This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3451627

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

GreenLand: A Secure Land Registration Scheme for Blockchain and AI-enabled Agriculture Industry 5.0

FESHALBHAI NAGUJI[1], NILESH KUMAR JADAV[1], SUDEEP TANWAR[1] *SENIOR MEMBER, IEEE*, **GIOVANNI PAU[2], GULSHAN SHARMA[3], FAYEZ ALQAHTANI[4], AMR TOLBA[5]** *SENIOR MEMBER, IEEE*

[1]Department of Computer Science and Engineering, Institute of Technology, Nirma University, Ahmedabad, Gujarat, India- 382481 (e-mails: 21mcec05@nirmauni.ac.in, 21ftphde53@nirmauni.ac.in, sudeep.tanwar@nirmauni.ac.in)
[2]Faculty of Engineering and Architecture, Kore University of Enna, 94100, Enna, Italy(e-mail: giovanni.pau@unikore.it)
[3]Department of Electrical Engineering Technology, University of Johannesburg, Johannesburg 2006, South Africa (e-mail: gulshans@uj.ac.za)
[4]Software Engineering Department, College of Computer and Information Sciences, King Saud University, Riyadh 12372, Saudi Arabia (e-mail: fhalqahtani@ksu.edu.sa)
[5]Computer Science Department, Community College, King Saud University, Riyadh 11437, Saudi Arabia (e-mail: atolba@ksu.edu.sa)

Corresponding authors: Giovanni Pau (e-mail: giovanni.pau@unikore.it) and Sudeep Tanwar (e-mail: sudeep.tanwar@nirmauni.ac.in)

This work was funded by the Researchers Supporting Project No. (RSP2024R509) King Saud University, Riyadh, Saudi Arabia and also by Kore University of Enna, 94100, Enna, Italy.

ABSTRACT The main aim of the proposed system is to facilitate secure and protected land registry in the domain of agriculture Industry 5.0. Considering the outlook of issues associated with it, we considered the blockchain and AI-based technology to fulfill the purpose of secure land registry.

Purpose: Establishing and confirming land ownership is essential for the land registry system in ensuring the protection of ownership rights, particularly crucial in the contexts of agriculture and Industry 5.0. In these sectors, land serves as a crucial resource for sustainable development and industrial innovation. Most of the existing works rely on legacy and centralized system to store land records; which result in high incidences of forgery and fraud. Therefore, maintaining a robust land registry system is essential to fostering economic investments, promoting green practices, and facilitating equitable access to land resources in agriculture and Industry 5.0 ecosystem.

Methods: We proposed an AI and blockchain-enabled land registry system for agriculture and industry 5.0 that offers a more reliable, transparent, and efficient solution to the challenges of lack of transparency, data tampering, and inefficiency, which can result in disputes, fraudulent claims, and a lack of trust during the land registry. AI models, such as logistic regression (LR), support vector machine (SVM), random forest (RF), extreme gradient boosting (XGB), and light gradient boosting machine (LGBM), are employed to classify the fraud and non-fraud land data. Only the non-fraud land data is forwarded into the blockchain network, thereby reducing the computational overhead of the proposed land registry system. In the blockchain network, we designed various smart contracts that validate the land data with unparalleled efficiency and security. Further, the slither solidity source analyzer tool is used for smart contract vulnerability assessment. After the assessment, the smart contract is deployed using the Sepolia test network. The non-fraudulent land data is redirected to the interplanetary file system (IPFS) that stores the original data and forwards the associated hash into the blockchain's immutable ledger.

Results: The entire proposed system is evaluated with different performance parameters, such as AI statistical measures including accuracy, ROC, log-loss score, blockchain scalability comparison, gas cost utilization, and bandwidth utilization. Furthermore, the vulnerability assessment of the smart contract is analyzed using Slither to highlight the working of proposed system without any vulnerabilities.

Conclusion: The proposed blockchain and AI-based land registry system ensure a secure and intelligent pipeline to combat against land forgery activities.

INDEX TERMS Artificial intelligence, machine learning, land registry system, blockchain, Ethereum, Bitcoin, and smart contracts.

I. INTRODUCTION

The modern era of agriculture and industry 5.0 has revolutionized land use through cognizant technologies, such as precision farming, crop monitoring, and smart manufacturing [1]. These innovations optimize resource and operational efficiency, increase profits, and lessen environmental impact, reshaping the land for a more sustainable future. Countries worldwide have recognized the critical importance of implementing robust land registry systems to ensure the legitimate ownership and legal transfer of land rights for effective agriculture and industry 5.0 applications. These systems are designed not only to prevent disputes and fraudulent claims but also to establish a foundation of trust and reliability in property transactions. Nevertheless, contemporary land registry systems grapple with several pressing challenges, each of which threatens their efficacy and public confidence. The proposed study presents a robust strategy to build a more reliable and effective land registry system by fusing blockchain and artificial intelligence (AI) technology. AI plays a crucial role in the modern agriculture section, where it can offer precision farming by analyzing soil data in real-time, which further allows the performance of precise monitoring and field management. It is also used for prediction, such as yield and pest prediction, where AI models are used to train on historical and real-time datasets to give informed decisions to farmers for selling crops, harvesting, and efficient plantation. Further, AI has a big role in supply chain optimization, sustainable agriculture, and drone-based farming [2].

A contemporary community must have land registry systems because they are necessary for ensuring that land is legitimately owned and transferred, securing property rights, and preventing disputes and fraudulent claims. In the past, government-maintained centralized databases that have been used to construct land registry systems. The government body is in charge of confirming and documenting land transfers and sales as well as issuing land titles. Additionally, many land management systems were proposed in the past; for example, to solve an unknown ownership issue in Portugal, the Cadastral Information and Property Management System (SICAP) platform was proposed [3]. It aims to support land registry management and provide useful functionalities to perform analysis operations and visualize metrics and statistics of the properties data. However, this approach uses a centralized database to store all the data, which may lead to single node failure and vulnerable to many security attacks.

Further, to address the issues of the paper-based land registry system in Bangladesh, a multi-channel model was proposed that leverages digital land management technology to improve the existing systems [4]. This approach aims to improve efficiency, transparency, and accessibility and address the limitations of paper-based systems. However, this approach uses a centralized MySQL database, which may become vulnerable to security attacks, such as data manipulation and SQL injection, that can degrade the performance of their proposed work. Next, to address issues of corruption and inefficiency in Bangladesh, a web-based land management system was proposed [5]. Their approach aimed to improve transparency, efficiency, and accessibility, along with reducing the cost. However, this approach also relies on a centralized database, which may become vulnerable to highly confidential data like land or property records.

In order to solve the issues of a centralized database, blockchain was introduced for the land registry system. It facilitates transparency, immutability, decentralization, traceability, and elimination of third-party users, which offers security against various confidentiality and integrity attacks. In the blockchain, every land ownership transaction is recorded on a safe and tamper-proof ledger that is accessible to all stakeholders by blockchain, which acts as a decentralized ledger to ensure a high level of transparency and immutability. Blockchain removes the need for intermediaries like governmental organizations by allowing peer-to-peer transactions, which can lead to a more effective and less dishonest land registry system. Authors of [6] proposed a blockchain-based secure land registry system, wherein they developed a smart contract for land registration to record the land owner, land details, and land ownership details. Further, the smart contract was deployed on the Ethereum blockchain, and a decentralized application was built to interact with different landowners. They achieved around 3 ms computation time for processing the user's request and a number of blocks. Their scheme did not perform a cost analysis of different smart contract functions to validate the performance.

Later, [7] addressed the above issue and performed a cost analysis of their smart contract. They implemented a decentralized application using the Ethereum blockchain and Web 3.0 with the Ganache platform. They provided the graphical user interfaces of each webpage associated with different users. Moreover, they measured the gas cost and Ether cost consumption for the number of documents. They did not calculate the network performance of their proposed scheme. To this concern, authors of [8] implemented a secure land registration using the Ethereum blockchain. They designed a smart contract, which includes newRegistration(), addProperty(), propertyVerification(), searchProperty(), buyProperty(), and removeOwnership() functions to record the land details in the blockchain. Moreover, they measured the transaction throughput, latency, response time, and computation time to check the ability of their proposed scheme. They achieved a 9-second average latency for 500 transactions, around 450 ms response time, and around 530 ms computation time for 500 users. However, they were not concerned about the scalability of their proposed scheme.

With this aim, [9] proposed a land registration using the Ethereum blockchain and IPFS. They implemented a pinata API and recorded the land-related data to an IPFS-based decentralized file system, and only the hash of that data is stored in the blockchain. Using IPFS-based data storage takes very minimal time in transaction verification, commitment, and consensus achievement. This minimum time processed more transactions and improved the scalability of their proposed scheme. However, they did not check the vulnerability of their smart contract. The existing work uses blockchain for large amounts of data storage without validating the intent of the data, whether the data is fraudulent or non-fraudulent data. It makes blockchain computationally heavy and does not provide a scalable solution.

To address the aforementioned issues, we integrated blockchain and Artificial intelligence (AI)-based land registration, where agricultural land and other land data and ownership data are validated using AI-based classification algorithms. Moreover, the verification and processing of land ownership transactions can be automated by AI, eliminating the need for human processing by government employees. This can speed up the process and lower the cost of land ownership during transfer ownership. Further, only valid/non-fraudulent data is stored in the IPFS, and only hash data is stored in the blockchain, which improves the scalability and network performance of the land registration system. Moreover, the blockchain provides security and privacy to land registration transactions that help to keep the data secure from vulnerabilities. Additionally, the proposed scheme uses a Slither smart contract vulnerability assessment tool to check the validity of the smart contract code before deploying it to the Ethereum blockchain. This integration has two benefits: the AI-based classification bifurcates valid and invalid data for a blockchain. Further, blockchain-based data storage improves the security and privacy of land registry systems. On the other hand, AI also reduces the computational overhead of the blockchain network since wrong data is discarded from the land registry systems. This dual approach offers significant improvement in terms of efficiency, reliability, and computation to the land registry deployed in agriculture and industry 5.0.

