# LABELCHAIN: A DECENTRALIZED DATA LABELING PLATFORM FOR WEB3

Mrityunjay Kumar Gupta[1], Md Nomaan Zeya[1], Amarkant Upadhyay[1], Anmol Kumar[1], Smita Dash[1], and Nushrat Praveen[1]

[1] Department of Computer Science and Engineering,
R.V.S College of Engineering and Technology, Jamshedpur, India
mjguptacse@gmail.com
zeyamdnomaan@gmail.com
harshroy428@gmail.com

**Abstract.** This study introduces LabelChain, a decentralized platform for managing data labeling tasks efficiently and securely. By leveraging blockchain technology and cloud storage.LabelChain is a decentralized platform designed to streamline data labeling tasks while ensuring security, efficiency, and transparency. By leveraging blockchain technology and cloud storage, it provides a robust solution for managing and verifying labeling tasks without reliance on intermediaries. Users can post labeling tasks along with associated assets, which are securely stored on AWS S3 using pre-signed URLs. This ensures data integrity while allowing seamless access for workers.Workers can browse available tasks, accept them, and submit completed work for verification. Upon approval, payments are processed in Solana (SOL), benefiting from the blockchain's high-speed, low-cost transactions. The use of the Solana blockchain not only accelerates payment processing but also enhances transparency, ensuring that all transactions are securely recorded and easily auditable. To safeguard user data and private keys, LabelChain incorporates strong security measures, fostering a reliable and trustworthy environment. This decentralized approach significantly improves the efficiency of data labeling by removing bottlenecks associated with traditional task management systems. With its scalable infrastructure and blockchain-powered verification process, LabelChain offers a future-ready alternative that reduces costs, enhances security, and streamlines workflow management, making it an ideal solution for large-scale data annotation needs.

**Keywords:** Decentralized Data Labeling, Blockchain, Cryptocurrency, Web3

## 1 Introduction

Managing data labeling tasks efficiently and securely is a critical challenge in the field of artificial intelligence (AI) and machine learning (ML). High-quality labeled data is essential for training accurate and reliable AI models, yet traditional centralized solutions often suffer from inefficiencies, high operational costs, and security vulnerabilities. These challenges hinder scalability and create bottlenecks in the data annotation pipeline, limiting the effectiveness of AI-driven applications in mainstream adoption. As AI continues to shape industries such as healthcare, finance, and autonomous systems, the demand for a secure, cost-effective, and scalable data labeling solution has become more pressing than ever.

This study introduces LabelChain, a decentralized platform designed to streamline the data labeling process by leveraging blockchain technology and cloud storage. By utilizing Solana for fast, transparent, and cost-effective payments and AWS S3 for secure asset storage, LabelChain creates a seamless and scalable ecosystem where users can post tasks, workers can efficiently complete labeling assignments, and transactions are executed without intermediaries. Unlike traditional platforms that rely on centralized servers, LabelChain ensures decentralization, immutability, and trustless collaboration, making it a more secure and transparent alternative for large-scale data annotation needs.

The platform enhances trust and efficiency by ensuring secure task verification, transparent payment processing, and reduced transaction costs. Furthermore, LabelChain implements robust security measures to protect user data and private keys, ensuring a secure and tamper-proof labeling process. By bridging the gap between AI data requirements and decentralized solutions, LabelChain presents a transformative approach to data annotation, paving the way for mainstream adoption of a more secure, efficient, and scalable labeling ecosystem. As the AI industry continues to expand, solutions like LabelChain will play a crucial role in democratizing access to high-quality labeled data, fostering innovation, and accelerating AI advancements across industries.

## 2 Literature Survey

The rapid expansion of artificial intelligence (AI) and machine learning (ML) has increased the demand for high-quality labeled datasets. Traditional data labeling platforms face several challenges, including high costs, inefficiencies, security risks, and lack of transparency in payment and task verification. Blockchain technology has emerged as a potential solution for addressing these limitations by enabling decentralized, transparent, and tamper-proof mechanisms for task execution and payment. Several studies have explored the benefits of decentralized task management for AI applications.

W. Gan et al.[1] introduced a blockchain-powered crowdsourcing system to improve task execution transparency and ensure fraud-resistant payments. The researchers emphasized that decentralization could eliminate intermediary fees, making data labeling tasks more cost-effective and accessible to a global workforce. Additionally, the immutability of blockchain ensures a tamper-proof history of labeled data, providing a trust layer for AI model developers.

