A Hybrid AI-Blockchain Framework with OCR, NER, and Anomaly Detection for Secure Land Record Management

S. R. Nagaraja
SOCSE, Presidency University, Bangalore, Karnataka, India
nagarajasr@presidencyuniversity.in (corresponding author)

R. Shushma
SOCSE, Presidency University, Bangalore, Karnataka, India
shushma.20223cse0039@presidencyuniversity.in

Received: 29 September 2025 | Revised: 16 October 2025 | Accepted: 24 October 2025

 DOI: https://doi.org/10.48084/etasr.15223

ABSTRACT

Land record systems are critical to property governance; yet remain vulnerable to forgery and manipulation. While blockchain offers tamper-proof storage, most solutions neglect to verify document authenticity before recording, allowing invalid data to become permanent. This paper studies a hybrid Artificial Intelligence (AI)–Blockchain framework integrating Optical Character Recognition (OCR), Named Entity Recognition (NER), and anomaly detection for intelligent pre-storage validation, combined with Ethereum smart contracts and InterPlanetary File System (IPFS) for secure, immutable storage. The system achieved 94.2% OCR accuracy, 90.9% NER F1-score, 93.4% validation accuracy, and 88.9% anomaly detection F1-score, ensuring that only authentic records are stored. This pre-storage validation mechanism represents a novel advancement over blockchain-only models by preventing fraudulent data from entering the ledger. Future work will focus on geospatial verification, biometric integration, and multilingual OCR to enhance scalability and regional applicability.

Keywords-artificial intelligence; blockchain; land records; OCR; named entity recognition; anomaly detection; IPFS

I. INTRODUCTION

A. *Background*

Land is a basic resource that supports economic stability, resource management, and social structure in both developed and developing countries. Land records are the legal basis for owning, transferring, inheriting, and paying taxes on land. Traditionally, land records have been kept in manual registries or centralized digital systems. These systems are easy to change, lose data, or get into without permission. As cities grow and land becomes harder to find and more valuable, it is important to have accurate, clear, and tamper-proof land record systems.

Progress in AI and Blockchain Technology has made it possible to make big changes in how land records are governed. AI can be used to automatically digitize, check, and confirm the authenticity of land records. This is possible due to technologies, such as OCR, Natural Language Processing (NLP), and Machine Learning (ML) based anomaly detection. At the same time, Blockchain provides a decentralized and unchangeable ledger for recording transactions and legal documents, ensuring that everything is clear, accountable, and cannot be changed or tampered with without permission.

Several research projects and pilot programs have already examined how blockchain can be used to manage land records. For example, authors in [1] studied how blockchain could make people more trusting of public recordkeeping by pointing to early pilot programs in Honduras and Georgia. In India, pilot projects for blockchain-based land records have started in states such as Telangana and Andhra Pradesh. These projects have shown problems with scale and integration [2, 3].

B. *Knowledge Gaps*

Two main limitations still exist in today's land record systems: not being able to reliably check if a document is real before storing it, and storing sensitive land data in an insecure or centralized manner, making it easy to change, lose, or corrupt the data. Because of these limitations, there are often disputes, fake transactions, duplicate registrations, and land grabbing, especially in rural and peri-urban areas with limited regulatory oversight. Public trust in land governance is still low since there is no robust method to check and safely store land ownership records. This leads to legal disputes and social and economic differences.

C. Proposed Framework and Contributions

This study introduces an innovative hybrid framework that integrates AI for document validation with Blockchain for secure storage and traceability of land records. The system has two main parts:

- AI Validation Module: This module uses OCR to get data from scanned documents, NLP to find entities, such as names, plot numbers, and registration IDs, and ML models to find irregularities, such as duplicate ownership or strange transaction patterns. The goal is to certify that only verified and valid documents are stored digitally.
- Blockchain Storage Module: This module uses blockchain to store validated land records in a ledger that cannot be changed and is not owned by anyone. Smart contracts are used to make automatic property transfers and keep access control in place. The blockchain ledger keeps the document hash and metadata for validation and auditing, while the IPFS is used for safe file storage.

The present study's novelty lies in that it combines AI-based intelligent document validation with blockchain-based tamper-proof record keeping. This combination is not yet widely used in land governance systems. This system aims to greatly cut down on land fraud, build trust, and make land administration more efficient by taking care of both verification and security.