A. MOTIVATION

Systems for keeping track of land ownership, which is a core component of property rights, are essential. However, there are a number of issues with the current land registry systems, including a lack of transparency, fraud, and slowness, which reduces the operational efficiency of the applications maintaining land records in agriculture and industry 5.0. These issues resulted in conflicts, phoney claims, and a lack of confidence in the land registry system, which has negative economic and social consequences. To address these issues, the authors of [6] [7] [8] [9] used a blockchain-based land registry system. This system offers transparency, end-to-end traceability of land registration details, security and privacy. These scheme uses different test networks of the Ethereum blockchain to create and deploy smart contracts. Their schemes did not verify the smart contract before deploying it in the blockchain. The existing scheme did not offer a scalable solution and lower network performance. Moreover, the state-of-the-art work did not verify the intent of the data prior to adding it to the blockchain. These challenges motivate us to design a reliable, transparent, and effective land registry system. The integration of AI and blockchain increases the overall effectiveness and integrity of the land registration system, AI can automate the verification and processing of land ownership transactions as well as provide predictive analytics to detect and avoid false claims. The goal of the proposed work is to fully understand and implement how this solution can address the problems with current land registry systems while additionally examining the possibility of a blockchain-based AI-enabled land registry system. With the proposed system, citizens and the economy as a whole might get access to a more reliable and accessible land registry system for agriculture and industry 5.0.

B. RESEARCH CONTRIBUTIONS

The study enhances the subject of land registry systems in the following ways.

- We studied and investigated the problems associated with the current land registry systems and proposed a secure land registry system using AI and blockchain technology for agriculture and industry 5.0.
- To apply classification algorithms on the standard land registry dataset that classify falsified and non-falsified land records. For that, we applied logistic regression (LR), support vector machine (SVM), random forest (RF), extreme gradient boosting (XGB), and light gradient boosting machine (LGBM).
- Further, we proposed a blockchain-based land registry system that allows only legitimate land records to get stored in the blockchain's immutable ledger. In that view, we designed smart contracts that verify and validate the land record data and simultaneously forward them to the interplanetary file system (IPFS) for an efficient data storage and retrieval process.
- The proposed system is evaluated by considering different performance metrics, such as statistical measurements - accuracy, precision, recall, F1-score, receiver operating characteristic (ROC) curve, IPFS bandwidth utilization, and blockchain gas consumption.

C. ORGANIZATION

The remaining paper is organized as follows. Section II introduces the related works. Section III presents the system model and problem formulation. Section IV presents the proposed AI and blockchain-enabled secure land registration system. Section V discusses the results analysis of the proposed system. Finally, Section VI gives the concluding remarks.

II. RELATED WORKS

This section discusses various state-of-the-art approaches proposed by various authors for the land registry system. Table 1 shows the relative comparison table of various

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3451627

TABLE 1: A relative comparison of the proposed scheme with the state-of-the-art land registry systems.

Author	Year	Objective	(a)	(b)	(c)	(d)	Pros	Cons
Kusuma *et al.* [10]	2023	To make the land registration process more transparent, efficient, and less prone to fraud	✓	-	✓	-	In detail implementation is provided	Slow storage used without IPFS, no vulnerability assessment proof is provided, No use of AI
Shrivastava *et al.* [11]	2023	Use blockchain technology to establish a system for land registration.	✓	-	✓	-	Implementation based, Secure, fast	No use of fast IPFS storage and AI, Provided very fewer implementation results
Ncube *et al.* [12]	2022	Using a Distributed Ledger for the Land Registry System	✓	-	✓	-	Implementation based, high reliability, high security	Works in permissioned environment only, slow storage used, No AI-based approach
Khalid *et al.* [13]	2022	Increase the reliability of land registration system.	-	-	✓	-	Decentralized, reliable, proof-of-concept system, secure	Conceptual framework without any kind of implementation
Suganthe *et al.* [14]	2021	Land Registration Digitization Using Blockchain	✓	-	-	-	Implementation based, transparent, reliable	Involvements of a third party, slow storage used, no AI used
Mishra *et al.* [15]	2021	Digitalization of Land Records using Blockchain Technology	-	-	✓	-	Reduced fraud cases, transparent, fully decentralized	Any kind of implementation is not provided
Nandi *et al.* [16]	2020	Secure method for resolving land registration problems.	✓	-	✓	-	Decentralized system, Implementation based, Secure.	Slow and high-cost data storage, No use of AI
Khan *et al.* [17]	2020	To quicken India's land registration procedure.	✓	-	-	-	Implementation based, Increased security, decreased frauds.	Involves third party, No use of AI
Gollapalli *et al.* [18]	2020	Peer to peer system that doesn't require any middleman	✓	-	✓	-	Security, different functionalities, transparent, scalable	Works in permissioned environment only, slow storage used, No AI is used
Shinde *et al.* [19]	2020	Secure storage for property papers.	✓	✓	-	-	Tamper-proof, fast IPFS storage.	Process of land registration done in a person
Singh *et al.* [20]	2020	Digitization of land record of the Indian scenario	-	-	✓	-	Peer to peer, fast, decentralized, trustworthy, transparent	No SC implementation provided
The proposed system	2023	AI-Enabled Secure land registry system using blockchain	✓	✓	✓	✓	Tamper-proof, fast IPFS storage., fully decentralized	-

(a)Implementation, (b) IPFS, (c) Full decentralized, (d) AI incorporation

existing state-of-the-art approaches for land registry systems. Kusuma *et al.* [10] proposed a blockchain-based approach to strengthen the existing land registration process in terms of transparency, money, and time and reduce instances of fraud. Their proposed system is fully decentralized without any third-party involvement, and their approach features smart contract-based implementation. However, their approach uses Ethereum nodes to store all the data, which makes the process of data storing/fetching very slow compared to IPFS storage, and the cost of storing all the data in Ethereum nodes is very high compared to IPFS storage. Their approach also lacks in the usage of Artificial Intelligence, which can help to classify fraud vs. non-fraud transactions and save the cost of storing fraud data into blockchain nodes. Shrivastava *et al.* [11] utilized blockchain technology to create a land registry system. Their proposed approach is fully secure, fast, and implementation-based. However, their proposed approach lacks the usage of fast IPFS storage, and no AI is used in their proposed approach, which may allow fraud transactions to be stored in blockchain.

Furthermore, with the advancement and innovation in the modernization of agriculture and industry 4.0 through the usage of cognizant and advanced technologies such as precision farming, crop monitoring, and smart manufacturing for which the authors in [1] proposed a survey on various image processing techniques for detection and classification of citrus plant diseases. Their main focus is to ensure the full automation of detection and classification procedure. Later, Ncube *et al.* [12] proposed a land registry system using distributed ledger. Their proposed approach features smart contract-based implementation. Their proposed approach is highly secure and reliable. However, their proposed approach is implemented on a permissioned blockchain environment, which is only limited to certain users. Also, their proposed approach lacks in the usage of fast IPFS storage and AI technologies, which can largely enhance the land registry system. Khalid *et al.* [13] proposed an approach to increase the reliability of the land registration system. Their proposed approach is secure, decentralized, reliable, and proof-of-concept based. However, their proposed approach only features a conceptual framework without any implementation.

Suganthe *et al.* [14] proposed blockchain-enabled digitization of land records to improve the land registration process in India. Their proposed work is based on high transparency and reliability, which makes the land registration process highly secure and fast in India. However, their proposed approach involves a third party that opens the system to several vulnerabilities, and also their proposed approach does not feature any AI integration with blockchain. Mishra *et al.* [15] digitized the land Records using blockchain technology to reduce human errors, corruption, and human intervention. Their proposed approach helps to reduce digital fraud by using features of blockchain, such as transparency. However, it is observed that their proposed work lacks in providing implementation details of their work and also lacks in the usage of AI technologies.

Nandi *et al.* [16] address the land registration issues in the traditional land registry system and provide an implementation-based system for a secure land registry. Their proposed system is fully decentralized and secure. They also provided smart contract implementation for their proposed approach. However, they are using Ethereum blockchain nodes to store all their user's data and transactions, which makes data processing very slow and costly as Ethereum blockchain nodes are very slow and hard to maintain for large amounts of data. Also, their proposed approach does not use AI, which can classify fraud transactions and save the cost of storing fraud transactions in the blockchain. Further, Khan *et al.* [17] analyzed the increasing fraud in India's traditional land registry systems. They proposed a secure land registry system to accelerate the land registration process in India. Their proposed approach is highly secure, fast, and secure against digital fraud. However, their approach involves a third party, making it vulnerable to various security attacks (e.g., data manipulation attacks) and not a fully decentralized system. Gollapalli *et al.* [18] proposed a peer-to-peer system that doesn't require third-party involvement. Their proposed approach offers high security, scalability, transparency, and sustainability. However, their proposed approach only works in permissioned environments as it is developed in a hyperledger fabric blockchain network and also lacks the usage of AI technologies. Shinde *et al.* [19] proposed secure storage for property papers using blockchain technology. Their proposed approach is tamper-proof and provides IPFS storage, making the entire process fast. However, their proposed approach to the land registration process must be made in person as it only works to store the property papers. Singh *et al.* [20] proposed an approach to digitizing land records for the Indian scenario. They proposed a peer-to-peer, fast, decentralized, trustworthy, and transparent system. However, their proposed approach doesn't feature any implementation. Therefore, there is a requirement for an amalgamation of AI and blockchain-enabled, tampered-proof, reliable, and scalable system that tackles the aforementioned issues in traditional land registry systems.