F. Almeida et al. [2] explored the integration of smart contracts for fair payment distribution in annotation tasks. The research demonstrated how Ethereum-based smart contracts could be leveraged for automated payments to workers, reducing disputes over compensation. However, the study highlighted the limitations of Ethereum's high gas fees, suggesting that Solana or Polygon could provide faster and cost-efficient alternatives for microtransactions in data labeling ecosystems.

Q.Wang et al. [3], researchers proposed an incentive-based system where annotators' reliability scores determined their eligibility for high-value tasks. The study showed that weighted voting mechanisms improved annotation accuracy by filtering out unreliable labels.

Statista et al. [4] studied the use of reputation-based models to improve the consistency and reliability of labeled datasets in decentralized environments. By employing trust metrics, these models evaluate the credibility of contributors and their labels, ensuring higher data quality. This approach helps mitigate inconsistencies caused by malicious or inaccurate labeling, a common challenge in decentralized systems. The reputation system incentivizes honest behavior by rewarding trustworthy contributors and reducing the influence of unreliable ones.

Hassanein et al. [5] introduced an ML-based framework for assessing worker credibility in annotation tasks, leveraging historical task performance data to adjust payment structures dynamically. This aligns with LabelChain's goal of creating a secure, trust-based ecosystem for labelers and data providers.

S.Zheng et al. [6] examined how AWS S3 pre-signed URLs could be used to facilitate secure file uploads while maintaining scalability. The study compared centralized and decentralized storage solutions, ultimately concluding that hybrid approaches (e.g., combining AWS S3 with blockchain verification) offer the best balance between performance and security.

S. I. Kim et al. [7] explored end-to-end encryption in AI data pipelines, ensuring that labelers only access necessary metadata while preserving data privacy. These findings support the privacy-centric approach adopted by LabelChain, where secure access mechanisms prevent unauthorized usage of sensitive datasets.

Renduchintala et al. [8] evaluated various blockchain networks for micropayments, demonstrating that Solana's low-latency architecture could provide near-instant settlements with minimal fees. This aligns with LabelChain's use of Solana for transparent, on-chain payment processing, reducing payment disputes and transaction costs compared to traditional fiat-based labeling platforms.

R. Khan et al. [9] examined the role of stablecoins in AI task marketplaces, arguing that pegged digital assets could reduce volatility in worker payments. While LabelChain primarily utilizes SOL, future extensions might incorporate stablecoin-based payouts to improve financial stability for workers.

Ethereum et al. [10] investigated how smart contracts could automate task verification by integrating machine learning models to validate annotations. The study demonstrated that multi-party dispute resolution involving validators, task posters, and labelers significantly reduced annotation fraud.

M. Alharby et al. [11] proposed a trustless arbitration model where blockchain-based voting mechanisms resolved disputes without centralized authorities. These findings reinforce LabelChain's approach to transparent and automated task verification, where blockchain-based validation systems minimize conflicts.

B. Shrimali et al. [12] analyzed Layer 2 solutions designed to scale blockchain-based microtasks, focusing on how rollups and sidechains can significantly enhance throughput. As blockchain networks face challenges with scalability, Layer 2 solutions have emerged as an effective way to offload the transaction processing from the main blockchain (Layer 1) to secondary layers, thus improving efficiency. Rollups, which bundle multiple transactions into a single one, reduce the burden on the main chain while maintaining security. There are two main types of rollups: Optimistic Rollups, which assume transactions are valid by default and only check them when disputed, and ZK-Rollups, which use zero-knowledge proofs for more immediate validation.

A.Ringen et al. [13] explored the concept of cross-chain interoperability, which allows different blockchain networks to communicate and share data seamlessly. The study suggested that integrating multi-chain settlement options into LabelChain could enhance its functionality by enabling transactions across various blockchain platforms. This approach would provide users with more flexibility, allowing them to transact using different cryptocurrencies or digital assets, regardless of the blockchain network. By incorporating cross-chain capabilities in future versions, LabelChain could expand its reach, improve user experience, and offer broader market opportunities, addressing the growing demand for multi-chain compatibility in decentralized applications.

M.Naeem et al. [14], is the integration of on-chain reputation scores to dynamically assign labeling tasks based on worker expertise. This could be a potential enhancement for LabelChain, ensuring that high-value tasks are allocated to top-performing labelers, increasing overall dataset quality.

