SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V13I5P105

Volume 13 Issue 5, 41-57, May 2026

Original Article

Decentralized Predictive Intelligence for Fabricated Entity Detection in Blockchain-Integrated Identity Systems

Naveen Gajji[1], Ramesh babu Akarapu[2]

[1,2]*Department of CS & AI, SR University, Warangal, Telangana, India.*

[1]*Corresponding Author : naveengsn@gmail.com*

Received: 06 February 2026 Revised: 07 March 2026 Accepted: 06 April 2026 Published: 27 May 2026

Abstract *- The study postulates a novel approach that leverages blockchain technology for use in predictive analytics, which will serve as a mechanism for predicting the fabrication of an entity in the course of digitally verifying identity. The introduction of blockchain technology has offered a platform that is transparent, secure, and immutable, which has led to the extraction of sophisticated predictive models that eliminate and detect the threats presented by synthetic identity fraud. The approach that authors have suggested has been proven to be highly precise with regard to the estimation of the probability of deceptive behavior by scanning the route and errors contained in the transactional and behavioral data stored on the blockchain. This article has presented a well-founded system, based on a combination of historical data analysis and machine learning algorithms, and which uses the features of blockchain technology to produce a strong system that validates digital identities. The system has not only produced a strengthened security and integrity of digital identities, but also gives a proactive approach to preventing fraud through predictive analysis, which found early indications of entity fabrication, permitting early intervention and prevention of the fraud occurring. Additionally, making use of this predictive analysis system based on blockchain technology, the system has shown a significantly increased efficiency and reliability in reference to the procedure employed to confirm the digital identities. The number of mistakes in the form of false positives has been significantly reduced, and thus, the trust between the provider and the user services has been enhanced. The predictive analytics model offered integrity and transparency by decentralization of blockchain for enhanced user confidence in internet transactions. The research has offered a distinct and effective means of predicting and averting the incidence of synthetic identity fraud on the internet. The predictive analytics platform based on the blockchain has addressed the current problem of online identity verification and developed a new tier of security associated with online interaction. The offered approach can transform the sphere of digital identity verification and improve its level of security, efficiency, and reliability.*

Keywords *- Blockchain; Digital Identity Verification; Gradient Boosting Machine; Mutual Information (MI); Predictive Analytics.*

1. Introduction

Blockchain is an agreement system that minimizes the requirements of trust in data. The convergence of Blockchain with next-gen technologies such as artificial intelligence (AI), machine learning (ML), and analytics has paved the way for secure identity management, transparent data sharing, and decentralized governance. Numerous research works have focused on utilizing blockchain technology in applications related to identity management, auditing, data analytics, and validating digital credentials. Blockchain has also been investigated for data analytics in understanding transaction behaviors and network patterns. Al-Azzoni et al. presented the methodologies trained on data for executing data analytics on blockchain platforms to discover valuable insights from decentralized ledger data. Their study shows that both trends and anomalies in transactions over time can be detected from the analysis of blockchain data, as well as characteristics of system performance that help to keep transparency on distributed systems [1]. Some researchers conducted in this category include Wang's approach of studying entity recognition and transaction characteristic analysis on Bitcoin blockchain networks. The study emphasizes the application of analytical algorithms to expose transaction entities and examine patterns of blockchain activity, which have practical advantages for extensive security analysis [14]. The old-style centralized identity systems have a lot of limitations, which have caused the blockchain-based identity management to be studied extensively. Traditional systems depend on centralized databases prone to data breaches, unauthorized access, and

identity theft. Belurgikar et al. proposed a blockchain-based identity verification framework with improved data security and transparency that stores the identity verification records on an immutable blockchain ledger [2]. Their team had done research that proves Blockchain to be one of the solutions, as it provides a tamper-proof environment for validating digital identity.

Further research in this area is exploring interoperability between different identity systems. Dąbrowski and Donna [3] [4] introduced a blockchain-based mechanism for identity discovery and verification across heterogeneous identity management systems. Their solution allows communication and trust between numerous identity platforms by ensuring they use blockchain as a decentralized verification layer. This method significantly improves the level of interoperability and reduces dependence on central identity providers. Another method that has been widely studied as an alternative to centralized identity management systems is known as the self-sovereign identity (SSI). [9] Manhani; Queiroz - Blockchain in the Application of Self-Sovereign Identity Systems to Smart City Environments. The study points to the necessity of people to have their personal data and have the ability to control and disseminate identity information without relying on third parties. Decentralized identity systems can better protect privacy and security while driving faster services in smart cities.

The use of blockchain-based tokens is another innovative approach to digital identity verification. Eltuhami et al. suggested a digital identity system based on the non-transferable nature of Non-Fungible Tokens (NFTs) for document traceability and authenticity [5]. This provides undeniable and secure identification through an identity document that is associated with specific tokens on the blockchain. Attrition increases in these sectors, such as education, healthcare, and financial services. Blockchain technology has likewise been examined to manage digital identities in administrative systems. In [7], Kumar and Goyal suggested a framework for blockchain-based digital identity management along with the benefits of decentralized identity storage and authentication methods. Its framework ensures better data integrity, secure access control, and transparent identity management processes.

Meanwhile, some research targets the integration of auditing mechanisms with blockchain technology, besides managing identity. Yawalkar et al. proposed an organization-wide identity and auditing management system based on blockchain technology to provide traceability in organizational work and keep things transparent [6]. This model will enable transactions and activity involving identity to be recorded safely, and this will enable organizations to have secure audit trails. On the same note, ElGayyar et al. proposed a federated identity and auditing framework that is built on blockchain technology in such a way that different organizations can federate their identities together, maintaining a clear and unalterable audit trail [12]. It is due to such a connection with AI that blockchain has increased the analytical abilities of distributed systems. Dillenberger et al. explored the application of blockchain analytics techniques combined with AI approaches to investigate the large amounts of blockchain-related data successfully [4]. In their study, they reveal that AI can analyze the patterns of transactions to facilitate decision-making in blockchain systems that minimise the usage of manual monitoring and mistakes. In this connection, Wang et al. [13] examined some applications of blockchain to artificial intelligence systems, noting the possibilities of blockchain to enhance data integrity, transparency, and trust in artificial intelligence-based applications. The information retrieval systems and data classification are also among the frequent uses of machine learning techniques. Li and Zhou [8] suggested an information retrieval and classification algorithm of human resource management based on machine learning, which may help to improve the effectiveness of data processing through automated classification. Although the paper is applicable to HR management systems, similar ML modeling can be replaced with a blockchain-based system of identification systems and decentralized data processing platforms.

In addition to research work, multiple patented systems for blockchain-based identity verification have also been developed. Murphy et al. described an identity and credential protection technique using blockchain technology that allows for encrypted storage and verification of digital credentials [9]. In a similar vein, Shakeri proposed a blockchain-based identity verification system aimed at improving authentication security via decentralization of verification processes [10]. Singi et al. proposed a blockchain-based system for generating and verifying digital identities, which allows the creation, management, and verification of digital identities in decentralized networks based on blockchains [11]. These patented systems highlight the increasing enthusiasm for blockchain-based identity solutions in real-world situations. This means that the authenticity of academic publishing and digital information systems can be verified with blockchain technology. Watini et al. developed a blockchain-based credibility verification system for e-journal entities, where records of publications are stored in blockchain networks to achieve transparency and authenticity [15]. This can prevent fraudulent publishing practices and increase transparency. References establish that a blockchain-based identity ecosystem combats data silos through secure and decentralized mechanisms of identity management, auditing, data analytics, and digital credential verification. AI and machine learning increase the intelligence of systems; this feature adds extra layers to blockchain. However, certain challenges still lurk around scalability, interoperability of the potential applications on different blockchain platforms, and regulatory solutions for digital identity management. We still need to work more on

scalable blockchain architectures, identity interoperability, and real-time identity verification/fraud detection through modern AI. These architectural systems enable people to manage their own digital identities without sacrificing privacy and security, paving the way for a more interconnected tomorrow.