III. SYSTEM MODEL AND PROBLEM FORMULATION

In this section, we provide a novel strategy for land registry systems that combines blockchain and AI to ensure improved data security and integrity. Traditional land registry systems often rely on centralized databases, denoted as $\mathcal{C}$, managed by a single authority (a) or a consortium of authorities $\{a_1, a_2, \ldots, a_k\} \in A$ for storing land ownership data.

$$\mathcal{C} \subseteq \{d_1, d_2, \ldots, d_6, \ldots, d_N\} \in D \tag{1}$$

$$A \xrightarrow{\text{manages}} \mathcal{C} \tag{2}$$

The dataset D in this instance includes land records (L), whereas (A) can be either a single governing authority or a group of authorities. We have (N) users in our architecture, each of whom is linked to a certain set of land record data, as shown in Eq. 3.

$$\{u_1, ..., u_i, ..., u_m\} \rightarrow \{d_1, ..., d_i, ..., d_n\} \tag{3}$$

$$\forall u_i \xrightarrow{\text{has}} d_i, 1 \leq i \leq n \tag{4}$$

$$u_i \in U, d_i \in D \tag{5}$$

The terms $u_i \in U$ and $d_i \in D$ highlight the crucial functions of users and land records in conventional land registry systems. However, because these systems are centralized, they are susceptible to security breaches. Malicious actors (α) can use these weaknesses to change land records.

$$\alpha \in \{A \text{ or } a_h\} \tag{6}$$

Here, a_h represents a potential attacker within the authority set $\{a_1, a_2, \ldots, a_k\} \in A$. These attackers can compromise data integrity, resulting in unauthorized data tampering:

$$\alpha \xrightarrow{\text{manipulates}} \mathcal{C} \tag{7}$$

$$\alpha \xrightarrow{\text{tampered}} \{d_1^\alpha, d_4^\alpha, \ldots, d_8^\alpha, \ldots, d_f^\alpha\} \in D^\alpha \tag{8}$$

Eq. 7 illustrates how α first manipulates $\mathcal{C}$ through privilege escalation attacks, followed by tampering with specific land records, ranging from d_f to d_f^α. To address these data integrity challenges and enhance security, we propose the integration of AI and blockchain technologies. In the proposed system, we envision before user transactions are stored on the blockchain, an AI categorization module examines them. Transactions are categorized as fraudulent or not fraudulent using supervised learning algorithms. This classification algorithm makes use of crucial transaction characteristics such as receiver account addresses, flags, and transaction amounts. Based on the aforementioned facts related to the land registry system, an objective function (O) has been formulated to take into account how well the AI classification module detects fraudulent transactions,

$$O = \max_{1 \leq i \leq N} S_i, n \times A \tag{9}$$

Here, S_i denotes the system's security, n is the number of transactions, and A represents the accuracy of the AI classification module. It is crucial to use a trustworthy dataset when developing the algorithm in order to guarantee the efficacy of the AI classification module. The chosen attributes must be pertinent to the land registry setting, and the dataset should include both fraudulent and non-fraudulent transactions. The dataset structure is represented as.

$$D = (x_i, y_i)_{i=1}^{n} \tag{10}$$

In this equation, D constitutes the dataset, x_i represents the pre-processed data for transaction i, and $y_i \in 0, 1$ indicates the label for transaction i, where 0 signifies a legitimate transaction, and 1 represents a potentially fraudulent transaction. To prepare the transaction data for the AI classification module's training, pre-processing is required. This entails feature extraction and encoding, which converts transaction data into a numerical format and includes features such as transaction type, receiver account, and amount.

$$t_1, ..., t_i, ..., t_n \rightarrow x_1, ..., x_i, ..., x_n \tag{11}$$

$$\forall_i t_i \rightarrow x_i, 1 \leq i \leq n \tag{12}$$

$$t_i \in T, x_i \in X \tag{13}$$

Here, $t_i \in T$ represents transaction data, while $x_i \in X$ signifies pre-processed data. For the purpose of creating and analyzing algorithms, the dataset is split into training and validation sets. The training set is used to optimize the model's parameters, while the validation set is used to evaluate the model's effectiveness and prevent overfitting. The goal of the optimization procedure is to reduce the difference between the anticipated and actual transaction labels or the loss function. The loss function is written as follows.

$$L(\hat{y}, y) = -\frac{1}{n}\sum_{i=1}^{n} y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i) \tag{14}$$

Here, $\hat{y}_i$ denotes the predicted label for transaction i, and y_i represents the true label for transaction i. Decision trees and other supervised learning algorithms are used to categorize transactions as valid or potentially fraudulent while minimizing the loss function to optimize the parameters.

$$\hat{y} = f(x) \tag{15}$$

In Eq. 15, $\hat{y}$ signifies the predicted class of a transaction, while $f(x)$ represents the trained algorithm. By contrasting the projected class with the actual class of the transaction, which can be determined either through expert judgment or past transaction classifications, the model's accuracy is evaluated. Ultimately, the proposed approach incorporates the AI classification module to analyze transactions in real-time, effectively preventing potentially fraudulent transactions from being recorded on the blockchain. This integration not only enhances security but also reduces the costs associated with storing fraud.

IV. THE PROPOSED SYSTEM

FIGURE 1 shows the sequential workflow of the proposed system, i.e., AI-enabled secure land registry system using blockchain for agriculture and industry 5.0. The proposed system includes an AI classification that analyzes transactions for fraudulent activity before deciding whether to store them in the blockchain. This is designed to prevent fraudulent transactions from being stored in the blockchain, thus saving the cost of storage and making the approach much more secure. The proposed system is divided into 3 layers, namely (i) the data layer, (ii) the AI layer, and (iii) the Blockchain layer. The description of each layer is as follows.

A. DATA LAYER

The data layer in the proposed system consists of unprocessed transaction data about buyers, sellers, and lands up for sale. Eq. (16) is used to represent the set of all buyers in the transaction data. The set of all sellers involved in the transactions is represented in Eq. (17) in a similar manner. Similarly, Eq. (18) is used to represent the set of all lands that are for sale. Eq.(19) shows the buyer-seller pair, which is used to analyze the specific transactions. This representation captures the relationship between the buyer and seller in each transaction. These unprocessed raw transaction data include details about the land, the buyer and seller's identities, and financial transaction information. The second layer, i.e., the AI layer, receives the raw transaction data, which processes the raw data and identifies patterns that distinguish fraudulent and legitimate transactions using machine learning (ML) algorithms.

$$\mathrm{B} = buyer_{i\,i=1}^{\,N} \tag{16}$$

$$\mathrm{S} = seller_{i\,i=1}^{\,N} \tag{17}$$

$$\mathrm{L} = land_{i\,i=1}^{\,N} \tag{18}$$

$$(buyer_i, seller_i) \quad \text{where} \quad 1 \leq i \leq N \tag{19}$$

B. AI LAYER

The raw transaction data from the data layer is processed by the AI layer in the proposed system. AI identifies patterns that help to distinguish fraudulent and non-fraudulent transaction data. This layer consists of different AI elements, which are described below.

1) Dataset Description

The dataset [21] used in this study includes transaction data that was collected from the Ethereum blockchain. The initial dataset has 51 columns and 9841 rows. Eq. 20 represents the dataset used in the proposed system, where X is a matrix of real numbers with dimensions 9841×51.

$$\mathrm{D} = X \in \mathbb{R}^{9841 \times 51} \tag{20}$$

Each column in D represents a piece of transaction-related information, such as an address, flag, sent or received transactions, or the number of newly created contracts. The dataset also contains other significant columns such

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3451627

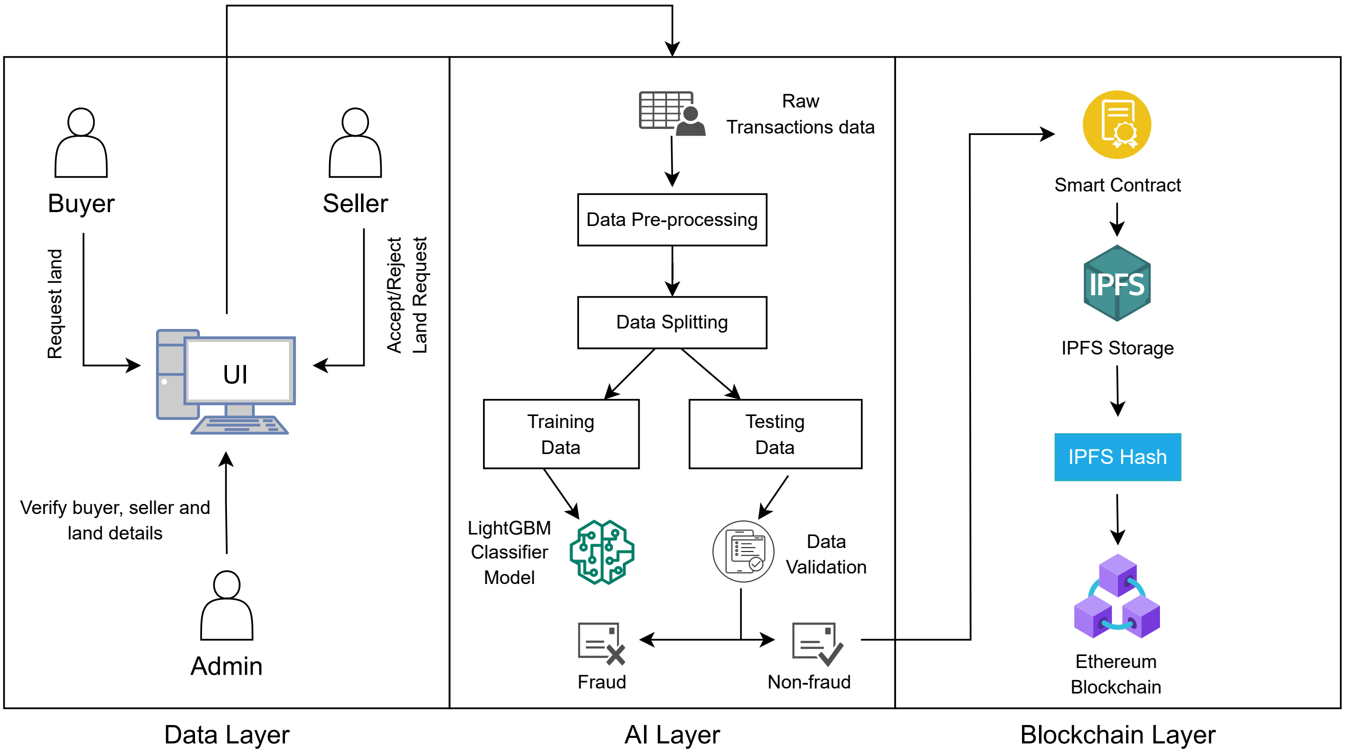

FIGURE 1: AI-enabled secure land registry system using blockchain for agriculture and industry 5.0.

as transaction hash, gas price, gas limit, and value. The sender's or recipient's address is shown in the transaction's address column. The flag (target variable/class label) column specifies the binary value, which is marked as 0 or 1, indicating whether the transaction was fraudulent or not. The number of transactions sent by the Ethereum address in the past is shown in the sent transaction column. The number of transactions received using the Ethereum address is shown in the received transaction column. Before feeding data into the model, there is a requirement to preprocess the data. For that, we used different preprocessing techniques, including data cleaning, feature selection, and data normalization.