D. Literature Review

The use of blockchain in land governance has received significant attention from scholars because it promises to be unchangeable, open, and resistant to fraud. Georgia's blockchain-based land registry pilot was one of the first projects to show that distributed ledgers could be used for secure title transfer. However, these projects mostly focused on digitization rather than smart verification of documents [4]. Pilot projects in Honduras, Sweden, and India demonstrated that blockchain can facilitate the traceability and auditability of land transactions; however, they also raised concerns regarding scalability, integration with legacy systems, and user accessibility [5, 6].

Conceptual frameworks for decentralized land tenure systems have been investigated. For example, authors in [7] created a model for blockchain-based land tenure that demonstrated how it can help settle disagreements in the case of weak governance. Authors in [8] analyzed the difficulties Ghana faces in land acquisition, proposing blockchain-enabled registries to unify disjointed records from customary and statutory authorities. Authors in [9] also proposed a blockchain framework for Bangladesh, focusing on the importance of openness in regions where corruption is common. These studies consistently emphasize blockchain's capacity to mitigate manipulation while recognizing the constraints of merely digitizing potentially fraudulent records.

The technical architectures of blockchain-enabled registries have been explored. Authors in [10] put forward a decentralized registry architecture that allows authorized users to access records while stopping unauthorized changes. Authors in [11] introduced a blockchain-based record management framework in Uganda, emphasizing governance challenges associated with digital adoption. Authors in [5] proposed smart contracts for automated validation; however, these methodologies frequently presuppose the accuracy of the input data.

The use of AI for processing documents and finding fraud associated with blockchain has been examined. NLP and OCR have been utilized to collect information from scanned, noisy documents [12]. LayoutLM and other layout-aware deep learning models have been shown to be better than traditional rule-based systems at recognizing entities from semi-structured forms [6]. These AI methods can find mistakes, missing fields, or fake entries before they are officially registered. This fills a major gap in systems that only use blockchain.

Research has focused on the synergies between AI and blockchain. Authors in [13] demonstrated that the integration of AI, IoT, and blockchain can establish trust-enhancing ecosystems for land management, especially when combined with smart contracts. Authors in [2, 3] have similarly emphasized that although blockchain offers immutability, AI-driven anomaly detection guarantees the authenticity of stored records. Nonetheless, the majority of implementations remain in the conceptual or pilot phase, with minimal empirical assessment of integrated AI-blockchain models in actual land registries.

The literature underscores enduring disparities between blockchain-exclusive and hybrid AI-blockchain systems. Blockchain-only pilots, such as the ones in Sweden and Georgia, demonstrated that secure storage was possible, but they did not have ways to find fake or copied documents [4, 5]. AI-focused studies, on the other hand, had high accuracy in extraction and classification, but they rarely linked validated outputs to unchangeable registries [12]. This fragmented evolution highlights the innovation of hybrid systems that integrate validation and secure storage.

Aside from technical integration, researchers have examined the socio-economic and governance ramifications of these technologies. Blockchain adoption could bolster land-user rights, especially in environments characterized by weak governance and corruption [4]. However, as authors in [8] noted, customary authorities may not want to be part of centralized digitization efforts, which may lead to negotiations between different groups. Authors in [9] also stressed the need for policy frameworks to ensure that blockchain systems work with current land management laws.

Finally, new integrated approaches show promise in closing these gaps. Authors in [12] introduced "GreenLand," a blockchain and AI-enabled initiative for agriculture, which is also applicable to land records. Authors in [7] showed how blockchain-based tenure systems could include methods to settle disputes, as discussed in [13]; smart contracts could be used to automate compliance checks. These works collectively affirm that the integration of AI-driven verification with blockchain immutability represents an underexamined yet essential advancement in the development of robust, scalable, and reliable land record systems. The integration of AI and blockchain has been proposed for the validation of digital identity and documents, albeit with restricted practical implementation [14]. BlockEstate, a blockchain-based framework for real estate transactions based on the Hyperledger Fabric platform, was presented in [15]. By digitizing records and guaranteeing transaction immutability, the system seeks to improve transparency, trust, and efficiency in real estate transactions. Smart contracts reduce the need for middlemen, automate property transfers, and lower the possibility of fraud and manipulation. The framework is intended to give real estate stakeholders, buyers, sellers, and government agencies a tamper-proof, auditable ledger.