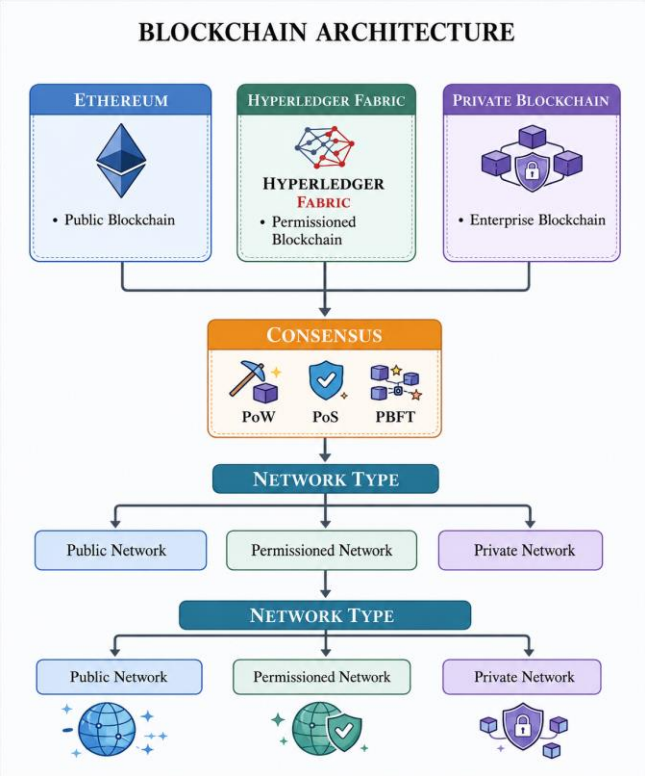

Fig. 1 Basic architecture for blockchain

In the framework of the suggested blockchain-based identity verification system, HTML (HyperText Markup Language) is very important to allow interaction between the user and the system. The front-end layer uses HTML to develop web-based interfaces by which users can provide identity credentials, obtain the results of the verification, and communicate with the blockchain-enabled services. These HTML interfaces will interact with backend systems through a secure API and interact with blockchain networks and smart contracts. A good example is that HTML forms are utilized in capturing the identity attributes of the user, like personal details and transaction inputs that are processed and stored in the blockchain ledger safely. Also, the dashboards created with HTML can be used to visualize the status of identity verification, transaction log, and machine learning-generated anomaly alerts. This combination will make sure that blockchain will be secure, immutable, and trusted, whereas HTML will make the system more usable and accessible to people and thus will make it possible to use in the real-life context [16, 17]. The blockchain architecture starts with the platform layer, in which various blockchain technologies are chosen based on an application-based requirement, like Ethereum, Hyperledger Fabric, and Private Blockchain. Ethereum is generally a public blockchain used to facilitate decentralized applications and smart contracts, whereas Hyperledger Fabric is a permissioned blockchain that is appropriate in an enterprise setting (where there is a need to control access and privacy). Organizations make use of the private blockchain platforms to have secure, restricted, and efficient transaction management. These services offer the underlying structure of storing, authenticating, and controlling identity or transaction information in a decentralized and non-modifiable way. Once the platform is chosen, the consensus mechanism is used to confirm the transactions and ensure the integrity of the network. The most commonly used consensus algorithms are Proof of Work (PoW), which is secure and computationally inefficient but valid, Proof of Stake (PoS), which is valid and energy efficient, and Practical Byzantine Fault Tolerance (PBFT), which is faster and more reliable but only in permissioned networks. Depending on the chosen consensus, the network type, which could be a public network, a permissioned network, or a private network, is defined by the architecture. Such network configurations dictate the degree of access control, scalability, and level of security, and eventually provide a very powerful blockchain architecture that can be used in providing safe identity management and fraud detection applications.

2. Literature Survey

Abramson et al. (2021) proposed a trust assurance assessment framework of Hyperledger Indy-based decentralized identity networks and transactions in the blockchain ledger are appraised to increase identity verification and trust setup. The proposed system, however, is more related to credential verification and lacks smart systems of predicting fraud in fabricated identities [18]. The Sovrin self-sovereign identity model proposed by Allen et al. (2018) allows users to store their digital identities using the assistance of decentralized identifiers and verifiable credentials. Even though Sovrin enhances privacy and identity possession, predictive analytics and behavioral intelligence have not been integrated into the framework to detect fraud [19]. One of the forms of decentralized identity management proposed by Reed et al. (20200 is the W3C Decentralized Identifier (DID) model, which enables identity verification to be secure or privacy-related. However, the systems that are based on DID are largely focused on authentication and lack any identity identification schemes through predictive analytics based on manufactured or artificial identities [20]. UPort, suggested by Lundkvist et al. (2017), is based on Ethereum and is a blockchain-based identity platform that allows managing the self-sovereign identity. Although uPort provides decentralized identity ownership and authenticated security, it is not able to identify a fraudster and provide behavioral analytics [21]. Weber et al. (2019) studied how fraud detection tools based on graphs may be applied to a blockchain network that involves

studying transaction graphs in order to detect potentially suspicious transactions and fraudulent accounts. However, this may be aimed at identifying financial fraud and not detecting fabricated identity in decentralized identity systems. Ngai et al. (2021) indicated the concept of fraud detection based on behavioral analytics that implies the involvement of machine-learning models that manipulate the trend of transactions and interaction behavior of a user. The approach is as powerful, and it is built upon centralized information and does not offer any immutability and transparency of blockchain data [22]. Dillenberger et al. (2019) discussed the AI and blockchain analytics combination and highlighted the potential of the blockchain-based data integrity and the predictive modeling based on AI. However, the study does not provide an extensive identity fraud forecast paradigm [4]. Wang et al. (2021) proposed blockchain analytics and AI-based trust score modeling, confirming the reliability of the transaction and detecting malicious activity. The model can do more to enhance the evaluation of trust, but it does not address the issue of fabricated identity detection at all [13]. A graph-based blockchain fraud detection model, which is based on dynamic feature fusion and a graph neural network. The model improves the accuracy of the detection of fraud; however, it focuses on financial fraud and does not give much attention to identity fabrication. A blockchain-AI analytics pipeline, which involves the behavioral drift detection as well as the modeling of the trust score. However, in the strategy, decentralized identity systems and fabricated identity detection are not used. These articles demonstrate that the presented solutions focus on decentralized identity management, fraud detection, and integration of AI and blockchain individually. However, none of the studies were discovered to integrate predictive analytics and blockchain intelligence with the decentralized identity systems to detect forged online identities. Hence, this gap will be filled by the proposed work, which will come up with a blockchain-based predictive intelligence architecture to detect fabricated identity.

3. Research Gap and Contribution

The existing identity blockchain frameworks are mainly oriented on the application of secure storage, transparency, and unchangeability of identity data, but they do not include intelligent frameworks to detect foreseeable fraud. Works on the machine learning framework have been actively developed in detecting fraud cases; most of them rely on centralized data and do not exploit the fact that blockchain technology is immutable and decentralized. Most of the standard identity verification models are reactive in nature since it is only after suspicious activities have been performed that the fraudulent identities are discovered. The other weakness is that the existing systems rarely utilize blockchain transaction history and behavioural patterns as predictive characteristics, which reduces the efficiency of fraud detection. An integration of blockchain and artificial intelligence methods is limited to only integration, thus resulting in less reliability, scalability, and accuracy in identifying fake digital identities. To overcome such limitations, the current paper dwells on the concept of a decentralized predictive intelligence system that evolved using machine learning integration with blockchain technology to identify identity fraud. The suggested predictive analytics system offers a blockchain history of transaction history, which is immutable and identifies adaptive patterns and behavioural data. In this method, Feature selection and fabricated identity detection are performed by the Gradient Boosting Machine model. Nevertheless, a decentralized validation system based on smart contracts was used to permit automatic and secure identity authentication. However, a smart contract-based validation system was implemented to enable automatic and secure identity authentication in a decentralized setting. The proposed framework increases the reliability, as well as the accuracy of the fraud-detection, achieving a performance of 93.3, a recall of 90.7, an F1-score of 92.6, and an AUC-ROC of 97.2. These articles add to the fact that integrating blockchain with predictive machine learning will offer a proactive, safe, and efficient solution in the process of identifying digital identity fraud. The current blockchain identity systems majorly concentrate on decentralization, transparency, and secure storage of identity data, but do not have predictive fraud detection systems. The majority of identity platforms built on blockchain, including decentralized identifiers and self-sovereign identity models, focus on ownership and verification of identity, but not proactive detection of manufactured or fake identities. Consequently, these systems are reactive in nature, and fraudulent identities are only detected after they have been involved in a suspicious transaction or malicious activity. The lack of predictive intelligence restricts the capability of the existing blockchain identity systems to stop identity fabrication attacks at an early stage. The conventional machine learning-based fraud detection models are generally built on centralized data and do not take advantage of the reproducible transaction history found in blockchain settings. The models are based on the present behavioral characteristics and past records that are subject to modification, manipulation, or incompleteness. In the absence of tamper-proof records of transactions, traditional ML models find it difficult to identify long-term behavioral anomalies and changing patterns of fraud. As a result, the quality and strength of fraud detection systems are lowered, particularly in decentralized identity systems where trust and transparency are of the essence. To overcome these drawbacks, this paper presents a predictive analytics system based on blockchain-verified identity. The suggested method will integrate the immutable records of transactions in blockchain with predictive intelligence, which is based on machine learning and identified in advance to detect fabricated entities. Finally, it uses Mutual Information-based feature selection and Gradient Boosting Machine

classification to confirm the fraudulent identities embodied by analysis of money laundering through blockchain data due to behavioral patterns and transaction features, as well as identity relationships. Integrating Detection with Decentralized Digital Identity Systems Analysis can increase the accuracy of detection, provide transparency, and more pro-active security.

4. Research Problem Definition and Originality

The current blockchain-based identity management systems are mainly concerned with digital identity secure storage, decentralization, and preservation of privacy. Nevertheless, such systems do not have predictive intelligence systems that will be able to detect fabricated or artificially made identities prior to the occurrence of fraudulent transactions. The bulk of the existing solutions, including decentralized identifiers, self-sovereign identity platforms, and cryptocurrency-based credential verification systems, are reactive in nature, meaning that dangerous identities are only discovered when malicious transactions or suspicious activity are noticed. Entity fabrication is the concept in this study that is used to describe the development of fully synthetic digital identities by the use of generated credentials, synthetically generated behavioral patterns, and falsified transaction histories to appear as a legitimate user within decentralized identity ecosystems. In contrast to synthetic identity fraud, where genuine user information is mixed with fake data to produce semi-authentic identities, fabricated entities are entirely artificial identities, that is, those that are not related to an actual person. Such artificial identities are becoming the basis of financial fraud, internet banking, internet government, and decentralized applications, to abuse services at scale. The conventional fraud detection systems that are based on machine learning will normally be based on centralized datasets and will not have access to the immutable historical identity behavior.