2) Data Pre-Processing

In the data preprocessing, we initially prepared the raw data using data cleaning, feature selection, feature scaling, and data splitting sub-steps. A detailed explanation of these sub-steps is as follows.

1) Data cleaning - In the data cleaning, we identify and remove inaccurate, incomplete, or irrelevant data from a dataset. In the proposed system, data cleaning eliminates duplicate or missing values and inspects the data for any unusual values or irregularities from the dataset.
2) Handling missing data - After that, missing data were handled in some columns by imputing the column mean for continuous variables and the mode for categorical variables, as shown in Eq. 21 and Eq. 22.

$$\text{Mean} = \frac{\sum_{i=1}^{n} x_i}{n} \quad (21)$$

$$\text{Mode} = \text{Frequent value in the } D \text{ column} \quad (22)$$

3) Feature selection - Feature selection is the process of choosing a subset of relevant characteristics from a dataset. These columns contain useful information for determining fraud transactions, which are chosen as part of the proposed system's feature selection process. Columns like the transaction value, transaction timestamp, and the number of contract creations have the potential to be included in this.
4) Feature scaling - The entire raw data is in varied range and type. To standardize the data, we used feature scaling, which converts each feature between the range of 0-1. Each feature has a mean of 0 and a standard deviation of 1, which is used to scale the continuous variables.

$$\text{Standardization} = \frac{x - \text{Mean}(x)}{\text{Standard Deviation}(x)} \quad (23)$$

After the successful data preprocessing, we used data splitting, where we used an iterative functional loop that gives a suitable ratio to improve training accuracy. In the functional loop, we used train_test_split() from $fromsklearn.model\_selectionimporttrain\_test\_split$ library for data splitting.

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3451627

TABLE 2: Feature importance using Kruskal Wallis technique.

Feature name	Importance score
TimeDiffBetweenFirstAndLast_Mins_	374.0391
AvgMinBetweenReceivedTnx	72.3072
totalTransactions_includingTnxToCreateContract	52.2632
ReceivedTnx	33.3506
SentTnx	32.3142
avgValSent	22.0028
UniqueSentToAddresses	12.0114
UniqueReceivedFromAddresses	6.4828
AvgMinBetweenSentTnx	5.7579
TotalERC20Tnxs	4.2194
ERC20MinValSent	4.1324
ERC20AvgValSent	4.0494
ERC20MaxValSent	4.048
ERC20TotalEtherSent	3.9773
maxValSent	3.6485
minValueReceived	3.448
Training accuracy	83.19%

TABLE 3: Feature importance using Chi-squared technique.

Feature name	Importance score
AvgMinBetweenSentTnx	741.9552
AvgMinBetweenReceivedTnx	740.5483
totalTransactions_includingTnxToCreateContract	636.7954
totalEtherSent	630.6941
SentTnx	615.008
Index	614.36
minValueReceived	553.4257
totalEtherBalance	517.4098
maxValSent	480.5881
UniqueSentToAddresses	378.1511
avgValSent	375.0698
ReceivedTnx	316.0069
UniqueReceivedFromAddresses	272.5227
minValSent	113.3185
NumberOfCreatedContracts	32.9488
ERC20UniqSentAddr	24.6741
ERC20UniqSentTokenName	20.3062
Training accuracy	92.86%

5) Data Splitting : The preprocessed dataset was split into an 80:20 ratio of training and a testing set.

$$\text{Training Set} = 0.8 \times \text{Total Dataset} \tag{24}$$

$$\text{Testing Set} = 0.2 \times \text{Total Dataset} \tag{25}$$

In essence, we have used 80% of the data as training data and the rest, 20% of the data, is used to validate the AI training as testing data. Though we used different ratios to evaluate our training accuracy, with an 80:20 ratio, we achieved a notable improvement in training accuracy; hence, we used the aforementioned split of data in the AI training.

The preprocessing steps helped to ensure that the dataset was ready for classification and that the algorithm was able to interpret the data accurately. We applied different feature extraction techniques, such as Kruskal Wallis, Analysis of Variance (ANOVA), and Chi-squared test to reduce the dimensionality of the dataset from 51 columns to 16 columns as displayed in Tables 2,4, and 3.

We used features from Tables 2,4, and 3 for AI training using different classification algorithms. From the feature importance, we realized that the ANOVA technique has extracted essential features that have improved the training accuracy from 92.86% to 98.74%.

TABLE 4: Feature importance using ANOVA technique.

Feature name	Importance score
totalEtherSent	557.0353
SentTnx	514.342
TotalERC20Tnxs	214.5823
minValueReceived	429.725
maxValSent	417.3484
ERC20AvgValRec	398.9564
ERC20MaxValRec	369.9763
ERC20TotalEtherReceived	345.3494
totalTransactions_includingTnxToCreateContract	345.0733
TimeDiffBetweenFirstAndLast_Mins_	339.2868
UniqueSentToAddresses	316.6674
totalEtherBalance	311.6946
avgValSent	284.5691
ERC20UniqRecContractAddr	232.2537
ERC20UniqRecTokenName	232.1797
AvgMinBetweenSentTnx	228.8379
ERC20UniqRecAddr	228.6403
Training accuracy	98.74%

$$\text{D'} = X' \in \mathbb{R}^{9841\times 16} \tag{26}$$

After applying the data preprocessing techniques, the processed dataset D' now consists of 9841 rows and 16 columns, represented as $X' \in \mathbb{R}^{9841\times 16}$. Equation (26) represents the processed dataset. The resulting processed dataset will be used to train AI classification models.

3) Training the AI classification model

The AI classification model is trained on a dataset of non-fraud and fraud transactions. Each transaction is represented as a feature vector x_i with n dimensions, where n is the number of features used to describe the transaction. The dataset is divided into two sets, which are (i) the training set and (ii) the testing set. The AI model is trained using a supervised learning technique, i.e., classification. In the supervised learning phase, the model is trained using labelled data, where each transaction is labelled as either legitimate or fraudulent. We used different AI models, such as LR, SVM, RF, XGB, and LGBM, for this classification task. Among them, LGBM has offered a better training accuracy that has been elaborated in Section V. The reasons LGBM has better training accuracy are as follows:

- LGBM has an implicit optimization that can improve the speed and performance and is suitable for large dimensional datasets. Unlike SVM and LR, they take large training time to give their classification results
- It uses histogram-based decision tree learning to handle high dimensional data for efficient memory usage, thereby enhancing scalability.
- It has many hyperparameters that can be tuned to improve the performance of the AI model. We specifically used boosting_type = 'gdbt', which is a gradient decision tree, and a learning rate of 0.05 to improve training accuracy from 95.31% to 98.74%.

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3451627

Due to the aforementioned advantages, we used LGBM, among other AI models, for a comparative analysis on the basis of various statistical measures. The model (LGBM) learns to identify patterns in the data that are indicative of fraudulent activity and uses these patterns to make predictions on new, unlabeled data. The training process involves minimizing a loss function L, which measures the difference between the model's predictions and the true labels. The loss function in LGBM is typically a cross-entropy loss, which is defined in Eq. 27.

$$L = -\sum_{i=1}^{N} y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \tag{27}$$

where N is the number of transactions in the training set, y_i is the true label of transaction i, and $\hat{y}_i$ is the predicted label of transaction i. The model is trained using gradient descent, which iteratively updates the model's parameters in a direction that helps to minimize the loss function. The update rule is given in Eq. (28)

$$\theta_{t+1} = \theta_t - \alpha \nabla_\theta L \tag{28}$$

where θ is the model's parameters, α is the learning rate, and $\nabla_\theta L$ is the gradient of the loss function with respect to the model's parameters.

4) Classifying Transactions

Once the AI classification model has been trained, it is used to classify new transactions as legitimate or fraudulent. Each transaction is represented as a feature vector x and passed through the trained model to obtain a prediction $\hat{y}$. If the predicted label $\hat{y}$ is greater than a threshold value, the transaction is classified as legitimate and stored in the blockchain. If the predicted label is below the threshold, the transaction is classified as fraudulent and is not stored in the blockchain. The threshold value can be adjusted to balance between false positives and false negatives, depending on the desired level of security and efficiency. The classification process can be summarized as follows.

1) Given a transaction represented as a feature vector x, predict the label $\hat{y}$ using the AI classification model.
2) If $\hat{y} > \theta$, where θ is the threshold value, store the transaction in the blockchain and notify the user that the transaction has been approved.
3) If $\hat{y} \leq \theta$, do not store the transaction in the blockchain and notify the user that the transaction has been rejected due to fraudulent activity.