The literature highlights the transformative potential of blockchain technology in AI data labeling. Studies have demonstrated that decentralized task management, secure cloud storage, transparent payments, and smart contract-based verification significantly enhance the efficiency, security, and trustworthiness of annotation platforms. By leveraging Solana's high-speed blockchain, AWS S3's secure storage, and smart contract-driven task execution, LabelChain addresses the challenges of traditional data labeling platforms while paving the way for a scalable, decentralized ecosystem.

Future research could explore multi-chain interoperability, stablecoin-based payments, and AI-driven quality control mechanisms to further optimize decentralized labeling workflows. LabelChain's decentralized approach positions it as a pioneering solution in the evolving landscape of blockchain-powered AI data marketplaces.

## 3 Methodology

LabelChain relies on a combination of hardware and software components to implement its decentralized data labeling solution. The backend is developed using Node.js and Express.js to handle API requests and business logic. The

Solana blockchain, integrated via Web3.js, facilitates transparent and trustless transactions. AWS S3 is used for cloud storage, with pre-signed URLs ensuring secure asset management. PostgreSQL serves as the database for task and user management, providing structured and efficient data handling. The frontend is built using React and Redux, enabling a responsive and interactive user interface. Authentication and cryptocurrency transactions are managed through Phantom Wallet, ensuring seamless and secure interactions.

### 3.1 System Configuration

The system configuration for LabelChain is designed to deliver a robust, secure, and scalable platform. The front-end is developed using React and Redux, providing a dynamic and user-friendly interface while efficiently managing application state. The back-end is powered by Node.js and Express.js, serving as the core for handling API requests, server-side logic, and seamless communication between the client and the blockchain. Blockchain integration is achieved through Solana using Web3.js, enabling decentralized functionalities such as asset transfers and smart contract interactions. For asset management, AWS S3 is

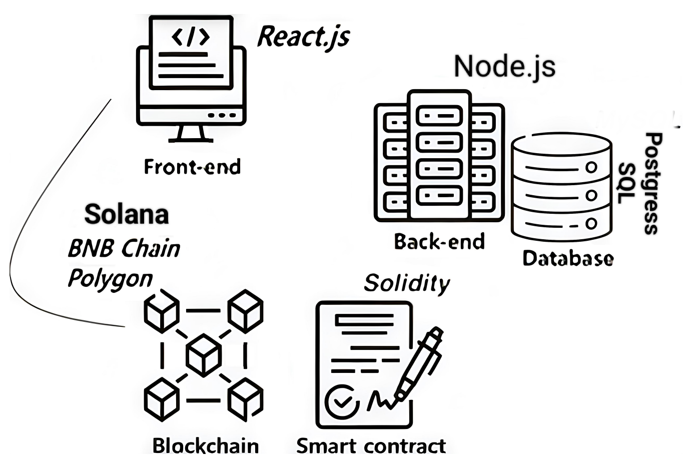


**Fig. 1.** System Configuration

utilized with pre-signed URLs, ensuring secure, scalable, and efficient storage and retrieval of digital assets. The database layer is built on PostgreSQL, which manages tasks, user data, relationships, and critical application metadata, ensuring data consistency and reliability. Authentication and cryptocurrency transactions are facilitated by the Phantom Wallet, enabling secure and seamless blockchain interactions for users. The system also incorporates advanced security measures, including best practices for private key management, encrypted communications, and secure transaction handling, ensuring the safety and integrity of all user and system data.

Furthermore, the integration of automated monitoring tools and performance optimization strategies ensures high availability and quick transaction processing. The use of Kubernetes and Docker for containerization ensures that the system is scalable, fault-tolerant, and able to handle increased user load efficiently. With the inclusion of advanced cryptographic techniques and decentralized identity management, the system guarantees a higher level of privacy and transparency for all user interactions. Additionally, future scalability is ensured through regular updates and system improvements based on evolving industry needs, enabling LabelChain to remain competitive and meet growing user demands.ling decentralized functionalities such as asset transfers and smart contract interactions.

### 3.2 System Architecture

The overall system architecture and flow for LabelChain is shown in figure 2. Users access the system through Phantom Wallet for authentication and transaction signing, securely linking their wallet address to the system. Once authenticated, users can log in and perform blockchain-based actions like labeling tasks, storing assets, or managing cryptocurrency transactions using Solana and Web3.js. The system ensures that all user actions are verified and executed securely on the blockchain network.