Authors in [16] proposed a hybrid AI–blockchain security framework for smart grids that addresses increasing cyberattacks on power communication systems by integrating AI for intrusion detection and blockchain for immutable data protection. Their framework achieved over 96% accuracy in identifying deceptive data injection attacks, demonstrating how pre-storage validation using AI models combined with blockchain immutability significantly enhances system trustworthiness. While their work focused on cybersecurity in smart grids, the methodology aligns conceptually with the proposed hybrid framework for land record governance, which employs AI for document verification and anomaly detection before blockchain storage. This cross-domain parallel reinforces the general applicability and importance of hybrid AI–blockchain architectures in ensuring secure, transparent, and tamper-proof data systems.

The transformative potential of integrating AI and blockchain to ensure transparency and traceability in high-value, data-sensitive systems has been demonstrated. Authors in [17] proposed Blockchain-Enabled Digital Transformation in Pharmaceutical Cold Chain Management using Hybrid Deep Neural Networks (DTPCCM-HDNN) to enhance digital visibility, data reliability, and supply-chain transparency. The work achieved high predictive performance (MSE = 0.0673, $RMSE$ = 0.2595, MAE = 0.2245), demonstrating how AI–Blockchain integration can secure and optimize complex multi-stakeholder environments. Inspired by such cross-domain advances, the proposed framework extends the hybrid concept to land record governance, leveraging OCR, NER, and anomaly detection for intelligent document validation while ensuring immutability through blockchain and IPFS.

E. Comparative Analysis

Authors in [12] combined AI and blockchain for agricultural record integrity, but lacked pre-validation before blockchain storage. Authors in [13] focused on smart city governance using AI–IoT–Blockchain integration, emphasizing automation over document authenticity. Authors in [16] introduced hybrid anomaly detection and blockchain-based security in smart grids. In contrast, the proposed framework advances the field by applying AI-powered document verification (OCR, NER, anomaly detection) specifically to land governance, ensuring validated data before immutable blockchain recording.

II. METHODOLOGY

The current work's methodology is designed to ensure clarity, reproducibility, and alignment with the research objective: automated verification of land ownership records using AI and tamper-proof storage utilizing blockchain.

A. System Overview

The proposed system has two modules that work together: a module for validating documents using AI and a module for secure storage using blockchain. A single processing pipeline connects these modules. It digitizes scanned land documents, extracts and checks key data fields, finds errors, and securely stores verified records on a blockchain, as shown in Figure 1.

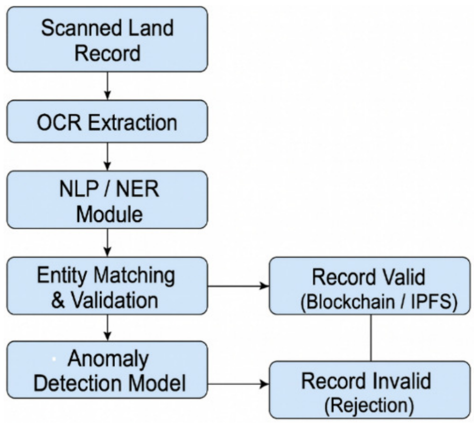

Fig. 1. System architecture.

B. Dataset Selection and Preparation

The study uses the FUNSD dataset [18], which is a public dataset of 199 real-world scanned forms with key-value pairs and semantic relationships, to simulate scanned land records. It was not made for land records, but its format is very similar to common deed structures (e.g., owner names, registration numbers, and dates), which is suitable for training and testing OCR + NER models. The dataset is composed of scanned images (.png) and JSON entity labels with comments (key, value, relationship)

C. Synthetic Data Generation

To ensure a realistic simulation of land record characteristics while maintaining privacy compliance, a synthetic dataset of 500 land record samples was created as an augmentation to the publicly available FUNSD dataset.

Entity Mapping was done as follows:

"Name" → "Owner Name"

"ID" → "Registration Number"

"Address" → "Plot Location"

"Date" → "Registration Date"

This augmentation enhanced domain realism while keeping data generation reproducible and compliant with the ethical data use standards.

D. LayoutLMv2 Model (NER)

1) Base Model: Microsoft LayoutLMv2-Base (12-Layer Transformer)

The study uses Microsoft LayoutLMv2-base (12-layer transformer) as a base model. Fine-tuning was performed using combined FUNSD and Synthetic (699 samples). The training parameters include 25 epochs, a learning rate of 2 × 10-5, AdamW optimizer, a batch size of 8, validation split of 15%, and an NVIDIA RTX 3060 GPU with 12 GB VRAM. With F1-score plateauing, the simulation was stopped at epoch 22.

