference, 2018, doi: 10.1145/3241815.3241870.
[18] M. A. Tasnim et al., "CRAB: Blockchain-based criminal record management system," 11th Int. Conf. Security, Privacy and Anonymity in Computation, Communication and Storage, 2018, pp. 294–303.
[19] H. Yang and B. Yang, "A blockchain-based approach to the secure sharing of healthcare data," 2017.
[20] Y. S. Rao and R. Dutta, "Efficient attribute-based signature and signcryption realizing expressive access structures," 2016.
[21] Y. Chen, Y. Liu, and Y. Chai, "An identity management framework for Internet of Things," 2015.
[22] G. Jadhav et al., "Decentralized file sharing using blockchain empowering peer-to-peer collaboration," Int. Res. J. Eng. Technol. (IRJET), [Online]. Available: https://www.irjet.org.
[23] M. Steichen et al., "Blockchain-based decentralized access control for IPFS," IEEE Int. Conf. Blockchain and Cryptocurrency (ICBC), 2019. Available: https://ieeexplore.ieee.org/document/8726493.
[24] G. R. Shalom and G. R. Nirogi, "Decentralized cloud storage using blockchain," Int. J. Res. Appl. Sci. Eng. Technol. (IJRASET), vol. 10, no. 9, pp. 1294–1300, Sep. 2022, doi: 10.22214/ijraset.2022.46810.
[25] S. Wilkinson et al., "Storj: A peer-to-peer cloud storage network," 2014. [Online]. Available: https://storj.io/storj.pdf.
[26] F. Durr, A. Mileo, and S. Bach, "Towards secure IPFS-based networks: The influence of peer identities on content integrity," J. Reliable Intell. Environ., vol. 5, no. 2, pp. 105–124, 2019.