The back-end, powered by Node.js and Express.js, continuously monitors blockchain activities via a daemon, which parses transaction data and stores it in PostgreSQL. The front-end retrieves data from the database to display updates, transaction history, and usage activity in real-time. Additionally, AWS S3 handles secure asset storage, ensuring scalability and efficient access to digital assets. Robust private key management and encryption protocols safeguard user data and maintain system integrity throughout all operations.

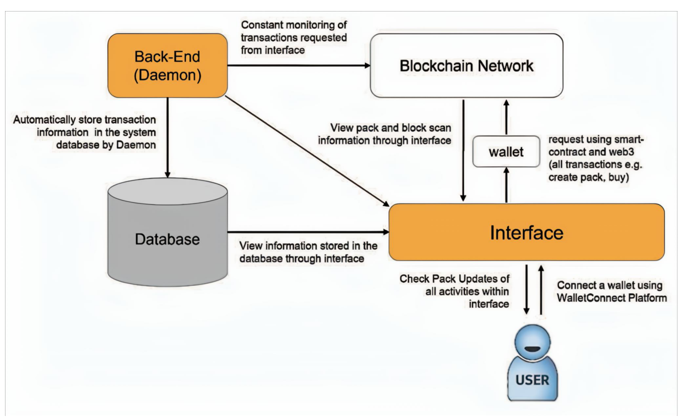


**Fig. 2.** System Architecture and Flow

### 3.3 System Data Flow and Management

The system data flow and management for LabelChain is shown in figure 3. The client initiates actions through Phantom Wallet to interact with the blockchain, sending transactions for smart contract execution. A POST request is then made to the back-end to store transaction data in the PostgreSQL database. The back-end continuously monitors blockchain events related to the smart contracts. Upon detecting relevant events, it extracts the event logs and transforms them into a format that the client can use. The database is updated with the latest transaction and asset information. The system ensures that all transactions are processed efficiently and securely, maintaining the integrity of the data.

When the client requests specific data, such as transaction history or asset details, the back-end retrieves the corresponding data from the database and returns it to the client. This ensures that the client has access to accurate and up-to-date information at all times. The process of fetching data from the database is optimized for speed, ensuring low-latency access for real-time interaction with the blockchain. This seamless communication between the client, back-end, and blockchain ensures a smooth user experience and efficient operation.

Similarly, when the client sends new data, such as asset details or transaction requests, the back-end first stores this information in the database. It then communicates with the blockchain node to fetch the latest details, ensuring that the database is always up-to-date with accurate and real-time information. This two-way communication ensures consistency across all layers of the system, preventing any discrepancies between the blockchain and the database. Additionally, the system is designed to handle high volumes of data, ensuring scalability as user demand grows.

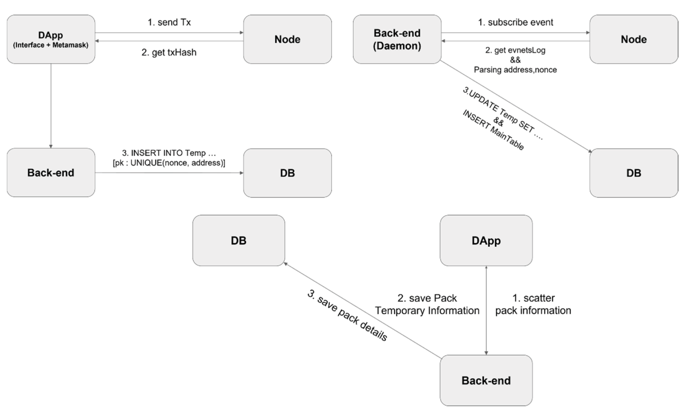


**Fig. 3.** System Data Flow and Management

### 3.4 Implementation

The proposed decentralized application, LabelChain, a Web3-based data labeling platform, was implemented and tested in the environment shown in Table 1. The DApp consists of two sub-servers and one main server, where the sub-servers are divided into users and administrators, and the main server is configured to host the web server and the database server.