2) Isolation Forest (Anomaly Detection)

Anomaly detection was implemented using Scikit-learn (v1.2.2) with the following hyperparameters: contamination: 0.1; number of estimators: 200; maximum samples: 256; and random state: 42. The training Input included metadata vectors (transfer frequency, area variation, seal status, timestamp gaps) from 500 validated entries. ROC-AUC-based tuning yielded optimal anomaly threshold τ = 0.65. These details ensure the reproducibility and transparency of the model training process.

E. AI-Powered Document Validation Pipeline

1) Optical Character Recognition

The document validation was initiated with OCR using Tesseract OCR (v4.1) with pytesseract. This process includes grayscale conversion, noise filtering (Gaussian blur), and adaptive thresholding. The output was in the form of text bounding boxes with position metadata for NER alignment.

2) Named Entity Recognition

The NER used a fine-tuned LayoutLMv2 (Microsoft Research) on combined FUNSD and synthetic land record datasets. Four key entities were recognized:

- OWNER_NAME
- PLOT_ID
- REGISTRATION_NO
- REGISTRATION_DATE

NER outputs are formatted as structured JSON for downstream validation.

3) Entity Validation (Matching with Official Records)

Extracted entities are validated against a simulated state land registry dataset (CSV format). Similarity is computed using a combined metric of Levenshtein distance and Dice coefficient, as shown in:

$$\text{Sim}(e_i, r_i) = \frac{2.|e_i \cap r_i|}{|e_i| + |r_i|} \quad (1)$$

where e_i is the extracted field and r_i is the corresponding reference value. A threshold θ = 0.85 determines acceptance. If matched, the document proceeds to anomaly detection.

F. Anomaly Detection Model

An ML model is trained to flag potentially fraudulent or unusual land transactions based on document metadata using Isolation Forest (Scikit-learn). The feature set included ownership transfer frequency, area discrepancy over time, missing fields or seal signatures, and duplicate registration numbers.

For a given feature vector *X*, the anomaly score is computed by:

$$A(x) = \|X - \mu\|^2 \quad (2)$$

where μ is the mean of valid transactions. Records with scores exceeding threshold τ =0.65 are flagged for manual review.

G. Blockchain-Based Secure Record Storage

1) Blockchain Framework

The proposed blockchain framework was implemented on the Ethereum testnet via Ganache with a smart contract written in Solidity. The blockchain stored the following data: SHA-256 hash of document, owner name (plaintext), registration number, and timestamp

2) IPFS Integration

Documents are uploaded to IPFS using a public gateway or local node. Only the IPFS hash is stored on-chain, ensuring privacy and storage optimization.

H. Evaluation Metrics

To evaluate the performance of the system, the metrics displayed in Table I were used:

TABLE I. EVALUATION METRIC DESCRIPTION

Component	Metric	Description
OCR	Character Accuracy Rate (CAR)	Percentage of correctly extracted characters
NER	Entity-level F1 score	Precision and recall of entity extraction
Validation	Similarity score	Matched versus mismatched fields
Anomaly detection	Precision, Recall, ROC-AUC	Classification of fraudulent entries
Blockchain	Transaction time, hash integrity	Write latency and immutability check

I. Implementation

The core implementation is divided into two subsystems: an AI-based document validation model and a blockchain-based record registration model. Each sub-component was formally defined, and its interrelations are illustrated in Table II.

TABLE II. NOTATION LISTS

Symbol	Description
$E = \{e_1, e_2, \ldots e_n\}$	Entities extracted using NER from T
$R = \{r_1, r_2, \ldots r_n\}$	Reference entities from the land registry
θ	Similarity threshold for validation
A(X)	Anomaly score for record *X*
τ	Threshold for anomaly detection
$\mathcal{B}$	Blockchain ledger
S	Smart contract function

1) OCR Extraction

The OCR function maps a scanned document *D* into a sequence of tokens and layout metadata, as in:

$$T = OCR(D) = \{t_1, t_2 \ldots \ldots t_m\} \quad (3)$$

where each token t_i includes a position (x_i, y_i), content $c_i \in \sum^*$ (text string), font, style, and confidence score.