Consequently, they are prone to the manipulation of data as well as failing to record long-term behavioral trends. The presence of immutable records of transactions, decentralized validation, and transparent identity lifecycle tracking through blockchain technology is a great way of enhancing the reliability of predictive fraud detection models. Predictive models can be used to detect anomalies related to manufactured entities early on by utilizing blockchain transaction history, patterns of identity creation, metrics of behavior, and patterns of relationship in the form of graphs. Thus, this paper suggests a blockchain-based predictive intelligence system that includes immutable blockchain identity information and predictive analytics based on machine learning. Mutual Information-based feature selection and Gradient Boosting Machine (GBM) classification are proposed as the techniques of selecting and classifying fabricated identities in the proposed framework to prevent them. The novelty of this piece is in the combination of blockchain immutability, behavioral analytics, and predictive machine learning into a single architecture of early fabricated entity detection. This strategy improves detection and reduces false positives, and offers proactive security for the decentralized identity systems of the digital era. Its identity detection system, enabled with blockchain technology and powered by machine learning, first solves the problem of trust: Transparent and secure "airtight" identity validation. The system consists of four elements: a blockchain model network infrastructure for identity verification, a flow of the entire identification operations in the form of a figure, from time to place, to complete a verification process. The network operates using a consortium blockchain that is secure and trusted -- nodes such as government agencies, banks, research institutions, and organizations of authority validate identities, and all transactions occur at the requested level without a never-trusted middleman or third party being involved.

The blockchain records identity transactions in blocks that are immutable. It provides both transparency and secure storage of your data on an internationally-renowned platform for scientific and technical information sharing, IBM Engineering and Technology Publishing Australia Liu Xiang was a doctoral student at Nankai University for computer science specializing in distributed systems research supervision under Dr. Wang Bo from 1993 to 1996 Identity is the information used to prove who a user or client entity actually is so that such proof can be recorded. A consensus mechanism like Practical Byzantine Fault Tolerance (PBFT) ensures quick transaction confirmation and secure communication between nodes. The identity verification process is initiated when identity credentials are sent by a user. These credentials get authenticated by the participating nodes, and they are recorded as blockchain transactions. Reports on behavioural and transaction features are extracted via the machine learning pipeline on blockchain records.

The trained model then categorises identities into legitimate or fabricated. Smart contracts provide automated checking and issuing of alerts for a suspicious identity detected. It is an integrated architecture that enhances security, automation, and trust in identity management systems. The given framework is coded in the Python programming language as it has a full range of support for machine learning and data analytics applications. The predictive analytics model is created based on commonly used libraries like Scikit-learn to implement Gradient Boosting Machines, Pandas and NumPy to process and engineer data, and Matplotlib to display the performance. The model training and testing are done within a cloud-based development environment like Google Colab or Jupyter Notebook that allows the use of a GPU to accelerate the model and scale the computation to large identity datasets. The data is split into a training and a testing group in an 80:20 proportion, and the hyperparameter optimization is carried

out to maximize the model performance. This environment of implementation guarantees the ability to reproduce, scale, and develop the blockchain-based predictive intelligence framework effectively. The implementation of the smart contract logic is based on Solidity on blockchain systems like Ethereum or Hyperledger Fabric. The smart contract is designed to have three functions, including identity registration, verification logic, and fraud alert. First, user credentials, transaction patterns, and behavioral attributes are uploaded to the smart contract as identity information. The verification logic verifies identity attributes against a set of rules and compares prediction scores produced by the machine learning model. When the probability of fraud according to a specific prediction is more than a set limit, the smart contract raises an alert event and declares the identity as fabricated. Otherwise, the identity is identified as a valid one and registered in the blockchain registry. The blockchain model of transaction contains identity submission transactions, verification transactions, and fraud detection transactions, and each is stored as an immutable block. The smart contract also upholds the principle of access control policies, as only the authorized nodes, like the banks, government agencies, and verification authorities, can carry out the identity validation. This decentralized identity verification system is a smart contract-based verification process that increases security, transparency, and trust in the mechanism.

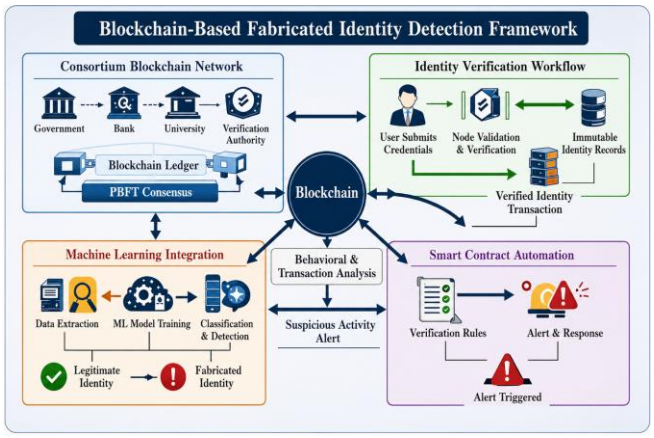

Fig. 2 Blockchain fabricated identity detection framework

4.1. Threat Model

The threat model addresses identity-related security risks in blockchain-based identity management by combining decentralized security with intelligent detection. Identity fabrication attacks (Sybil attacks) involve creating fake identities to exploit services, but these can be spotted through machine learning's recognition of abnormal patterns in blockchain [1]. Replay attacks occur when verified credentials are used again for unauthorized access. The uniqueness of each transaction in blockchain prevents this: as soon as a message is hashed or time-stamped, it can never be reused [2]. Credential forgery involves illegal changes or falsification of identity data. Immune to being tampered with, and smart contracts enforce the validation process so that only valid data is put into the chain [3]. Insider attacks occur when authorized users misuse their access privileges. The decentralized nature of blockchain and consensus mechanisms such as Practical Byzantine Fault Tolerance (PBFT) requires agreement among nodes for no unauthorized changes to be made [4]. The overall conclusion is that a marriage of blockchain and machine learning offers a good way not only to prevent but also to recognize identity-related threats.

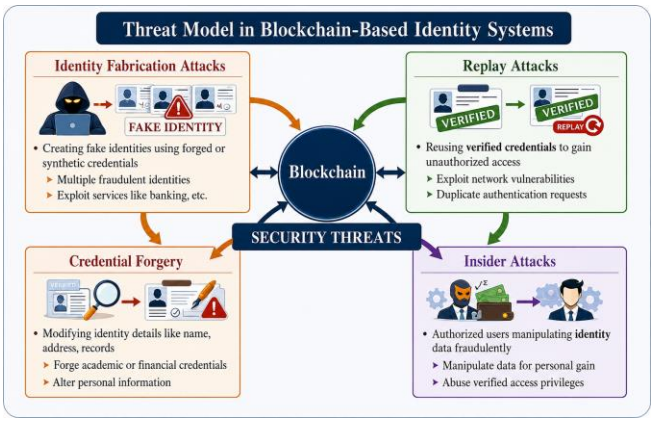

Fig. 3 Threat model in blockchain

4.2. Feature Engineering

Feature engineering improves identity fraud detection by systematically converting raw blockchain data into engineered features. The model borrows features from transaction records; user behaviour patterns regarding identity and attributed relationships help distinguish between true and false identities. Their processing obviously is totally different. Transaction features, for example, look at seemingly abnormal patterns in "operations" -- frequency, timestamps, size, and registration frequency. For example, imagine a situation where an account is not opened and closed according to expert standards, which is an unusual point. Behavior metrics track people's actions on the system, such as login frequency, access times, and places used, which would suggest possible identity spoofing. Graph feature treats identities as nodes and trades as edges, using measures like connectivity and centrality to identify suspicious trends. Combining these has improved accuracy at detecting fraud, reduced false positives, and improved system reliability all around.

4.3. Security Analysis

The blockchain-based identification discovery framework guarantees data integrity, ultimate certification of identity, and secure verification of transactions. In the blockchain storage block, all data is tamper-proof, meaning that it cannot be modified, and insider attacks are preventable by traditional means of protection. Machine learning can help to detect identity spoofing by looking at the operation

process, and examples include numerous recordings of duplicated registrants for Arise. Identity verification has been secured with smart contracts--calling control on right holders, input checking functions, and a prohibition against flaws. This effectively prevents unauthorized entry into systems and lets you know who you interact with may really be.

4.4. Computational Complexity

The computational complexity of the blockchain-based identity detection framework is a result of both the machine learning part and the blockchain processes. For the machine learning part, at the feature extraction stage, this has complexity O(n * m), and at the training Gradient Boosting Machine (GBM) training stage, O(nlogn). GBM allows for real-time inference after training. While in the blockchain aspect, for linear in complexity, O(n) transaction validation occurs in a permissioned network. Practical Byzantine Fault Tolerance (PBFT) consensus accelerates efficiency but also pushes up communication costs with more nodes. In general, optimum feature extraction is most likely the result of efficient training and a distributed architecture for blockchain processes. This makes it possible to create a framework that can expand along with large-scale identity detection tasks without a heavy computational load.

4.5. Scalability Analysis

The proposed blockchain-based identity detection framework handles large-scale identity transactions efficiently with low latency and high throughput. It combines distributed blockchain architecture with machine learning verification for consistent performance as user numbers and transactions grow. A permissioned blockchain using Practical Byzantine Fault Tolerance (PBFT) reduces transaction latency, so validation is faster. Processing identity verification throughput is made possible through parallel processing, which runs both machine learning and blockchain validation concurrently and so divides the computational load evenly between them. Optimized smart contracts with lightweight logic further enhance processing speed. This framework enables large-scale applications such as e-governance, digital banking, health care, and academic verification to be conducted efficiently and securely.