**Table 1.** IMPLEMENTATION AND TEST ENVIRONMENT

| **Item** | **Content** |
|---|---|
| OS | Windows 11 |
| Smart Contract | Solidity |
| Distributed Computing Platform | Solana |
| Database | PostgreSQL |
| Server | Node.js, Express.js |
| Web | React.js |

The system is built around a smart contract written in Solidity, designed for seamless interaction with the Solana blockchain. The Hardhat framework was utilized for efficient development, testing, and automated deployment of the smart contracts. The user interface was implemented using the React.js framework, while the back-end was developed using Node.js and Express.js in a single web server architecture.

The deployment process consists of three stages: development, staging, and production. During the development phase, the system was deployed and tested on a local development server within a secure VPN environment, with all changes version-controlled in a VCS. In the staging phase, the smart contract was deployed to the Solana testnet, and the service was hosted on a Kubernetes engine within AWS for further testing. For the production phase, the smart contract was deployed to the Solana mainnet, and the service was containerized using Docker and deployed on a dedicated Kubernetes engine for end-users.

Deployment automation was handled using Jenkins, ensuring a smooth CI/CD pipeline. Docker was employed to containerize the application, bundling necessary dependencies and ensuring consistency across environments. Kubernetes managed the containerized services, providing scalability, disaster recovery, and efficient resource utilization. Additionally, the integration with cloud services enabled rapid scaling to handle high traffic while optimizing costs.

The React.js library was utilized to build the front-end, enabling an intuitive interface and efficient event handling. figure 4 Solana Wallet Setup figure 5 demonstrates the page for transacton approval , while figure 6 displays the interface for viewing completed tasks and their associated blockchain data. Users can share links to completed tasks via social media, allowing others to view details and interact with the system securely and transparently.

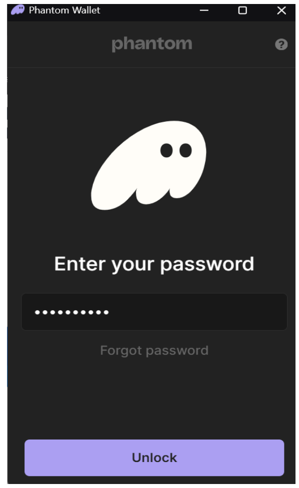


**Fig. 4.** Solana Wallet Setup

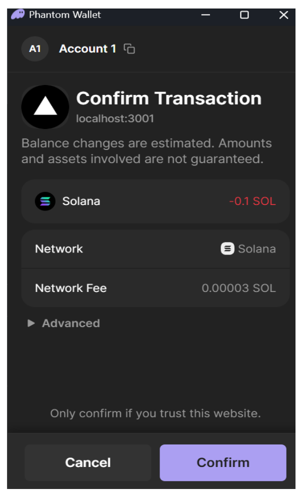


**Fig. 5.** Transaction Approval From Wallet

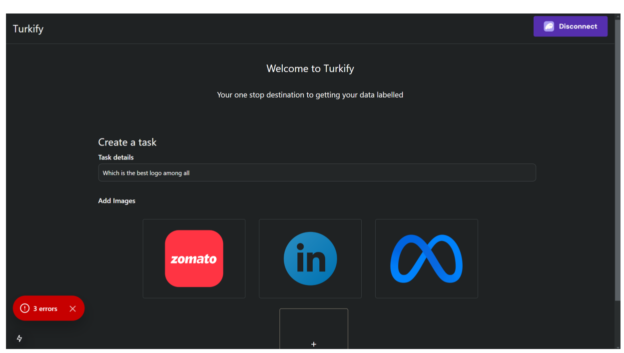


**Fig. 6.** Task Post UI

### 3.5 Evaluation

The related work proposed in this paper leverages the Solana blockchain network to enhance security, transparency, and efficiency in the data labeling process. LabelChain achieves higher security and transparency by storing task information and transaction records immutably on the Solana blockchain, which is enforced by smart contracts. While the speed of operations is slightly slower due to blockchain confirmation time and wallet interactions, and transaction fees add additional costs, the benefits outweigh these limitations. Forgery is impossible as all records on the blockchain are immutable, whereas traditional applications rely heavily on centralized databases and are prone to tampering.

In terms of performance evaluation, LabelChain on Solana demonstrates significant advantages. Solana's high throughput ensures near-instant transaction confirmations, with block times averaging around 400 milliseconds. Transaction fees are exceptionally low, often less than 0.01, making the system cost-effective for creators and users. The system automatically converts transaction fees into Solana's utility token, streamlining the payment process.