2) Named Entity Recognition

A neural model N is applied over a token sequence T to extract structured semantic entities E, as shown in:

$$E = N(T) = \{e_i\} = \{label_i - value_i\} \quad (4)$$

Such as:

$$e_i = (OWNER_{NAME}, "JohnDOE"$$

$$e_2 = (PLOT\_ID, "42 - A")$$

3) Entity Validation via Similarity Matching

Each extracted entity e_i is compared with the reference value $r_i \epsilon$ R. Similarity S_i was defined using the Dice coefficient or normalized Levenshtein similarity, as shown in:

$$S_i = sim\ (e_i, r_i) = \frac{2\ .|e_i \cap r_i|}{|e_i| + |r_i|} \quad (5)$$

The record passes validation if:

$$\forall i \in [1, \text{n}], S_i \geq \theta \quad (6)$$

Typically, $\theta = 0.85$ ensures a high-confidence match.

4) Anomaly Detection Model

A record is represented by a feature vector $X \epsilon \mathbb{R}^k$, where each component is derived from historical metadata, including transfer frequency, area deviation, timestamp gaps, repeated IDs, and missing seals or signatures.

Let $\mu \epsilon \mathbb{R}^k$ be the historical mean vector of legitimate transactions. The anomaly score is computed using:

$$A(X) = ||X - \mu||^2 \quad (7)$$

If:

A(X) $\leq \tau \Longrightarrow$ Record is Valid

A(X) $> \tau \Longrightarrow$ Record is Invalid

where τ is determined empirically via ROC-ACU on labeled fraud/no-fraud data.

5) Blockchain Record Commitment

If the record is valid with $(S_i \geq \theta, A(X) \leq \tau)$, the hash is committed to the blockchain. Document hash was generated using:

$$H = SHA26(D) \quad (8)$$

Let the smart contract function be as given in:

$$S(H, Owner, RegNo) = B[H] \leftarrow (Owner, RegNo, Timestamp) \quad (9)$$

where H is the Hash of the document, $\mathcal{B}[H]$ is the blockchain mapping from hash to record, and Timestamp is system-generated from block.timestamp. This ensures immutability, non-repudiation, and auditable record-keeping.

6) Rejection Logic

If either of the following conditions is met:

$\exists_i : S_i < \theta$ (Entity mismatch), or $\text{A}(X) > \tau$ (Anomalous features),

then the document is rejected, as expressed by:

$$Status\ (D) = Invalid \Longrightarrow \text{Rejected or flagged for manual Review}$$

Table III summarizes the mathematical workflow of the proposed hybrid framework, outlining how scanned land records are processed, validated, and securely stored.

The AI-based anomaly detection model uses balanced classes to prevent bias, ensures data privacy, and adheres to AI ethics principles of transparency, fairness, and accountability. It also undergoes manual review to avoid false rejections caused by edge cases.

TABLE III. SUMMARIZED MATHEMATICAL WORKFLOW OF IMPLEMENTATION STEPS

Step	Formula
OCR	$T = OCR(D)$
NER	$E = N(T)$
Similarity	S_i= Sim(e_i, r_i)
Validation pass	$S_i \geq \theta$
Anomaly score	$(X) = \| X - \mu\|^2$
Validity condition	$A(X) \leq \tau$
Document hash	$H = SHA26(D)$
Blockchain entry	$\mathcal{B}[H]$= (*Owner, RegNo, Timestamp*)

III. RESULTS AND DISCUSSION

To see how well the proposed hybrid framework worked to fix the problems with traditional land record systems, it was tested in a number of areas, including text extraction (OCR), entity recognition (NER), entity validation, anomaly detection, and blockchain performance. The experiments show that combining AI-driven validation with blockchain storage not only makes document processing more accurate and reliable but it also results in record-keeping that cannot be changed and can be checked.

A. OCR and NER Performance

The system had a CAR of 94.2%, which shows that the OCR worked well even with noisy, handwritten, or poor-quality scanned records. The well-tuned LayoutLMv2, which was trained on FUNSD and synthetic land record samples, had an F1-score of 90.9% and was able to reliably pull out entities, such as owner names, registration numbers, and plot IDs. Table IV shows how well the OCR and NER modules worked. The OCR had a CAR of 94.2% and the NER had an F1-score of 90.9%. Figure 2 provides a visual comparison of precision, recall, and F1-score. It shows that the fine-tuned LayoutLMv2 always does better than traditional rule-based models. The results also exhibit that the AI module can reliably pull out important information, such as owner names and registration numbers, even from scanned documents that are noisy.