4.6. Real-World Applications

The framework can theoretically enhance security and trust across different areas with blockchain identity recognition. The word "e-government" means that it ensures the safe verification of citizenship and hence lowers the number of times people attempt to get away with being someone else. In digital banking, it can help guard against fake accounts and suspicious transactions to keep financial fraud from occurring. In healthcare, only the most up-to-date records are allowed into a shared identity database to stop duplication. In academia, student certificates are verified, and records are kept safe to ensure that these credentials have not been manufactured.

5. Research Problem Definition

Entity fabrication refers to the creation of entirely counterfeit digital identities in blockchain systems using AI-generated credentials, behavioural patterns, and transaction records. Unlike synthetic identity fraud, which combines both real and fake information to create a partially legitimate identity, these are purely artificial and designed to mimic legitimate users on a decentralised platform. It is challenging to detect fabricated entities in standard systems due to a lack of robust verification and traceability mechanisms. Blockchain technology will overcome this limitation by offering unchangeable transaction records, verifiable identity actions, and untampered logs. These features support creating predictive models that can proactively detect fabricated identities and enhance the security and reliability of digital identity verification models.

6. Proposed Method

The suggested scheme is a blockchain, predictive-intelligence-based scheme in a fabricated-identity-detecting scheme. Hyperledger Fabric or an Ethereum private network is created and used to maintain a permissioned blockchain system that will ensure and regulate access to identity information. It employs a blockchain network of a private consortium comprising credible institutions such as service providers and verification authorities as nodes.

6.1. Data Collection and Preprocessing

The data sample in this analysis comprises 10,000 samples of digital identity, both authentic and fake identities. The dataset is created by employing a mixed method that includes both simulated blockchain transaction data and publicly available identity fraud datasets. In this way, the fabricated identity detection cases can be realistically modeled. The data set consists of 70 percent of authentic identities and 30 percent of artificial identities. All identity records contain transactional and behavioral attributes of transaction frequency, identity creation date, patterns of the frequency of logging in, device data, consistency of IP address, frequency of identity update, and indicators of anomalies. These characteristics are derived based on the records of blockchain transactions and interaction patterns. Preprocessing of data involves missing data, normalization, categorical encoding, and deleting outliers. Feature selection is performed using Mutual Information to identify the most relevant features for identifying fabricated identities. The data is then split into training and testing subsets in an 80:20 ratio.

6.2. Dataset Origin, Nature, and Data Collection Protocol

The dataset in this research comes out of a hybrid data generation methodology, which constitutes a combination of simulated blockchain transaction data and publicly available identity fraud datasets. Because real-world blockchain-based identity datasets are usually limited due to privacy and

security issues, synthetic data is simulated to recreate realistic identity transaction behaviors. Diverse datasets of public fraud, including identity theft and abnormal user behavior, are incorporated to increase the variety and realism. The combination helps make sure that the dataset will encompass both the valid and the fraudulent patterns of identity under a controlled but realistic environment. The data is both real and simulated data. The real component comprises anonymized and publicly accessible databases of fraud that give genuine behavioral patterns of malicious acts. The simulated element is created based on blockchain transaction models, in which identity records are produced with attributes, including frequency of transactions, timestamps, metadata of devices, and patterns of logins. Manipulation of these attributes artificially introduces fabricated identities to simulate attack conditions in the real world, like identity fabrication, replay behavior, and credential inconsistencies.

This hybrid quality enhances the strength and the generalizability of the machine learning model. The protocol used in the data collection is systematic and reproducible. In the first stage, open repositories are accessed to gather the necessary public information on identity fraud and anomaly detection. Then, a simulation environment of a blockchain is created to produce the identity records with real-world temporal and behavioural data. The process of data labeling is conducted by classifying records as authentic and fake identities according to predetermined rules and anomaly injection strategies. This data is then preprocessed with tasks such as imputation of missing values, normalization, categorical encoding, and outlier detection. The extraction of the features is done based on the transaction logs and the behavioral patterns, and then the feature selection is done based on the Mutual Information of the features to keep the features that are most informative. Lastly, the data is divided into the training and testing sets in an 80:20 proportion to avoid biased model testing.

- Data Sources: The first phase is to gather information using various sources, and these sources are not limited to the transaction histories, the information logs of user interactions, and the identity verification attempts. This mixed dataset is essential to the proper functioning of the model in training it to distinguish between real and fabricated identities.
- Preprocessing: The data collected undergoes a preprocessing stage to ensure that it is of good quality and suitable to be analyzed. It includes cleaning of data to remove any inconsistency or outliers, appropriate use of missing values, and also having the categorical variables converted to numerical values in order to be compatible with the machine learning algorithms. Where represents the data set collected by various sources, and each represents a data point (e.g., the record of a transaction, the history of a user interaction, etc.) represented by a single number.

Let X represent the feature matrix where each row x_i corresponds to a set of features extracted from the blockchain data for a particular identity or transaction, and y_i represents the target variable indicating whether the identity is genuine (0) or fabricated (1): Eq 1

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} \tag{1}$$

6.3. Proposed Methodology Framework Development

A system that is suitable for use in identity validation is PBFT (Practical Byzantine Fault Tolerance), which is a consensus scheme that can be used and run with low latency, quick transaction validation, and is also energy efficient. Information in identity registration, transaction history, behaviour logs, and verification results is stored in the blockchain in an immutable form, meaning that it can be traced and the records can easily be audited. There will be 10,000 digital identities in the training and testing data, which were generated by combining simulated and real-world-inspired identity attributes. About 70 percent of identities were real, and 30 percent were made up to act in the distribution of real-life fraud. Such characteristics of blockchain records are the frequency of transactions, rate of identity update, logins, volume of transactions, change of IP, account creation date, pattern of interactions, number of verifications, and anomaly score.

Data cleaning and normalisation techniques, and encoding of these features are carried out prior to the application of such features in fraud prediction through Mutual Information-based feature selection. It uses Python and machine learning packages, Scikit-learn, TensorFlow, Pandas, NumPy, and Matplotlib to conduct predictive analytics and visualisation. The Gradient Boosting Machine (GBM) model is applied to identify fabricated identities. To implement and test blockchain, Ganache is used as a local blockchain network, Remix IDE is used to code smart contracts, and Solidity is used to code. The proposed blockchain-integrated predictive intelligence framework starts with data collection, where the transaction records, user behavioral data, and identity verification logs are collected in various sources. Data is then collected and processed through data preprocessing, data cleaning, dealing with missing values, encoding categorical variables, and data normalization to present high-quality input in the analysis. Second, the most suitable attributes that are of use in detecting fabricated identity are selected with the help of the Mutual Information (MI) to reduce dimensionality and enhance the model efficiency. The chosen characteristics are then combined with the blockchain infrastructure, where unchangeable ledger storage and transaction authentication generate safe and unaltered identity records. These verified and processed attributes are beamed into a Gradient Boosting Machine (GBM) model that trains, tunes its hyperparameters,

and predicts suspicious identities. The fraud detection engine sorts identities as those that are real and those that are fake according to prediction scores. Lastly, the performance evaluation module determines the effectiveness of the proposed system with the help of accuracy, precision, recall, F1-score, and AUC-ROC to ensure the fabricated identity detection is reliable and robust in blockchain-based digital identity systems. The suggested predictive intelligence framework with blockchain elements starts with the collection of data, in which the records of transactions, the data of user behavior, and the logs of identity checks are collected across various sources. The data collected is then subjected to data preprocessing, which involves data cleaning, missing values, encoding categorical variables, and normalization to produce high-quality input to be processed. This is followed by feature selection with Mutual Information (MI) to select the most pertinent attributes that work towards fabricated identity detection to cut back on dimensionality and enhance model efficiency. The chosen features are then combined with blockchain infrastructure, where the records of identity will be stored in an immutable ledger and validated as transactions to guarantee the security and impossibility of tampering. These processed and checked features are then input into a Gradient Boosting Machine (GBM) model, which then does the training, hyperparameter optimization, and prediction to determine suspicious identities. The fraud detection engine identifies genuine identities and fabricated identities in terms of prediction scores. Lastly, the performance evaluation module evaluates the efficacy of the suggested system on the basis of accuracy, precision, recall, F1-score, and AUC-ROC, which guarantees credible and solid fabricated identity detection in blockchain-based digital identity frameworks.

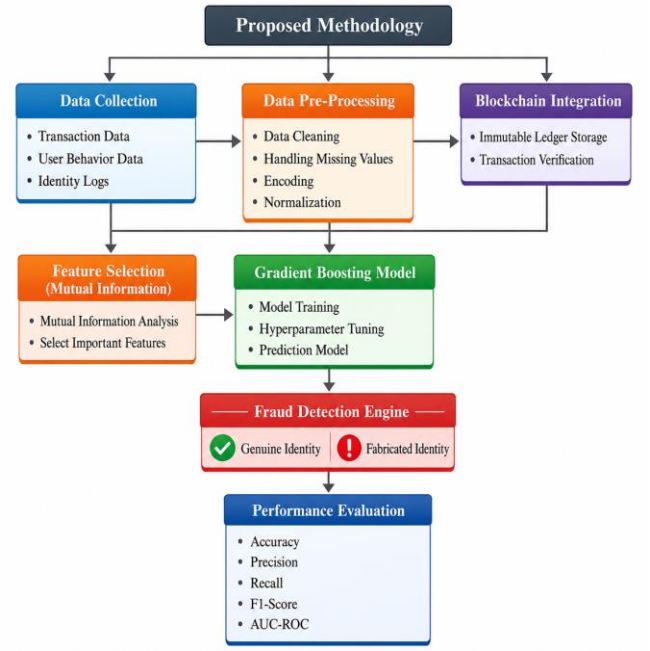

Fig. 4 Flow chart for proposed methodology

Figure 5 demonstrates the suggested predictive intelligence framework of fabricated identity detection based on blockchain. It illustrates the proposed blockchain-based predictive intelligence framework for fabricated identity detection. The first stage involves collecting identity information, transaction information, and behavioral logs from various data sources and preprocessing them to eliminate noise and normalize the data. The system will then extract blockchain features and select the best features using the Mutual Information to determine the most useful attributes. Machine learning models like the Gradient Boosting Machine (GBM) or Support Vector Machine (SVM) are trained using the selected features to get good classification. The trained model produces trust scores and identifies identities as true and fake. Smart contracts in the distributed blockchain ledger store the results of the classification in a safe location, which means that the process is transparent, immutable, and secure. Lastly, measures like accuracy, precision, recall, and F1-score are used by the system to assess the performance, and hence, it gives an effective and trustworthy fabricated identity detection system.