Algorithm 1 shows the step-by-step process of AI classification task for the proposed system.

5) Cost Savings and Security Benefits

By implementing an AI classification component in the blockchain-based land registry system, we save costs associated with storing fraudulent transactions in the blockchain. This is because fraudulent transactions are not stored in the blockchain, thus reducing the amount of storage space needed and lowering the overall cost of the system. Furthermore, the AI classification component enhances the security of the system by preventing fraudulent transactions from being stored in the blockchain. This reduces the risk of tampering with the land registry records and improves the trustworthiness of the system. In conclusion, the proposed system enhances the blockchain-based land registry system by adding an AI classification component that analyzes transactions for fraudulent activity before deciding whether to store them in the blockchain. AI implemented using an ML model has been shown to save costs and improve the security of the system.

Algorithm 1 Sequential flow of training the AI classification model.

Require: Transaction T represented as feature vector x
Ensure: Classification label y

```
Train the AI classification model on a dataset of legitimate and fraudulent transactions;
Obtain the threshold value θ;
Predict the label ŷ for transaction T using the trained model;
if ŷ > θ then
    Store transaction T in the blockchain and notify the user that the transaction has been approved;
    Set y = 1;
else
    Do not store transaction T in the blockchain and notify the user that the transaction has been rejected due to fraudulent activity;
    Set y = 0;
end if
return y;
```

C. BLOCKCHAIN LAYER

The last layer in the proposed scheme is the blockchain layer, which stores legitimate transaction information on the Ethereum blockchain. Initially, we implemented a smart contract containing different functions related to registering buyers and sellers, land registration, transfer ownership, approved data, and payment-related data. Before the deployment of this smart contract in the Ethereum blockchain test network. First, we performed a smart contract vulnerability assessment using the Slither tool to check the validity of the smart contract. We identify that the implemented smart contract is safe from different smart contract attacks. Further, we deploy the smart contract in the Sepolia test network using the Ethereum blockchain IDE platform. After deploying the smart contract, the proposed scheme first adds data to the IPFS-based decentralized file system. The reference (hash) of that data is included in the blockchain. Mathematically, the operations used in the blockchain layer can be represented using Eq. 29

$$T_n = t_n, H(t_n), IPFS(H(t_n)) \tag{29}$$

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3451627

Eq. (29) presents hashed transaction data addition process, where Tn represents the non-fraudulent transaction data, t_n represents the transaction data itself, $H(t_n)$ represents the hash value generated for the transaction data, and IPFS $(H(t_n))$ represents the transaction data stored on IPFS using the hash value. This hash value occupied minimum space in the blockchain, leading to higher performance in data retrieval and data processing. Additionally, it helps to process more transactions in a minimum time, leading to higher throughput and low latency. As a result, it improves the scalability of the proposed scheme.

The deployment of the entire proposed scheme in real-world scenarios is as follows.

- AI Layer - The land registry system has an intrusion detection system specifically meant for illicit transaction detection. We can deploy the best AI model (in our case - LGBM) on the detection system to efficiently classify illicit and non-illicit land registry transactions. The detection system discards illicit (fraudulent) transactions and allows only non-fraudulent transactions to be a part of the blockchain network.
- Blockchain layer - We have a web 3.0 application programming interface (API) to connect the AI model with the blockchain ledger. In that manner, the non-fraudulent transactions arrived at the blockchain layer. Before that, the smart contract verifies the authenticity of the land records, and based upon a successful verification, it forwards the land records to the IPFS. Here, IPFS is connected with the land registry system and the blockchain network, wherein each land record will get a unique ID for storage and retrieval purposes. IPFS will compute the hash of each land record and send the hash to the immutable ledger. Whenever the land administrator has to fetch the land record data, it calls the unique ID to retrieve the hash from the blockchain and, later, the raw data from the IPFS.

Consecutively, it offers a safe, reliable, and extremely fast platform for buying and selling land by first ensuring that all transactions are transparent and immutable. Second, it provides a distributed and decentralized platform for storing transaction data, which ensures that the data is accessible from anywhere in the world.

V. RESULTS AND DISCUSSIONS

This section discusses the details of the implementation, such as smart contract parameters, tools, libraries, and smart contract functions used to develop the proposed system. In addition, it shows the performance of the proposed system using statistical measures (e.g., training accuracy, precision, recall, F1-score, and log loss score), scalability, sustainability, and blockchain size comparison.

A. EXPERIMENTAL SETUP AND TOOLS

The proposed system consists of different smart contracts implemented in the Remix integrated development environment (IDE) [22] and different ML algorithm results implemented on the Ethereum Transactions dataset. The smart contracts are implemented on the public EB network using the solidity programming language with version $v0.8.16$.. It has various user-defined functions, such as $registrybuyer$, $registryseller$, $requestland$, $addland$, and many more. Next, the solidity compiler with version 0.8.17+commit.8df45f5f is used to compile the different smart contracts in the Remix IDE. The compiled smart contracts are forwarded for deployment on the EB network. For that, the injected provider Metamask environment is selected with a gas limit of 4000000. In that regard, a Metamask wallet is used to deploy smart contracts on the EB network and to perform different smart contract transactions. Furthermore, before deploying smart contracts to the EB network, they are validated using the slither analyzer tool for security and privacy concerns [23]. It finds the source code vulnerabilities from smart contracts so trivial code can be modified and protects it from malicious users. Additionally, IPFS storage is used for storing off-chain land registry data to enhance the security and performance of the proposed system. The proposed system uses 5 ML algorithms: Light Gradient Boosting Machine (LightBGM), Extreme Gradient Boosting (XGBoost), Decision Tree (DT), Logistic Regression (LR), and Support Vector Machine (SVM). We extract result-oriented data from the above-mentioned implementations, such as blockchain scalability, ROC curve, log loss score, accuracy comparison graph, precision, recall, and F1-score comparison graph, and visualize using the Python-based Matplotlib library.

B. AI MODELS ROC CURVE

The ROC curve is a graphical representation of the performance of a classification model. In this study, we compare the performance of five different models: LightGBM, XGBoost, DT, LR, and SVM. The ROC curve plots the true positive rate (TPR) against the false positive rate (FPR) for different classification thresholds. The TPR is the ratio of correctly classified positive instances to the total number of positive instances. At the same time, the FPR is the ratio of incorrectly classified negative instances to the total number of negative instances. In our analysis, as shown in FIGURE 2b, LightGBM achieved the best result among the five models, as it shows the highest TPR for a given FPR, indicating a better trade-off between sensitivity and specificity. This means that LightGBM is able to correctly classify more positive instances while minimizing the number of false positives. Compared to LightGBM, the other models showed varying levels of performance. XGBoost, DT, and LR showed comparable results, while SVM had the lowest performance. Overall, these findings suggest that LightGBM is a strong performer among other AI models tested in the proposed study.

C. AI MODELS ACCURACY COMPARISON

In this study, we compare the accuracy of five different classification models, namely LightGBM, XGBoost, DT,

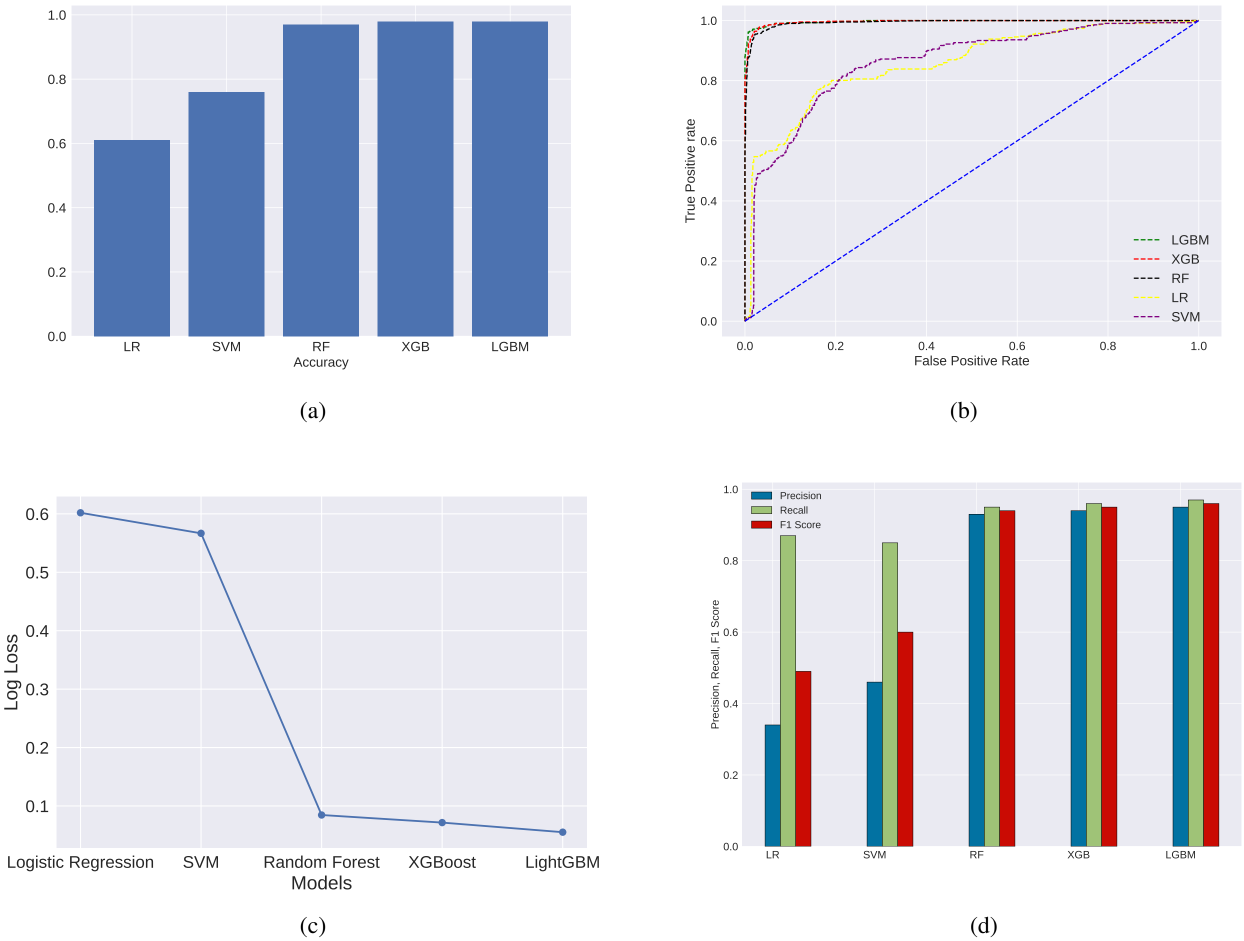

FIGURE 2: Comparison of AI models on the basis of statistical measures (a) accuracy, (b) ROC, (c) log-loss score, (d) AI models precision, recall, and F1-score comparison.