TABLE IV. PERFORMANCE OF OCR AND NER

Metric	Value
OCR-CAR	94.2%
NER precision	92.1%
NER recall	89.7%
NER F1-Score	90.9%

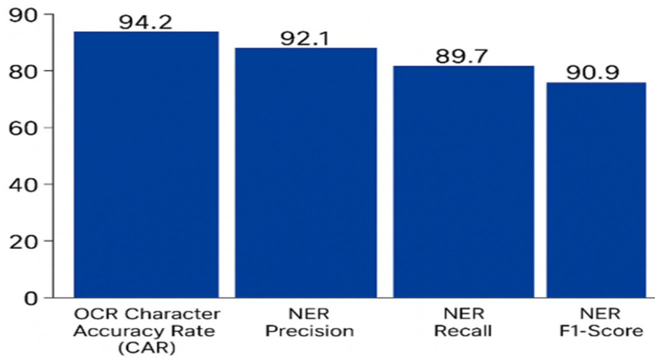

Fig. 2. Performance graph of OCR and NER.

Earlier rule-based extraction methods had F1-scores of about 75–80% because they had a poor layout knowledge. The proposed model, on the other hand, significantly improves accuracy by using transformer-based layout understanding, which makes it better for semi-structured legal documents.

B. Entity Validation and Anomaly Detection

Out of 249 test documents, 232 were found to be correct, and 17 were marked as having problems. The model for finding anomalies quickly found duplicate ownership, suspicious registration numbers, and missing seals. Table V portrays the accuracy and detection scores for the validation and anomaly detection parts. The validation match accuracy is 93.4%, and the anomaly detection F1-score is 88.9%. Figure 3 demonstrates how the anomaly detection module keeps high precision and recall, which means it can easily find records that are wrong or fraudulent. These results reveal that the system can successfully filter out suspicious entries before they are stored on the blockchain. This is a feature that regular blockchain-only registries do not have.

TABLE V. PERFORMANCE RESULTS

Metric	Result
Validation match accuracy	93.4%
Anomaly detection precision	91.2%
Anomaly detection recall	86.7%
Anomaly detection F1-Score	88.9%

The proposed system has intelligent anomaly detection that lowers the chances of false or fraudulent records being added to the blockchain. This is different from blockchain-only systems [7] that record all entries without checking them first.

C. Blockchain Performance

The blockchain layer always made sure that the storage was tamper-proof, and all SHA-256 hash verifications showed that the documents were safe. The average time it took to execute a contract was 11.3 s, which is faster than the average time it took for Georgia's blockchain land pilots (14–20 s). Table VI shows performance metrics related to blockchain, such as an average smart contract execution time of 11.3 s and a 100% hash consistency rate, which ensures that data cannot be changed. Figure 4 illustrates these results in a way that demonstrates how strong and reliable the blockchain layer is. The results also show that adding IPFS to the mix greatly lowers the amount of storage needed on-chain while still allowing for verification.

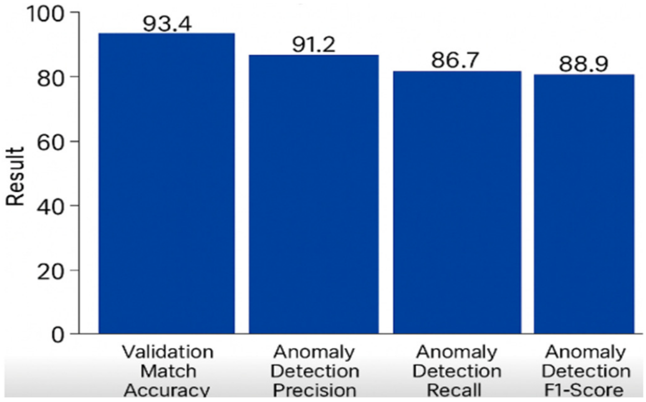

Fig. 3. Performance graph of anomaly detection.