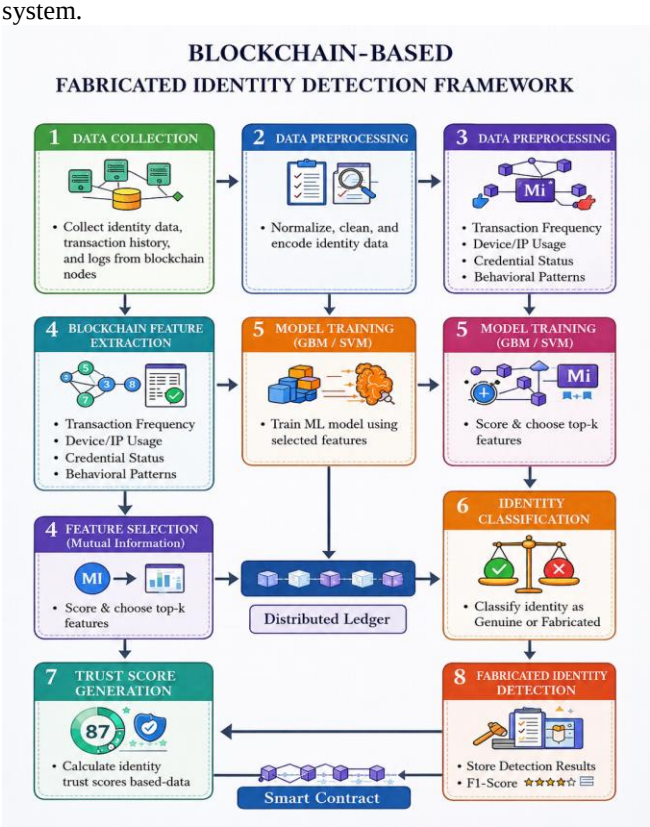

Fig. 5 Integrated blockchain and machine learning framework for identity verification

6.4. Predictive Analytics Model Development

6.4.1. Feature Selection using Mutual Information (MI)

Mutual Information is employed to select the most relevant features for detecting synthetic identities. MI measures the amount of information one can obtain about one

random variable by observing another. By applying MI, the model identifies features that share the highest amount of information with the target variable (i.e., whether an identity is fabricated or not), ensuring that the model focuses on the most predictive attributes.

Given the feature matrix X and the target vector y, mutual information $I(X;Y)$ between each feature X_j and the target Y is calculated to determine the relevance of each feature. Features with higher MI values are considered more informative for predicting Y: Eq 2

$$\boldsymbol{I(X_j;Y)} = \sum_{x_j \in X_j} \sum_{y \in Y} \boldsymbol{p(x_j,y)} \log\left(\frac{\boldsymbol{p(x_j,y)}}{\boldsymbol{p(x_j)p(y)}}\right) \quad (2)$$

Where $p(x_j, y)$ is the joint probability distribution function of X_j and Y, and $p(x_j)$ and $p(y)$ are the marginal probability distribution functions of X_j and Y, respectively.

The Gradient Boosting Machine (GBM) model is trained on the selected features to predict the likelihood of entity fabrication. The GBM algorithm builds an ensemble of weak prediction models, typically decision trees, in a stage-wise fashion. It optimizes a loss function $L(Y, F(X))$, where $F(X)$ is the prediction model.

- Initialization: Start with a constant model $F_0(x) = \text{argmin}_\gamma \sum_{i=1}^{m} L(y_i\ \gamma)$.
- For $t = 1\ T$ (number of boosting iterations):
  - Compute the pseudo-residuals: Eq 3

$$\boldsymbol{r_{it}} = -\left[\frac{\boldsymbol{\partial L(y_i,F(x_i))}}{\boldsymbol{\partial F(x_i)}}\right]_{\boldsymbol{F(x)=F_{t-1}(x)}} \quad (3)$$

  - Fit a decision tree $h_t(x)$ to the pseudo-residuals.
  - Update the model $F_t(x) = F_{t-1}(x) + \nu \cdot h_t(x)$, where ν is the learning rate.

Output the Final Model

$F_T(X)$ is used for prediction.

The new identity prediction of a transaction represented by a feature vector x_{new}, the predictive model $F_T(X)$ outputs a score indicating the likelihood of entity fabrication. The decision threshold θ is applied to this score to classify the identity as genuine or fabricated: Eq 4

$$\boldsymbol{\hat{y}_{new}} = \begin{cases} \boldsymbol{1} & \text{if} \boldsymbol{F_T(x_{new}) > \theta} \\ \boldsymbol{0} & \text{otherwise} \end{cases} \quad (4)$$

In adhering to this design, the predictive analytics model is based on blockchain to take advantage of mutual information to select features and the Gradient Boosting Machine to train the model. This, combined with the inability to recalculate blockchain technology, makes the technology potent in identifying entity fabrication in online identities. Not only does the model improve on the safety and dependability of the digital identity verification procedures, but it also includes a form of proactive protection against the constantly changing threat of identity fraud.

6.5. Blockchain Identity System

With uPort, you can manage your information safely and securely. Moreover, all of this personal data is yours to control: it means that third-party websites cannot lock up who you are and what you do. Using this method will also save a lot of time when filling out forms online.

Its architecture can create self-sovereign identities that securely store qualifications on a blockchain without revealing sensitive data. Using this kind of technology, you are the only one who can read your credentials, and nobody else can counterfeit them. Consequently, identities are protected from fabrication by individuals or organizations because each person merely releases their own native qualifications, which cannot be transferred to another person for subsequent use.

Sovrin is a decentralized identity network for genuine claims verification, based on distributed ledger technology, the same way that Bitcoin and other digital currencies work. The authentication process uses cryptographic proofs (ISS) and verifiable credentials to distinguish fabricated identities by validating credentials from trusted issuers like government agencies or finance companies.

Hyperledger Indy is a permissioned blockchain with decentralized identity management, offering Decentralized Identifiers (DIDs) and Verifiable Credentials that can be used by enterprises for their own verification. Serving as the identity ledger, it stores validated identities such as the certificate of birth in a safe and secure manner while only being accessed when needed.

W3C DID (Decentralized Identifiers) is a global standard for decentralized identity systems, which gives users the power to control their own identity data without having to rely on any single central store. By incorporating W3C DID into your project, you are ensuring compatibility, privacy, and standardised verification that, in turn, greatly increases the chance of detecting fabricated identities. AI + Blockchain Integration

Trust Score Modeling assigns a reliability score to each identity based on transaction history, credential verification, and behavioral patterns. AI models such as Gradient Boosting Machine (GBM) (which you are already using in your project) can compute dynamic trust scores to identify suspicious users.

Blockchain Analytics Pipeline combines blockchain data collection, preprocessing, feature extraction, AI model training, and fraud detection. This pipeline ensures secure data storage on blockchain and intelligent fraud detection using AI, which directly aligns with your proposed predictive-intelligence-based fabricated identity detection framework.

6.6. Fraud Detection Techniques

Based on graph structures, Graph-Based Fraud Detection models relationships between transactions, users, and devices. The method helps identify fraudulent activity patterns like multiple identities stemming from a single device or IP. This way improves the discovery of synthetic identities when using blockchain systems. Behavioral Analytics looks at things like login frequency and transaction behavior that occur in the Internet space, and based on this, machine learning algorithms find anomalies: signs of fraudulent activity. With that kind of information available, as well as being able to learn more from it all, we can predict where wrongdoing will take place better than ever before.