LR, and SVM. The accuracy of a classification model measures the percentage of correctly classified instances. In that context, FIGURE 2a revealed that both LightGBM and XGBoost achieved the highest accuracy among other AI models. This indicates that both models were able to accurately classify a larger number of instances compared to the other models. DT, LR, and SVM showed varying levels of performance, with LR being the least accurate of the models tested. The high accuracy of both LightGBM and XGBoost may be attributed to their ability to effectively handle high-dimensional datasets, as well as their ability to handle imbalanced class distributions. These features make them ideal for tasks in which high accuracy is critical. Overall, these findings suggest that both LightGBM and XGBoost are strong performers among the models tested in this study and may be well-suited for classification tasks in similar contexts where accuracy is a key metric.

D. AI MODELS PRECISION, RECALL AND F1-SCORE COMPARISON

We want to mention that accuracy alone may not be a reliable parameter for evaluating the performance of a classification model. Therefore, to support the accuracy parameter, we utilized other statistical measures, such as precision, recall, and F1-score, for all the adopted classification algorithms. From FIGURE 2d, we can infer that LightGBM achieved the highest precision, recall, and F1-score among all the models tested. Here, precision measures the proportion of true positives among all predicted positive instances, while recall measures the proportion of true positives among all actual positive instances. The F1-score is the harmonic mean of precision and recall. XGBoost and DT showed comparable results in terms of precision, recall, and F1-score, while SVM and LR had the lowest performance. The high precision, recall, and F1-score of LightGBM may be attributed to its ability to effectively handle imbalanced datasets and its ability to accurately classify both positive and negative instances. These findings suggest that LightGBM may be

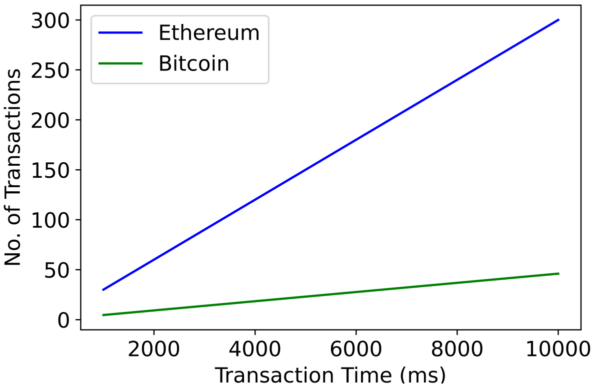

FIGURE 3: Blockchain (Ethereum vs Bitcoin) scalability comparison.

a good choice for classification tasks in similar contexts where high precision and recall are critical. In conclusion, our analysis underscores the importance of using multiple evaluation metrics in classification tasks to gain a more comprehensive understanding of a model's performance.

E. AI MODELS LOG LOSS SCORE

Further, we assessed the AI model's performance using log loss score, which is a widely used evaluation metric for binary classification tasks. It measures the performance of the AI model in terms of the probability estimates it provides for each class. By comparing the log loss scores of all the incorporated AI models, we can better understand their ability to provide accurate probability estimates, handle imbalanced datasets, and achieve efficient performance. FIGURE 2c shows that LR achieved the highest log loss score among all the utilized AI models, indicating that it had the lowest probability estimates for the correct class. LightGBM, on the other hand, had the lowest log loss score, indicating that it provided the most accurate probability estimates for each class. The superior performance of LightGBM in terms of log loss may be attributed to its ability to effectively handle imbalanced datasets and its capacity to provide accurate probability estimates for each class. Our findings highlight the importance of using multiple evaluation metrics when evaluating the performance of classification models. By assessing a model's performance using various metrics, we can obtain a more comprehensive understanding of its strengths and limitations, allowing us to make more informed decisions when selecting a model for a particular task.

F. ABLATION STUDY BASED ON HYPERPARAMETER TUNING

The meaning of ablation - process of systematically removing or altering components to assess the model's performance. The meaning is retained here since the components become our hyperparameters (components= hyperparameters); we evaluate the model's performance if that particular hyperparameter is used or not. This section shows the ablation study, wherein we used different hyperparameter tuning to analyze the improvement of training accuracy in different AI models. Table 5 shows the hypertuning parameters of the LR model where we first used regularization (penalty) L1 and L2. So, when we change the regularize from L1 to L2, the training accuracy of LR improves from 54.46% to 55.23%. The reason for this improvement lies in the L2 regularizer that offers a non-sparse solution and operates on squares of the LR model's parameters. Further, using the 'saga' solver, we have improved the training accuracy from 55.23% to 61.53%.

TABLE 5: Improvement in training accuracy from 54.46% to 61.53% using LR hyperparameter tuning.

Parameter	**Value**	**Training accuracy**
Penalty	L1 $\rightarrow$ L2	54.46% $\rightarrow$ 55.23%
Solver	lbfgs, newton-g	55.23%
	saga	61.53%
max_iter	100 $\rightarrow$ 1000	55.23% $\rightarrow$ 61.53%

TABLE 6: Improvement in training accuracy from 75.75% to 78.39% using SVM hyperparameter tuning.

Parameter	**Value**	**Training accuracy**
C	1.0 $\rightarrow$ 3.0	74.69% $\rightarrow$ 75.65%
gamma	svc $\rightarrow$ gamma	75.65% $\rightarrow$ 78.28%
kernel	linear $\rightarrow$ sigmoid	78.28% $\rightarrow$ 78.39%

Similarly, in Table 6 we used SVM hypertuning parameters, such as C value from [1-3] to optimize the margin between data points. C values indicate the percentage of misclassification to be avoided by the SVM model. Using C = 3, we have improved the training accuracy from 74.69% to 75.65%. Further, we used 'gamma' to make a decision boundary between the points. The higher the value of gamma, the closer the points are in the decision boundary, but it may sometimes be prone to overfitting. Hence, we used gamma = $\approx$ 0.35 and kernel = sigmoid to attain a final training accuracy of 78.39%. Table 7 shows the training accuracy improvement in XGB model. We used max_depth, which specifies the maximum depth of the tree, wherein a trade-off is required between bias and variance to reduce the chance of overfitting. Therefore, we used max_depth between 3-7 to achieve 96.38% training accuracy. Next, we used 'gamma' and 'scale_pos_weight' to deal with the unbalanced dataset. With these tuning parameters, we have improved the training accuracy of the XGB model from 96.38% to 97.23%. Table 8 displays the hypertuning parameters of RF to improve its

TABLE 7: Improvement in training accuracy from 96.38% to 97.23 using XGB hyperparameter tuning.

Parameter	**Value**	**Training accuracy**
max_depth	3-7	96.38%
gamma	0.1-0.3	97.23%
scale_pos_weight	1	97.23%

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3451627

TABLE 8: Improvement in training accuracy from 95.89% to 96.24% using RF hyperparameter tuning.

Parameters	Value	Training accuracy
n_estimators	100-120	95.89%
max_features	sqrt, log	95.89%
criterion	mse	95.89%
max_depth	5-7	95.89%
applying gridsearchCV	estimators = rf, cv = 4, n_jobs=1	96.24%

TABLE 9: Improvement in training accuracy from 92.08% to 98.74% using LGBM hyperparameter tuning.

Parameter	Value	Training accuracy
num_leaves	10	92.08%
min_data	100	92.08%
boost_from_average	True	92.08%
early_stopping_round	50	95.31%
learning_rate	0.05	98.74%
boosting_type	'gdbt'	98.74%

training accuracy. We tried multiple approaches to improve the training accuracy, but it was stuck at 95.89%. Then, we used gridsearchCV, which exhaustively searches over a specified parameter grid and evaluated the model for each combination of parameters using cross-validation (CV). We used parameters, such as estimators = 100-120, cross-validation (CV) = 4, and n_jobs = 1, to slightly improve the training accuracy from 95.89% to 96.24%. Table 9 shows the tuning parameters utilized while improving the training accuracy of the LGBM model for the classifications task. With num_leaves (maximum number of leaves in one tree) = 10, min_data (minimum number of samples a leaf must have) = 100, and boost_from_average (for improving convergence) = true, we achieved the training accuracy = 92.08%. Next, we used learning rate = 0.05 (control the step size to reduce the model errors) and boosting_type = 'gdbt', which optimizes the differentiable loss functions to improve the training accuracy, i.e., 98.74%. With that analysis, we finalized that the LGBM has the best classifier model (higher training accuracy - 98.74%) among other AI models to classify malicious and non-malicious land records.

G. SMART CONTRACT VULNERABILITY ANALYSIS

The vulnerability assessment of smart contracts is crucial before deploying them to the blockchain network. In this study, smart contracts developed in the Solidity programming language are assessed for their vulnerability using the Slither Solidity source analyzer tool [23]. It checks for various code vulnerabilities, such as deep stacks, integer underflows, and wrong naming methods. Solc v0.8.16 is used for the vulnerability assessment of smart contracts using Slither. FIGURE 4 shows that no vulnerabilities were found after

```
pradeepa@CHENDG3V0HR3:~$ slither /home/pradeepa/Smart\ Contract/LandRegistratationSystem.sol
/home/pradeepa/Smart Contract/LandRegistratationSystem.sol analyzed (1 contracts with 84 detectors), 0 result(s) found
```

FIGURE 4: Vulnerability assessment of smart contract.

assessing the smart contracts with Slither, indicating that the smart contracts developed for the proposed system are secure against all vulnerabilities.