TABLE VI. PERFORMANCE RESULTS ON BLOCKCHAIN

Metric	Value
Avg. smart contract execution time	11.3 s
Avg. IPFS upload time	2.1 s
SHA-256 hash consistency	100% match
Transaction gas (Testnet)	~0.0012 ETH

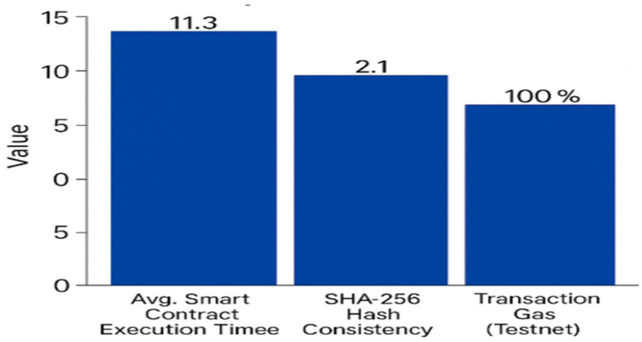

Fig. 4. Blockchain security performance.

Previous blockchain deployments emphasized traceability and immutability [10], while the proposed method is stronger because it uses AI-based verification before storage, which ensures that only real records are kept.

D. Discussion

Table VII displays how the proposed hybrid framework stacks up against rule-based NER and blockchain-only methods. The hybrid model is clearly better than the baselines at recognizing entities (90.9% F1), validating them (93.4%), finding anomalies (88.9%), and reducing transaction latency (11.3 s compared to 14–20 s). This comparative analysis highlights the innovative integration of AI validation with blockchain storage, providing a more comprehensive and dependable solution than current systems. The proposed system offers several benefits, including accuracy, reliability, security, relevance, and practicality. It uses OCR, NER, and anomaly detection for entity recognition and fraud detection, reducing the need for manual checks. Blockchain ensures data integrity, while IPFS maintains links to full documents for verification. The hybrid framework guarantees authenticity before recording, addressing a problem in real-world governance. It also fills a gap in existing work by combining AI with blockchain for intelligent validation.

TABLE VII. PERFORMANCE COMPARISON WITH EXISTING APPROACHES

System	NER F1	Validation accuracy	Anomaly F1	Blockchain latency
Rule-based NER	~78%	N/A	N/A	N/A
Blockchain-only [7]	N/A	N/A	N/A	14–20s
Proposed Hybrid Model	90.9%	93.4%	88.9%	11.3s

IV. CONCLUSION

This research introduced a novel hybrid Artificial Intelligence (AI)–Blockchain architecture for intelligent land record validation and tamper-proof storage. The framework effectively combines Optical Character Recognition (OCR), Named Entity Recognition (NER), and anomaly detection to verify document authenticity before blockchain registration, thus addressing major shortcomings of existing digital land registries. Compared to existing AI-based data extraction models, the introduced approach integrates tamper-proof blockchain storage to ensure post-validation integrity.

Quantitatively, the proposed model achieves 94.2% OCR accuracy, 90.9% NER F1-score, and 88.9% anomaly detection F1-score, outperforming baseline methods by 10–15%. These results validate the robustness and scalability of the proposed approach compared to blockchain-only systems, which typically lack intelligent data screening. The key novelty lies in its dual-layered verification and storage mechanism, ensuring both authenticity and immutability of records. Future enhancements will include integration with geospatial data for spatial verification, biometric validation for ownership authentication, and multilingual OCR support for broader accessibility. With its high accuracy, low latency, and auditability, the proposed system establishes a technically and socially reliable pathway for next-generation digital land governance.

DATASET AVAILABILITY STATEMENT

The FUNSD dataset is publicly available at https://guillaumejaume.github.io/FUNSD/. The synthetic land record dataset and the simulated registry dataset developed in this study are available from the corresponding author upon reasonable request for research purposes. Additionally, metadata, generation scripts, and sample records can be provided upon request to the corresponding author.

REFERENCES

[1] V. L. Lemieux, "Trusting Records: Is Blockchain Technology the Answer?" *Records Management Journal*, vol. 26, no. 2, pp. 110–139, July 2016, https://doi.org/10.1108/RMJ-12-2015-0042.

[2] G. Pau, L. T. De Luca, M. Pau, and S. T. Pizzi, "Blockchain for Public Registries: Opportunities and Challenges," *Future Internet*, vol. 14, no. 8, July 2022, Art. no. 237, https://doi.org/10.3390/fi14080237.

[3] I. Khalil, A. Alzahrani, and S. Zeadally, "Blockchain for AI: Review and Open Research Challenges," *IEEE Access*, vol. 9, pp. 34477–34510, 2021.