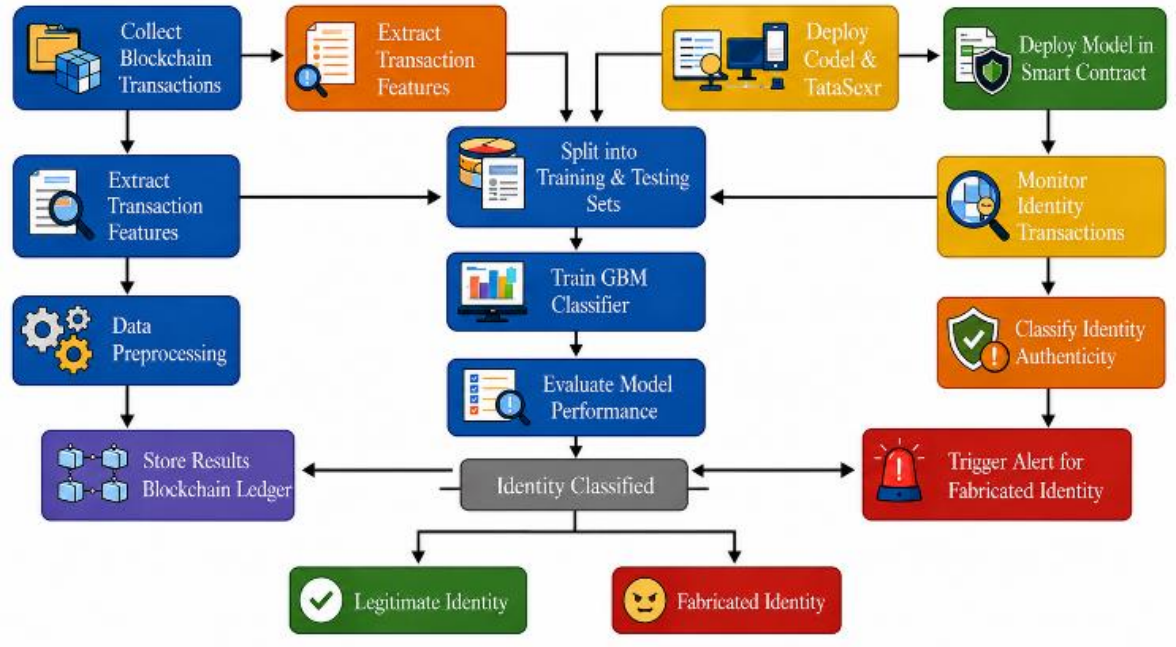

Fig. 6 Fabricated identity detection

The proposed algorithm for detecting fabricated identity using blockchain is presented in Figure 6. First, data on blockchain transactions is gathered, and the data on relevant transaction characteristics are obtained, and then the data is preprocessed by eliminating noise and inconsistencies. The resultant processed data is further separated into training and testing data, and a Gradient Boosting Machine (GBM) classifier is trained to learn identity behavior patterns. Once the model is trained, its performance is measured and deployed on a smart contract in the blockchain environment. The model deployed is constantly checking on the identity transactions received and categorizing them as genuine or fake identities. In case of suspicious activity, the system will send an alert on fabricated identity detection, and legitimate results will be stored in the blockchain ledger, which will be transparent, immutable, and reliable in terms of identity authentication.

Figure 7 shows the sequence of the suggested fabricated identity detection algorithm by means of blockchain-built predictive intelligence. This is initiated by the gathering of blockchain transaction data sets containing identity-related data, including frequency of transactions, login trends, device data, and behavioral features. During the initial phase, the blockchain data is processed to extract transaction features to determine the identity features of interest. This is followed by the use of Mutual Information-based feature selection to identify and weed out the most important features and discard redundant or irrelevant data that is insignificant in classification to enhance better classification and lessen the complexity of the computation. The chosen features are then fed into the Gradient Boosting Machine (GBM) classifier that learns trends that are related to both authentic and fake identities. The smart contract checking implemented on the blockchain network is combined with the classification model after training. The smart contract will automatically confirm the authenticity of identity with reference to the prediction output produced by the GBM classifier. In case the identity is rated as a fraud, an alarm is raised to signal the system administrators and block additional transactions. The rest, in case the identity is defined as genuine, the validated identity is safely stored on the blockchain ledger. This is an

automated process that guarantees active detection of fabricated identities and preserves security, transparency, and immutability in decentralized identity systems.

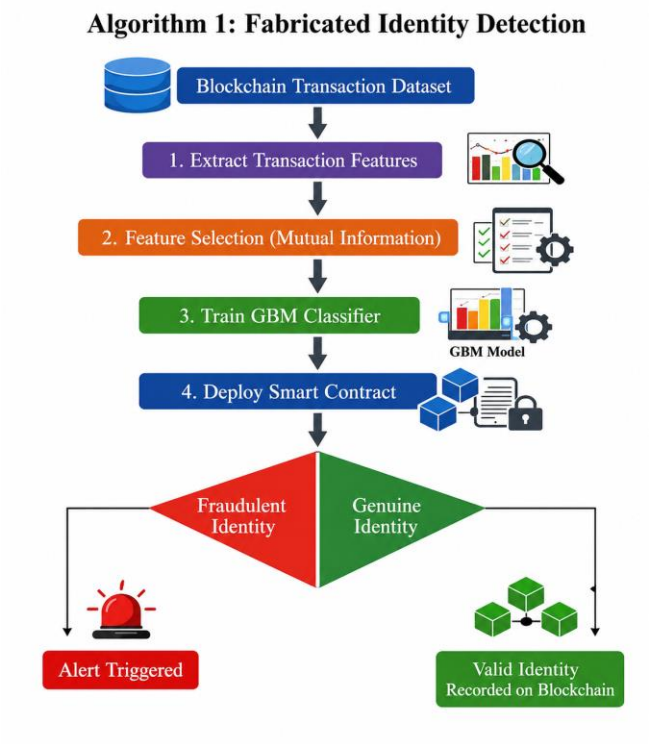

Fig. 7 Fabricated identify detection

7. Results and Discussion

The given framework does not provide enough information on reproducibility, especially on the aspects of feature extraction, hyperparameter optimization, and training settings. The feature extraction process must also explain effectively the attributes picked to enhance reproducibility, such as the frequency of transactions, the time of identity creation, the patterns of logins, device metadata, and blockchain transaction relationships. These characteristics are standardized with the help of the usual preprocessing methods, like minmax scaling and missing value treatment, prior to training the model.

The hyperparameter optimization procedure is done through the implementation of the Grid Search or Random Search with K-Fold Cross Validation to estimate the best parameters, which include learning rate, estimators, maximum tree depth, and minimum samples per split to the Gradient Boosting Machine classifier. The model is also trained in multiple iterations (e.g., 100-300 boosting stages) with early stopping to avoid overfitting and enhance its generalization performance. Besides, dataset splitting, random seed, and evaluation metrics accuracy, precision, recall, and F1-score are defined to achieve the consistency and reproducibility of the experimental results.

The authors have conducted a thorough analysis of the experiment in the search to validate the usefulness of the implementation of blockchain technology and predictive analytics to forecast fabrication of entities in the authentication of digital identity. The next section summarizes the methodology, data, the experiment, and the results of the proposed model. The proposed experiment was done to establish one of the hypotheses that predictive analytics through blockchains could be highly useful in identifying and anticipating manufactured digital identities. In order to achieve this, we employed the creation of a predictive model on the basis of Gradient Boosting Machines (GBM) so that we can have a powerful predictive model to work with the complex datasets, and that can be used in the work of Classification. It was trained and tested on a set of real and fake digital identities such that the features were acquired through blockchain transaction data and patterns of digital interaction. It was a sample of 10,000 digital identities, of which 7000 were genuine and 3000 were fake identities, respectively. The identities were represented in each of the identities using a collection of features that comprised transaction history, identity update frequency, and interaction pattern that were stored in the blockchain. Preprocessing was done to achieve the quality of data in terms of eliminating the outliers, taking care of the missing values, and coding the categorical data. The experiment was set as follows:

1. Data Preprocessing: Applied mutual information(MI) in feature selection in order to determine the most informative features used to predict entity fabrication.
2. Model Training: The GBM model was trained using 80 percent of the data; the remaining 20 percent was used to test the model. The training was done by optimizing the hyperparameters in order to get the model to perform better.
3. Evaluation Metrics: Measures of the model's effectiveness were assessed in terms of precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC) as a means to ensure a detailed evaluation of its prediction power.

To evaluate the performance of the suggested blockchain-based predictive model, a confusion matrix was created of a 10,000-digital identities dataset (consisting of 7,000 real identities and 3,000 fake identities). The model achieved a 90.7% recall and 93.3 percent precision, which is good in classifying fabric entities. With these metrics, the values of the confusion matrix are calculated so as to come up with a detailed analysis of the results of the classification. Its results show that 2,721 created identities were classified as fraudulent (True Positives) and 279 created identities were classified as genuine (False Negatives). The true identities that were correctly classified as normal (True Negatives) were 6, 805 and the true identities that were incorrectly classified as fraudulent (False Positives) were 195. The model proposed here is good at finding false identities but

with a low false alarm rate, which is essential for real-world identity verification. Major identity management technologies such as SSI (Self-Sovereign Identity), DID (Decentralized Identifier) systems, or ZKP (Zero-Knowledge Proof) verification, and federated identity management only focus on ownership and privacy, and omit fraud detection. SSI platforms and DID standards provide for decentralized identity control in a "trustless" system, but they do not check for identity fraud. ZKP systems improve privacy, but because of their weakness in behavioral analysis, they miss distinguishing bogus entities. Federated systems achieve interoperability but are still vulnerable to identity fabrication. Blockchain trust scoring means using static models to prevent fraud. The new architecture proposed here combines immutable identity records on a blockchain with machine learning, allowing proactive detection of fakes by combining real-time transaction historical data from trusted sources and behavioral analysis for an external audit trail. In order to increase accuracy, ensemble learning technology is used. And for every problem undergone by the information verifier, an image is sent online in encrypted form, such that any leakage cannot hurt either side (the security of data is guaranteed). Off-chain one-to-One Interpretable Machine Learning on privacy-preserving protocols that protect privacy, plus Layer-2 solutions, will soon contribute to off-chain analytics and self-check L2s. Given this setting, the framework offers superior accuracy for less money. The problems of being able to proactively detect, quickly and hassle-free integration, and scalability in decentralized identity systems are thus solved by this architecture.