After executing transactions, users can view details such as block time, transaction speed, and fees using Solana blockchain explorers like Solscan or Solana Beach by providing their wallet address. LabelChain effectively utilizes Solana's scalable and low-cost infrastructure to deliver a seamless and secure user experience, ensuring reliability and transparency in all operations.

### 3.6 Conclusion and Future Work

*LabelChain leverages blockchain to create a secure, decentralized data labeling platform, eliminating inefficiencies in traditional systems. Using Web3 wallets, users manage tasks without traditional authentication, with transactions recorded on the Solana blockchain for transparency. Validators and labelers earn Solana-based tokens, fostering a token-driven economy. Smart contracts and decentralized storage enhance security, while Solana's scalability ensures efficiency. Future developments include multi-chain support (Ethereum, Polygon), AI-driven quality control for improved accuracy, and decentralized identity (DID) for enhanced privacy and authentication. The platform will also integrate with decentralized data marketplaces for data monetization and launch mobile applications for broader accessibility. Enhanced tokenomics, including staking and rewards, will further drive user engagement, accelerating the adoption of decentralized applications and blockchain technology.*

## References

1. W. Gan, Z. Ye, S. Wan and P. S. Yu. 2023. "Web 3.0: The Future of Internet." arXiv preprint arXiv:2304.06032 (2023).

2. F. Almeida, J. D. Santos and J. A. Monteiro. 2014. “E-commerce Business Models In The Context of Web 3.0 Paradigm.” arXiv preprint arXiv:1401.6102 (2014).
3. Q. Wang, R. Li, Q. Wang, and S. Chen. 2021. “Non-fungible token (nft): Overview, evaluation, opportunities and challenges.” arXiv preprint arXiv:2105.07447 (2021).
4. Statista. Global spend on e-commerce platforms 2020-2025 — statista.com. https://www.statista.com/statistics/1385734/global-spend- on-e-commerce-platforms/. (Accessed on 10/01/2023).
5. A. A. Hassanein, N. E. Tazi and N. N. Mohy. “Blockchain, Smart Contracts, and Decentralized Applications: An Introduction, Implementing and Leveraging Blockchain Programming,” pp. 97-114, Springer, 2022.
6. S. Zheng, J. Luo, E. Dong, C. Chen and X. Liu. 2022. “SPENDER: A Platform for Secure and Privacy-Preserving Decentralized P2P E- Commerce.” arXiv preprint arXiv:2206.07215 (2022)..
7. S. I. Kim and S. H. Kim. “E-commerce payment model using blockchain,” Journal of Ambient Intelligence and Humanized Computing, pp. 1673-1685, Springer, 2020.
8. T. Renduchintala, H. Alfauri, Z. Yang, R. D. Pietro and R. Jain, “A Survey of Blockchain Applications in the FinTech Sector,” Journal of Open Innovation: Technology, Market, and Complexity, vol. 8, no. 4, 185, 2022..
9. R. Khan and T. A. Hakami, “Cryptocurrency: usability perspective versus volatility threat,” Journal of Money and Business, vol. 2, no. 1, pp. 16-28, 2021.
10. Ethereum. Introduction to DApps — ethereum.org. https://ethereum.org/en/developers/docs/dapps/. (Accessed on 10/05/2023).
11. [15] M. Alharby and A. V. Moorsel. 2017. “Blockchain-based Smart Contracts : A Systematic Mapping Study.” arXiv preprint arXiv:1710.06372 (2017).
12. B. Shrimali and H. B. Patel, “Blockchain state-of-the-art: architecture, use cases, consensus, challenges and opportunities,” Journal of King Saud University – Computer and Information Sciences, vol. 34, no. 9, pp. 6793-6807, 2022.
13. [22] L. A. Jaatun, A. Ringen and M. G. Jaatun. “Yet Another Blockchain- based Privacy-friendly SocialNetwork.” In 2022 IEEE International Conference on Cloud Computing Technology and Science (CloudCom), 13-16 Dec 2022, Bangkok, Thailand, pages 222-229. IEEE, 2022.
14. M. Naeem, M. Hameed and M. S. Taha. “A Study of electronic payment system.” In 2019 IOP 1st International Symposium on Engineering and Technology (ISETech), 23 Dec 2019, Perlis, Malaysia, IOP, 2019.