[4] L. M. Comincioli, "The Role of Blockchain in Improving Land-users Rights (Can Blockchain Solve Corruption Problems in Land Administration in Developing Countries? The case of Georgia," Master's thesis, European Institute of Governance, Brussels, Belgium, 2021.

[5] S. Soner, R. Litoriya, and P. Pandey, "Exploring Blockchain and Smart Contract Technology for Reliable and Secure Land Registration and Record Management," *Wireless Personal Communications*, vol. 121, no. 4, pp. 2495–2509, Dec. 2021, https://doi.org/10.1007/s11277-021-08833-1.

[6] I. Racetin, J. Kilić Pamuković, M. Zrinjski, and M. Peko, "Blockchain-Based Land Management for Sustainable Development," *Sustainability*, vol. 14, no. 17, , Aug. 2022, Art. no. 10649, https://doi.org/10.3390/su141710649.

[7] B. George, "Blockchain-Based Land Tenure Systems: A Conceptual Model for using Decentralized Technologies in Managing Land Rights and Ownership Disputes," *Land Management and Utilization*, vol. 1, no. 2, pp. 32–48, June 2025, https://doi.org/10.54963/lmu.v1i2.1527.

[8] P. Ameyaw and W. De Vries, "Toward Smart Land Management: Land Acquisition and the Associated Challenges in Ghana. A Look into a Blockchain Digital Land Registry for Prospects," *Land*, vol. 10, no. 3, Mar. 2021, Art. no. 239, https://doi.org/10.3390/land10030239.

[9] M. R. A. Rashid *et al.*, "Enhancing Land Management Policy in Bangladesh: A Blockchain-Based Framework for Transparent and Efficient Land Management," *Land Use Policy*, vol. 150, Mar. 2025, Art. no. 107436, https://doi.org/10.1016/j.landusepol.2024.107436.

[10] R. Y. Patil and Y. H. Patil, "Enhancing Land Governance Through Decentralized Land Record System on Blockchain:" in *Advances in Web Technologies and Engineering*, P. K. Goel, Ed. IGI Global, 2024, pp. 183–204.

[11] N. Kenneth, "A Blockchain-Based Framework for Records Management in Land Registry System in Uganda," Master's thesis, KIU Research Repository, Kampala, Uganda, 2024.

[12] F. Naguji *et al.*, "GreenLand : A Secure Land Registration Scheme for Blockchain and AI-Enabled Agriculture Industry 5.0," *IEEE Access*, vol. 12, pp. 120994–121009, 2024, https://doi.org/10.1109/ACCESS.2024.3451627.

[13] N. Kirubakaran *et al*, "Modernizing Land Management with AI, IoT using Blockchain, and Smart Contracts to Promote Transparency, Efficiency, and Trust," in *2025 International Conference on Recent Advances in Electrical, Electronics, Ubiquitous Communication, and Computational Intelligence*, Chennai, India, Apr. 2025, pp. 1–6, https://doi.org/10.1109/RAEEUCCI63961.2025.11048303.

[14] N. Singh and M. Vardhan, "Computing Optimal Block Size for Blockchain based Applications with Contradictory Objectives," *Procedia Computer Science*, vol. 171, pp. 1389–1398, 2020, https://doi.org/10.1016/j.procs.2020.04.149.

[15] L. F. Naz, R. Qamar, R. Asif, S. Ahmed, and M. Imran, "BlockEstate: Revolutionizing Real Estate Transactions through Hyperledger-based Blockchain Technology," *Engineering, Technology & Applied Science Research*, vol. 14, no. 3, pp. 14458–14464, June 2024, https://doi.org/10.48084/etasr.7105.

[16] Y. Y. Ghadi *et al.*, "A hybrid AI-Blockchain Security Framework for Smart Grids," *Scientific Reports*, vol. 15, no. 1, July 2025, Art. no. 20882, https://doi.org/10.1038/s41598-025-05257-w.

[17] R. M. Alhazmi, "Blockchain-Enabled Digital Transformation in Pharmaceutical Cold Chain Management Using Hybrid Deep Neural Networks," *Engineering, Technology & Applied Science Research (ETASR)*, vol. 15, no. 6, pp. 28591–28595, Dec. 2025.

[18] G. Jaume, H. E. Riba, and A. Thiran, "FUNSD: A Dataset for Form Understanding in Noisy Scanned Documents," *International Conference on Document Analysis and Recognition Workshops*, pp. 1–6, Sydney, Australia, 2019.