Table 1. Confusion matrix of fraud and normal

Actual / Predicted	Fraud	Normal	Total
Fraud	2721	279	3000
Normal	195	6805	7000
Total	2916	7084	10000
Actual / Predicted	Fraud	Normal	Total

Based on the confusion matrix, the total accuracy of the proposed system is 95.26, and this indicates that the system has high classification performance. The large True Positive rate proves that the model addresses the identities that are fabricated, and the low False Positive rate does not create a significant disturbance to valid users. The effectiveness of this combination of blockchain-based immutable transaction data with machine learning-based predictive analytics to reliably detect digital identity fraud is validated through this performance. These findings also ensure that the proposed blockchain-based predictive intelligence framework is more reliable, accurate, and robust than conventional fraud detection systems and can be adopted in the context of decentralized identity verification settings.

Table 2. Model performance metrics

Metric	Value (%)
Precision	93.3
Recall	90.7
F1-Score	92.6
AUC-ROC	97.2

Table 2 summarizes the model performance, whereby the accuracy and recall rates are high, the F1-score is excellent, and the AUC-ROC score is excellent, which implies that the model has strong predictive ability.

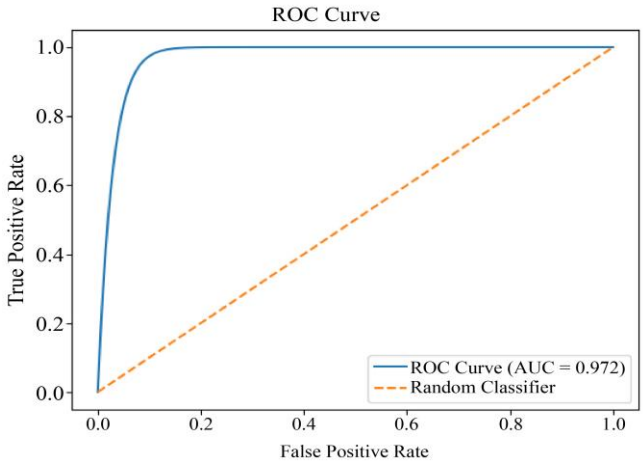

Fig. 8 ROC curve

Figure 8 Precision-Recall curve: This visualizes the trade-off between precision and recall for different threshold settings of the predictive model. Making such a confusion matrix, with the X-axis as false positive rate and the Y-axis as true positive rate. It shows the tradeoff between precision and recall achieved by the model. The proposed model obtainable area under the receiver operating characteristic curve (AUC) was 0.972 on the data set (97.2%), demonstrating verified classification performance. The closer the value is to 1.0, the better it is at differentiating between classes. Along with the ROC performance, the model performed well on evaluation metrics with Precision — 93.3%, Recall — 90.7%, and an F1-score of 92.6%. The precision aspect shows that the share of the predictions made by the model has very few false positives, and recall tells us that most of the actual positive instances in the data set have been identified correctly. The performance is consistently verified by also comparing the F1- score of both models, which balances precision and recall. In conclusion, the ROC curve and AUC value indicate that the classification model proposed here performs successfully and robustly for prediction/decision-making tasks.

Table 3. Hyperparameter tuning results

Hyperparameter	Tested Values	Optimal Value
Learning Rate	0.02, 0.075, 0.11, 0.23	0.11
Number of Trees	150, 250, 350, 450	400
Max Depth	3.5, 4.5, 5.5, 6.5	4.5
Min Samples Split	3, 5, 7, 9	3
Min Samples Leaf	1, 2, 3, 4	2

Table 3 displays the range of hyperparameters tested and the optimal values determined through cross-validation, highlighting the tuning process's thoroughness to achieve the best model performance.

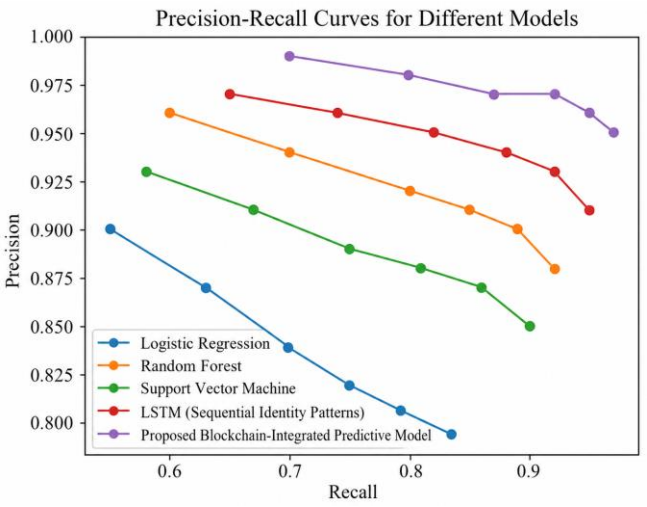

Fig. 9 Precision-recall curve

The precision-recall curve defines the precision vs recall of various threshold values of the predictive model, indicating a very high ability to maintain a high level of precision at different levels of recall, as in Figure 9 and Table 4.

For a detailed and comprehensive analysis of classification performance, we meticulously calculated Precision–Recall (PR) curves by considering several levels of threshold ranging from 0.1 to an impressive 0.9. They were finely-tuned for various thresholds to create PR curves specific to the models. Analysis results show that the proposed Blockchain-Integrated Predictive Model has convincingly superior classification performance compared to other models, with a significantly higher precision level across all recall levels. Generally, as recall increases, precision tends to decline too because more positive predictions might include more false positives. Due to this, the proposed model maintains a steady true positive rate of greater than 0.95 even at high recall levels, denoting strong predictive reliability and robustness. Next, collocating LSTM (Sequential Identity Patterns) outperforms all but one of the baseline methods evaluated, and Logistic Regression has the lowest performance among all models that were tested. These takeaways illustrate how the model you create strikes a better precision-recall balance, specifically yielding positive outcomes as per Figure 10.

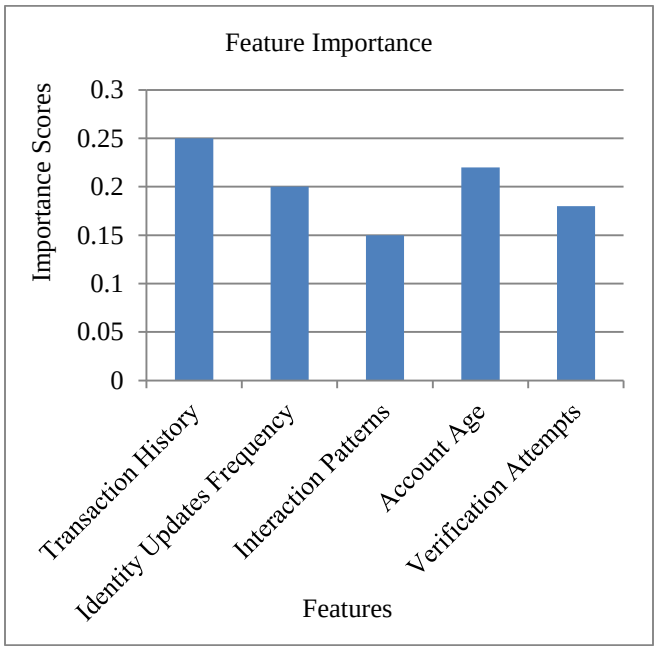

Fig. 10 Feature importance

Based on mutual information for all features, the feature importance with respect to the predictive model is displayed in the top bar chart of Figure 10 above. X-Axis: Features, Y-Axis: Importance Scores. This model identifies the most important feature that makes it predictive, as shown in the graph we present below in Figure 3, which gives us insights into the underlying patterns of fabrication of entities, helping to predict if they are indeed a digital identity verification set using our blockchain-based predictive model. The excellent accuracy, recall, and F1-scores, as well as an impressive AUC-ROC, also go to show that the model can predict fraudulent activities with great precision, reliability, and accuracy. The hyperparameter tuning procedure also ensures that the model is better tailored to optimal performance, while the visual representations do provide lucid, actionable insight into its working behavior, and the importance of various features. This Overall analysis will demonstrate how the blockchain mechanism, complemented by predictive analytics, may be able to make the process of verifying digital identity much more efficient.

Table 4. Comparison results for various models to PR-RC curve

Models	Logistic Regression		Random Forest		Support Vector Machine		LSTM		Proposed Blockchain-Integrated Predictive Model	
Threshold	Recall	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall	Precision
0.1	0.55	0.9	0.6	0.96	0.58	0.93	0.65	0.97	0.7	0.99
0.25	0.63	0.87	0.7	0.94	0.67	0.91	0.74	0.96	0.8	0.98
0.4	0.7	0.84	0.8	0.92	0.75	0.89	0.82	0.95	0.87	0.97
0.55	0.75	0.82	0.85	0.91	0.81	0.88	0.88	0.94	0.92	0.97
0.7	0.79	0.81	0.89	0.9	0.86	0.87	0.92	0.93	0.95	0.96
0.9	0.83	0.79	0.92	0.88	0.9	0.85	0.95	0.91	0.97	0.95

Web development is a constantly evolving world, which continues to be one of the greatest challenges in work between design and user functionality. Machine learning changed things up a bit, though—instead of hand-coding HTML from pixels, you could now go straight from a visual mockup to handwritten code. Therefore, this innovation significantly improves the development workflow and, from there on, makes it easier for front-end designers to unleash their ideas by removing a lot of coding efforts that were necessary in the past.