H. BLOCKCHAIN SCALABILITY COMPARISON

The scalability of the blockchain is a key factor in managing transactions over time without sacrificing security and performance. It assesses how well the blockchain scales, that is, whether it becomes slow, clogged, and vulnerable to security threats as the number of users and transactions rises. The scalability comparison between the proposed Ethereum-based system and the conventional blockchain network is shown in FIGURE 3. Here, the y-axis displays the total number of transactions, and the x-axis displays the transaction time in milliseconds. It can be seen from the

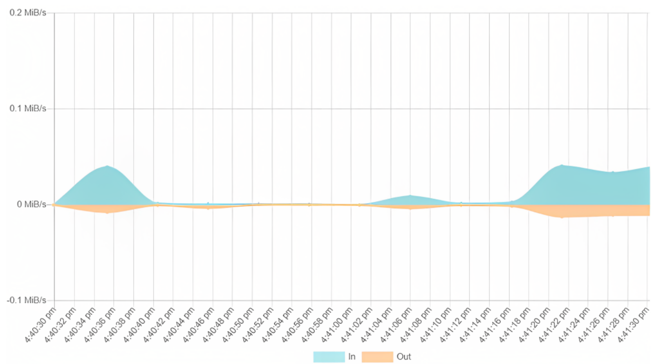

FIGURE 5: Input and output bandwidth utilization by IPFS.

FIGURE 3 that the proposed system is more scalable than the established blockchain network. This is so because the suggested system makes use of IPFS's core advantages, such as decentralized storage, lower storage costs, and increased file availability. As a result of its ability to distribute files across numerous network nodes, it is more resistant to risks associated with centralized storage, such as censorship and single points of failure. The Ethereum-based blockchain uses technology advancements that the traditional blockchain built on Bitcoin does not. As a result, it is less scalable than the Ethereum blockchain. Furthermore, it is imperative to emphasize that within the blockchain network, only non-fraudulent land data from the AI layer is selected for storage, and fraudulent data gets discarded. This results in reducing the computational overhead from the blockchain network since it has to only process non-fraudulent data and not all data. Thus, the reduction of computational overhead not only enhances response times but also augments the scalability of the blockchain network (as shown in FIGURE 3).

I. IPFS BANDWIDTH UTILIZATION

In the proposed system, IPFS storage is utilized to reduce the cost of storing all the data in the Ethereum nodes. This results in a faster process of storing and retrieving data due to its low bandwidth utilization. For the implementation of this approach, IPFS is installed into the local system and connected with peers in a decentralized network. Data can be stored and retrieved using a hash key. FIGURE 5 displays the input and output bandwidth for the data storage, with a sample time of 1 minute. The graph indicates a significantly low bandwidth usage in the proposed system.

J. SMART CONTRACT IMPLEMENTATION INTERFACE AND GAS COST

FIGURE 6 shows the smart contract functions for the secure land registry systems implemented in the Remix IDE and deployed on the public Ethereum-based blockchain. These functions act as data validation by allowing only valid data to be entered into the blockchain systems. It ensures that the data is accurate, consistent, and complete, which is essential for making informed decisions based on the data. Further, all the transactions (i.e., executing the smart contract functions) performed in the blockchain require gas costs to perform the transaction. The proposed system uses a PoW consensus mechanism, where miners execute smart contracts on the network. However, miners need compensation for their efforts, so users must pay a transaction fee, known as gas, to cover the cost of executing the smart contract. Gas cost is calculated using the gas profiler tool provided by the Remix IDE. FIGURE 7 shows the graph of transaction cost and execution cost of the main functions of the smart contract in gas units. Table 10 displays all the smart contract functions implemented for the proposed system. The graph shows transaction and execution costs without AI and also with AI. It is clearly seen from the graph that transaction and execution costs are reduced for all the smart contract functions using AI.

TABLE 11 presents the comparative analysis of the proposed scheme with different parameters, including blockchain used, AI used smart contract vulnerability assessment, and IPFS. We analyzed the existing land registration scheme's use of blockchain, where they developed a smart contract and deployed it in the Ethereum blockchain. The authors included the results related to interface and cost analysis. However, existing schemes did not verify the smart contract before deploying it. Moreover, [6], [7], [8] were not concerned about the scalability of their proposed approach, leading to lower performance of their land registration scheme. Additionally, the state-of-the-art work did not validate the intent of the data before storing it in the blockchain. To address these issues, we included blockchain and AI-based land registration schemes,

TABLE 10: Smart contract functions.

Code	Function	Code	Function
F1	addBuyer	F2	addSeller
F3	addLand	F4	verifyBuyer
F5	verifySeller	F6	verifyLand
F7	addLandRequest	F8	approveLandRequest
F9	rejectLandRequest		

TABLE 11: Comparison of the proposed scheme with the existing schemes.

Existing schemes	Blockchain	AI	Vulnerability assessment	IPFS
[6]	✓	-	-	-
[7]	✓	-	-	-
[8]	✓	-	-	-
[9]	✓	-	-	✓
Proposed scheme	✓	✓	✓	✓

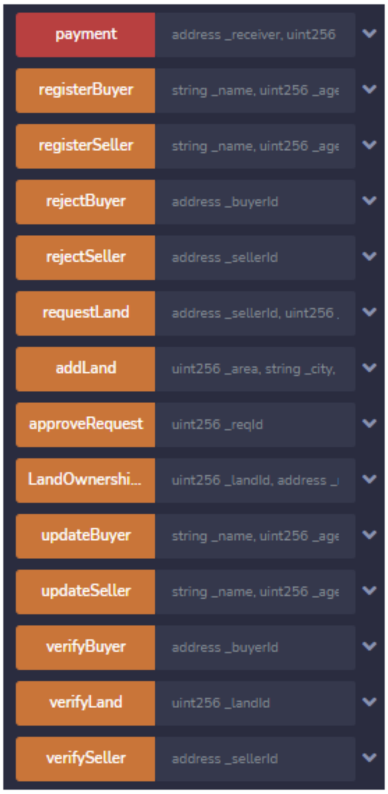

FIGURE 6: Smart contract user-defined functions.

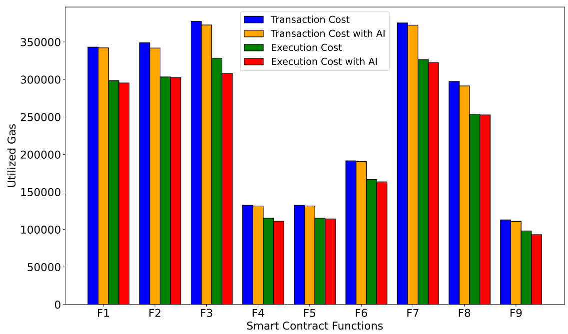

FIGURE 7: Gas cost of proposed smart contract functions in terms of transaction cost and execution cost.

where we validated the intent of the land registration data using AI, and only valid data was stored in the blockchain. Here, in the blockchain, we verified smart contracts using its vulnerability assessment to protect the smart contract from various vulnerabilities. Besides, we also used an IPFS-based decentralized file system that aids in improving the scalability of the proposed scheme. These features mark the novelty of the proposed scheme.

VI. CONCLUSION AND FUTURE WORK

This research proposes an AI and blockchain-based land registry system for agriculture and industry 5.0 to solve the challenges faced by traditional land registry systems. The proposed system provides a more secure, transparent, and efficient land registry system that can benefit citizens and the economy as a whole. First, a standard land registry dataset is used to efficiently bifurcate the fraudulent and non-fraudulent land data. For that, different AI algorithms, such as LightBGM, XGBoost, DT, LR, and SVM, are incorporated to perform binary classification on the land registry dataset. The fraudulent data is discarded from the proposed system, and only non-fraudulent data is forwarded to the blockchain-based land registry system. In the blockchain, we designed a land registry smart contract that efficiently validates the land ownership data. Once validated, the land data is forwarded to the IPFS, where a unique hash is computed for each land record, which can be accessed using a unique content identifier. The hashed land data is then relayed to the blockchain's immutable ledger for secure data storage. This reduces the risk of corruption and eliminates the need for intermediaries. The proposed system is evaluated by considering different evaluation metrics, such as AI's statistical measures (accuracy, precision, recall, F1-score, and ROC), blockchain's scalability, smart contract vulnerability assessment, and IPFS bandwidth utilization.

REFERENCES

[1] Z. Iqbal, M. A. Khan, M. Sharif, J. H. Shah, M. H. ur Rehman, and K. Javed, "An automated detection and classification of citrus plant diseases using image processing techniques: A review," Computers and Electronics in Agriculture, vol. 153, pp. 12–32, 2018.

[2] I. Haider, M. A. Khan, M. Nazir, A. Hamza, O. Alqahtani, M. T.-H. Alouane, and A. Masood, "Crops leaf disease recognition from digital and rs imaging using fusion of multi self-attention rbnet deep architectures and modified dragonfly optimization," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 17, pp. 7260–7277, 2024.

[3] F. D. Rodrigues, J. M. Alonso, and P. M. R. Castro, "Sicap - cadastral information and property management system," in 2022 17th Iberian Conference on Information Systems and Technologies (CISTI), pp. 1–6, 2022.

[4] S. N. Pereira, N. Tasnim, M. S. Rahman, S. N. Tumpa, M. M. Karim, R. S. Rizon, N. S. Khan, and R. A. Remon, "Multi-channel approach towards digitizing the land management system of bangladesh," in 2018 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE), pp. 68–71, 2018.