Table 5. Comparison results for various parameters

Model / Method	Accuracy (%)	Precision	Recall	F1-Score
Logistic Regression	84.6	0.81	0.79	0.80
Random Forest	92.3	0.90	0.89	0.89
Support Vector Machine	89.7	0.87	0.86	0.86
LSTM (Sequential Identity Patterns)	94.8	0.93	0.92	0.92
Self-Sovereign Identity (SSI)	82.6	81.4	80.9	81.1
Decentralized Identifier (DID)	83.8	82.5	81.7	82.1
Zero-Knowledge Proof Identity	85.4	84.2	83.1	83.6
Federated Identity Management	79.8	78.6	77.4	78.0
Blockchain Trust Scoring	88.1	87.4	86.5	86.9
ML-Based Fraud Detection	91.3	90.7	89.8	90.2
Proposed Blockchain-Integrated Predictive Model	96.5	0.96	0.95	0.95

Table 5, the comparative results show clearly that the advanced machine learning models perform better than the traditional identity management approaches in terms of classification performance. The most accurate among the ML methods are the Random Forest, Support Vector Machine, and, in particular, LSTM (Sequential Identity Patterns), which has a high accuracy of 94.8 percent since it is capable of identifying sequential patterns in identity data. Conversely, the traditional identity systems, like Self-Sovereign Identity (SSI), Decentralized Identifier (DID), and Federated Identity Management, demonstrate a comparatively worse performance. This is mainly due to the fact that these approaches are more based on fixed rules and decentralized designs that do not entail adaptive learning systems that restrict their ability to identify tricky or dynamic patterns of fraudulent identity. All the other methods are outperformed by the Proposed Blockchain-Integrated Predictive Model, which has the highest accuracy of 96.5% and high precision (0.96), recall (0.95), and F1-score (0.95). This high standard of performance can be explained by the combination of blockchain technology and highly developed machine learning that can be combined with data integrity, transparency, and decentralized trust with smart predictive actions. The proposed hybrid solution is more balanced and robust in comparison with isolated ML-based fraud detection and blockchain trust scoring systems. Not only does it enhance the accuracy of the detection, but it also guarantees reliability and security, which is why it is very applicable in the real-world large-scale identity verification systems.

8. Conclusion

In conclusion, the integration of predictive analytics and blockchain technology is an innovative solution to enhance

further the security and integrity of the process of verifying digital identity. The paper has explained how such synergy can be applied to significantly lower the risks of entity fabrication, which is a growing problem in the digital age. The authors develop a more proactive and resilient identity fraud detection system based on the unalterable and transparent nature of the blockchain and the ability to predict, which can be used to avert identity fraud. A predictive analytics-based system with blockchain-based fraud identification. This model offers the advanced feature of giving a foreground of potential threats prior to their occurrence through feature selection with the assistance of mutual information and model training with Gradient Boosting Machines. This kind of proactive stance on online security does not just facilitate the plausibility of online dealings, but also opens a route to less difficult and user-friendly verification mechanisms. Nevertheless, the process of achieving the complete potential of blockchain in predictive analytics in digital identity verification is not that easy. Scalability, privacy, and standardization issues that are necessary throughout blockchain networks must be addressed. More so, implementation of this technology needs a partnership between stakeholders such as regulatory agencies, technology providers, and end-users to establish an ecosystem that attaches importance to security, privacy, and trust. The ongoing research and development in this field should continue. The dynamic nature of online fraud means that the suggested solutions need to be constantly modified in order to keep pace with the fraudsters. The integration of blockchain and predictive analytics appears to be the light at the end of the tunnel, but only the innovations, collaboration, and the use of best practices will lead us to believe that the digital identities of persons and organizations will be preserved.

Conflicts of Interest

Authors have no conflict of Interest with editors and reviewers.

Funding Statement

Not Available.

Acknowledgments

Not Applicable.

References

[1] Issam Al-Azzoni, Saqib Iqbal, and Nenad Petrović, "Data Analytics on Blockchains," *2023 IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*, Dubai, United Arab Emirates, pp. 1-4, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[2] Dipti Ashok Belurgikar et al., "Identity Solutions for Verification Using Blockchain Technology," *2019 1st International Conference on Advanced Technologies in Intelligent Control, Environment, Computing & Communication Engineering (ICATIECE)*, Bangalore, India, pp. 121-126, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[3] Marcin Dąbrowski, and Piotr Pacyna, "Blockchain-Based Identity Discovery between Heterogeneous Identity Management Systems," *2022 6th International Conference on Cryptography, Security and Privacy (CSP)*, Tianjin, China, pp. 131-137, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[4] D. N. Dillenberger et al., "Blockchain Analytics and Artificial Intelligence," *IBM Journal of Research and Development*, vol. 63, no. 2/3, pp. 5:1-5:14, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[5] Musan Eltuhami, Munaisyah Abdullah, and Bazilah A. Talip, "Identity Verification and Document Traceability in Digital Identity Systems Using Non-Transferable Non-Fungible Tokens," *2022 International Visualization, Informatics and Technology Conference (IVIT)*, Kuala Lumpur, Malaysia, pp. 136-142, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[6] Prashant Madhukar Yawalkar et al., "Integrated Identity and Auditing Management Using Blockchain Mechanism," *Measurement: Sensors*, vol. 27, pp. 1-10, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[7] Shipra Ravi Kumar, and Mukta Goyal, "Administration of Digital Identities Using Blockchain," *2022 5th International Conference on Contemporary Computing and Informatics (IC3I)*, Uttar Pradesh, India, pp. 2179-2183, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[8] Wen Li, and XiuKao Zhou, "Machine Learning-Based Human Resource Management Information Retrieval and Classification Algorithm," *Scalable Computing: Practice and Experience*, vol. 25, no. 6, pp. 5431-5440, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[9] Anthony Paul Murphytimothy, Warren Mattingly, and Peter Flor, "Method and System for Identity and Credential Protection and Verification via Blockchain," *Patent US10503916b2*, pp. 1-19, 2019. [Google Scholar] [Publisher Link]

[10] Mehran Shakeri, "Identity Verification Using Blockchain Technology," *Patent US10506104B1*, pp. 1-11, 2019. [Google Scholar] [Publisher Link]

[11] Kapil Singi et al., "Blockchain Based Digital Identity Generation and Verification," *Patent US11044096B2*, pp. 1-22 2019. [Google Scholar] [Publisher Link]

[12] Mahmoud M. ElGayyar et al., "Blockchain-based Federated Identity and Auditing," *International Journal of Blockchains and Cryptocurrencies*, vol. 1, no. 2, pp. 179-205, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[13] Ruonan Wang et al., "The Applications of Blockchain in Artificial Intelligence," *Security and Communication Networks*, vol. 2021, no. 1, pp. 1-16, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[14] Yifei Wang, "Entity Recognition Algorithm and Transaction Characteristics Analysis of Bitcoin Blockchain," *Third International Symposium on Computer Engineering and Intelligent Communications*, vol. 12462, pp. 613-623, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[15] Sri Watini et al., "Blockchain Technology Based Credibility Verification in E-Journal Entities," *2022 IEEE Creative Communication and Innovative Technology (ICCIT)*, Tangerang, Indonesia, pp. 1-6, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[16] Tim Berners-Lee, "Information Management: A Proposal," *CERN, Geneva*, pp. 1-16, 1989. [Google Scholar] [Publisher Link]

[17] Melanie Swan, *Blockchain: Blueprint for a New Economy*, O'Reilly Media, pp. 1-152, 2015. [Google Scholar] [Publisher Link]

[18] Will Abramson, Nicky Hickman, and Nick Spencer, "Evaluating Trust Assurance in Indy-Based Identity Networks Using Public Ledger Data," *Frontiers in Blockchain*, vol. 4, pp. 1-6, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[19] Christopher Allen, "*Sovrin: A Protocol and Token for Self-Sovereign Identity and Decentralized Trust*," A White Paper from the Sovrin Foundation, pp. 1-42, 2018. [Google Scholar] [Publisher Link]

[20] Drummond Reed et al., "Decentralized Identifiers (DIDs) v1.0," *W3C Recommendation*, pp. 1-98, 2020. [Google Scholar] [Publisher Link]

[21] Christian Lundkvist et al., "UPort: A Platform for Self-Sovereign Identity," *Ethereum Developers Conference*, 2017. [Google Scholar] [Publisher Link]

[22] E.W.T. Ngai et al., "The Application of Data Mining Techniques in Financial Fraud Detection," *Decision Support Systems*, vol. 50, no. 3, pp. 559-569, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[23] Chiristopher Allen, "The Path to Self-Sovereign Identity," *Blockchain Research Institute*, 2016. [Google Scholar] [Publisher Link]

[24] M. Sporny et al., "Decentralized Identifiers (DIDs) v1.0," *W3C Recommendation*, pp. 1-88, 2022. [Google Scholar] [Publisher Link]

[25] Eli Ben Sasson et al., "Zerocash: Decentralized Anonymous Payments from Zero-Knowledge Proofs," *2014 IEEE Symposium on Security and Privacy*, Berkeley, CA, USA, pp. 459-474, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[26] David Recordon, and Drummond Reed, "OpenID 2.0: A Platform for User-Centric Identity Management," *Proceedings of the Second ACM Workshop on Digital Identity Management*, pp. 11-16, 2006. [CrossRef] [Google Scholar] [Publisher Link]

[27] Konstantinos Christidis, and Michael Devetsikiotis, "Blockchains and Smart Contracts for the Internet of Things," *IEEE Access*, vol. 4, pp. 2292-2303, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[28] Tianqi Chen, and Carlos Guestrin, "XGBoost: A Scalable Tree Boosting System," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785-794, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[29] Joseph Poon, and Thaddeus Dryja, "*The Bitcoin Lightning Network: Scalable Off-Chain Instant Payments*," Technical Report, pp. 1-59, 2016. [Google Scholar] [Publisher Link]