[5] E. Choudhury, M. Ridwan, M. A. Awal, and S. Hosain, "A web-based land management system for bangladesh," in 14th International Conference on Computer and Information Technology (ICCIT 2011), pp. 321–326, 2011.

[6] A. L. Shrivastava and R. Kumar Dwivedi, "Blockchain-based secure land registry system using efficient smart contract," in 2023 International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), pp. 165–170, 2023.

[7] G. Kusuma, C. Rupa, S. Reshma, and G. Rochana, "Secure storage of land records and implementation of land registration using ethereum blockchain," in 2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS), pp. 404–409, 2023.

[8] S. Dubey, D. Vinod, A. Gupta, and R. K. Dwivedi, "Secure land registration management via ethereum blockchain," in 2023 International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), pp. 185–191, 2023.

[9] R. Sidharthan and V. R. Balasaraswathi, "Secured land registration using ethereum blockchain and ipfs," in 2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC), pp. 1449–1454, 2024.

[10] G. Kusuma, C. Rupa, S. Reshma, and G. Rochana, "Secure storage of land records and implementation of land registration using ethereum blockchain," in 2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS), pp. 404–409, 2023.

[11] A. L. Shrivastava and R. Kumar Dwivedi, "Blockchain-based secure land registry system using efficient smart contract," in 2023 International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), pp. 165–170, 2023.

[12] N. Ncube, B. Mutunhu, and K. Sibanda, "Land registry using a distributed ledger," in 2022 IST-Africa Conference (IST-Africa), pp. 1–7, 2022.

[13] M. I. Khalid, J. Iqbal, A. Alturki, S. Hussain, A. Alabrah, and S. S. Ullah, "Blockchain-based land registration system: A conceptual framework," Applied Bionics and Biomechanics, vol. 2022, 2022.

[14] R. Suganthe, N. Shanthi, R. Latha, K. Gowtham, S. Deepakkumar, and R. Elango, "Blockchain enabled digitization of land registration," in 2021 International Conference on Computer Communication and Informatics (ICCCI), pp. 1–5, 2021.

[15] I. Mishra, Supriya, A. Sahoo, and M. Vivek Anand, "Digitalization of land records using blockchain technology," in 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), pp. 769–772, 2021.

[16] M. Nandi, R. K. Bhattacharjee, A. Jha, and F. A. Barbhuiya, "A secured land registration framework on blockchain," in 2020 Third ISEA Conference on Security and Privacy (ISEA-ISAP), pp. 130–138, 2020.

[17] R. Khan, S. Ansari, S. Sachdeva, and S. Jain, "Blockchain based land registry system using ethereum blockchain," Xi'an Jianzhu Keji Daxue Xuebao/Journal of Xi'an University of Architecture & Technology, vol. 12, pp. 3640–3648, 04 2020.

[18] S. A. Gollapalli, G. Krishnamoorthy, N. S. Jagtap, and R. Shaikh, "Land registration system using block-chain," in 2020 International Conference on Smart Innovations in Design, Environment, Management, Planning and Computing (ICSIDEMPC), pp. 242–247, 2020.

[19] D. Shinde, S. Padekar, S. Raut, A. Wasay, and S. S. Sambhare, "Land registry using blockchain - a survey of existing systems and proposing a feasible solution," in 2019 5th International Conference On Computing, Communication, Control And Automation (ICCUBEA), pp. 1–6, 2019.

[20] P. Singh, "Role of blockchain technology in digitization of land records in indian scenario," vol. 614, p. 012055, dec 2020.

[21] V. Aliyev, "Ethereum fraud detection dataset, version 1." https://www.kaggle.com/datasets/vagifa/ethereum-frauddetection-dataset, 2021. Retrieved March 20, 2023.

[22] "Remix solidity ide, 2022." https://remix.ethereum.org/. Accessed: 2022-05-05.

[23] "Slither security analysis tool." https://github.com/crytic/slither. Accessed: 2023-02-19.

FESHALBHAI NAGUJI is currently pursuing the M.Tech. degree with the Institute of Technology, Nirma University, Ahmedabad, Gujarat, India. His research interests include blockchain, Hyperledger Fabric, deep learning, and machine learning.

NILESH KUMAR JADAV is a Full-Time Ph.D. Research Scholar in the Computer science and Engineering Department at Nirma University, Ahmedabad, Gujarat, India. He received his Bachelor as well as Master of Technology from the Gujarat Technological University (GTU), Gujarat, India in 2014 and 2018, respectively. He has authored/co-authored publications (including papers in SCI Indexed Journal and IEEE ComSoc sponsored International Conference). Some of his research findings are published in top-cited journals and conferences such as IEEE INFOCOM, IEEE ICC, IJCS and many more. His research interest includes, Artificial Intelligence, Network security, 5G Communication Network, Blockchain Technology. He is also an active member of ST Research Laboratory (www.sudeeptanwar.in).

SUDEEP TANWAR (M'15, SM'21) is currently working as a Professor with the Computer Science and Engineering Department, Institute of Technology, Nirma University, India. He is a Visiting Professor at Jan Wyzykowski University, Polkowice, Poland; and the University of Pitesti, Pitesti, Romania. He has authored two books, edited 13 books, and more than 500+ technical papers, including top journals and top conferences, such as IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, IEEE WIRELESS COMMUNICATIONS, IEEE Network, ICC, GLOBECOM, and INFOCOM. He initiated the research field of blockchain technology adoption in various verticals, in 2017. His H-index is 74. He actively serves his research communities in various roles. His research interests include blockchain technology, wireless sensor networks, fog computing, smart grid, and the IoT. He is a member of the Technical Committee on Tactile Internet of the IEEE Communication Society. He is a Senior Member of CSI, IAENG, ISTE, and CSTA. He has been awarded the Best Research Paper Awards from IEEE GLOBECOM 2018, IEEE ICC 2019, and Springer ICRIC-2019. He has served many international conferences as a member of the organizing committee, such as the Publication Chair for FTNCT-2020, ICCIC 2020, and WiMob2019; a member of the Advisory Board for ICACCT-2021 and ICACI 2020; the Workshop Co-Chair for CIS 2021; and the General Chair for IC4S 2019 and 2020 and ICCSDF 2020. He is serving on the editorial boards for Frontiers of Blockchain, Cyber Security and Applications, Computer Communications, the International Journal of Communication Systems, and Security and Privacy.

GIOVANNI PAU is an Associate Professor at the Faculty of Engineering and Architecture, Kore University of Enna, Italy. Prof. Pau received his Bachelor's degree in Telematic Engineering from the University of Catania, Italy, and both his Master's degree (cum Laude) in Telematic Engineering and Ph.D. from Kore University of Enna, Italy. Prof. Pau is the author/co-author of more than 100 refereed articles published in journals and conference proceedings. He is a Member of the IEEE (Italy Section) and has been involved in several international conferences as a session Co-Chair and Technical Program Committee member. Prof. Pau serves/served as a leading Guest Editor in special issues of several international journals and is an Editorial Board member as Associate Editor of several journals, such as IEEE Access, Wireless Networks (Springer), EURASIP Journal on Wireless Communications and Networking (Springer), Wireless Communications and Mobile Computing (Hindawi), and Sensors (MDPI), to name a few. His research interests include Wireless Sensor Networks, Fuzzy Logic Controllers, Intelligent Transportation Systems, Internet of Things, Smart Homes, and Network Security.

GULSHAN SHARMA received the B.Tech., M.Tech., and Ph.D. degrees. He is currently working as a Senior Lecturer with the Department of Electrical Engineering Technology, University of Johannesburg. He is also a Y Rated Researcher with NRF South Africa. His research interests include power system operation and control and application of AI techniques to the power systems. He is working as an Academic Editor of International Transactions on Electrical Energy System (Wiley) and a Regional Editor of Recent Advances in E & EE (Bentham Science).

FAYEZ ALQAHTANI is currently a full Professor with the Software Engineering Department, College of Computer and Information Siences, King Saud University (KSU). He was appointed as the Director ofthe Computer Division at the Deanship of Student Affairs. He is also a member of anumber of academic and professional associations,such as the Association for Computing Machinery (ACM),the Australian Computer Society, and the Association for Information Systems. He has conductedresearch projects in several areas of information and communication technology, such as Web 2.0, information security, enterprise architecture, software process improvement, the Internet of Things, and fog computing. He has participated in several academic events.

AMR TOLBA received the M.Sc. and Ph.D. degrees from the Mathematics and Computer Science Department, Faculty of Science, Menoufia University, Egypt, in 2002 and 2006, respectively. He is currently a full professor of computer science at King Saud University (KSU), Saudi Arabia. He has authored/coauthored over 125 scientific papers in top-ranked (ISI) international journals (such as IEEE IoT, ACM TOIT, IEEE CEMAG, IEEE ACCESS, IEEE SYSTEMS, FGCS, JNCA, NC&A, JAIHC, COMNET, COMCOM, P2PNET, VCOM, WWWJ,... etc.). Prof. Tolba served as a technical program committee (TPC) member at several conferences (Such as DSIT 2022, CICA2022, EAI MobiHealth 2021, DSS 2021, AEMCSE 2021, ICBDM 2021, ICISE 2021, DSS 2020, NCO 2020, ICISE2019, ICCSEA 2019, DSS 2019, FCES 19, ICISE 2018, ESG'18, Smart Data'17, NECO 2017, NC'17, WEMNET'17, NET'17, Smart Data'16, ...). He has been included in the list of the top 2% of influential researchers globally (prepared by scientists from Stanford University, USA) during the calendar years 2020, 2021 and 2022 respectively. Prof. Tolba served as Associate Editor/Guest Editor for several ISI journals. He has translated four books into the Arabic language. His main research interests include artificial intelligence (AI), the Internet of Things (IoT), data science, and cloud computing. Prof. Tolba is a Senior Member of IEEE.