



A novel Hyperledger blockchain-enabled decentralized application for drug discovery chain management

Neetu Sharma^{*}, Rajesh Rohilla

Delhi Technological University, 110042, India



ARTICLE INFO

Keywords:

Drug discovery
Blockchain
Drug development
Machine learning
Hyperledger fabric
Chain management

ABSTRACT

The aim of the drug discovery process is to create a brand-new drug that is both safe and useful for treating disease in patients. The trustworthy collaboration of the organisations involved in the drug discovery process and the integrity of their contributions are essential for this process. Current drug discovery chains use centralised systems, which are susceptible to lockdown by cyberattacks. Blockchain, with its many characteristics such as accountability, immutability, integrity, privacy, and security, has the potential to be extremely useful in drug discovery chain management. The purpose of this work is to create a drug discovery framework utilizing the strengths of blockchain technology in combination with machine learning (ML). Such a system would provide a secure, efficient, and faster drug development life cycle. This study presents a novel Hyperledger fabric-based drug discovery application that empowers the permissioned organisations to upload, update, view, and verify contributions. ML is used to preprocess data and visualise features. In the proposed work, a unique identifier is assigned to each contribution asset using the secure hash algorithm (SHA-256). The proposed design also enables the regulatory authority to issue certificates proving the ownership of contributions to the contributing organisations. The blockchain ledger has been used to store the *meta*-data and InChIKey of drug contributions, and actual contributions are in off-chain storage. In this work, we have successfully built an end-to-end decentralised drug discovery application with a front-end interface and demonstrated chaincode algorithms. The end-to-end application is not available in any previous work. The Caliper tool has been used to investigate throughput, latency, and resource statistics. The performance and comparative analysis show that the proposed design is scalable and promising for drug discovery chain management.

1. Introduction

Drug discovery is a multi-stage process that involves finding new treatments and cures for diseases. The final medication must be effective for targeted treatment, risk-free, and compliant with all legal requirements. Drug discovery needs the positioning of a varied set of national and international structures supported by health institutes (Daizadeh, 2021). In the pharmaceutical industry, the process of discovering and developing a new drug is expensive and time-consuming. The average estimate for developing a new drug is around \$985 million (Wouters et al., 2020). It takes 12–15 years to develop the permitted medicine (Deore et al., 2019). The involvement of numerous organisations in the drug discovery process can sometimes pose a threat to the integrity of their contributions. Lack of privacy and security in the drug discovery chain can prevent a new drug from coming to market and generate research value from patents (Olsson & Toorani, 2021).

Technical solutions to share discovery chain data among multiple organisations in a transparent manner need to be explored (Andrews et al., 2015).

Pharma 4.0 is renovating the traditional approach to drug discovery through collaborating organisations (Hariry et al., 2022). There are numerous challenges in utilising emerging technologies across the pharmaceutical industry (Lewis & McCallum, 2020) and in creating a common, protected environment in which all stakeholders can work. Blockchain technology provides a decentralised ecosystem with various features. In a blockchain, data is encrypted by a unique hash value and is available at multiple locations instead of being in the custody of a single authority (Sharma & Rohilla, 2020).

The European Commission estimates that \$20 billion is wasted every year on discovering already developed innovations. Sharing research failures poses less of a threat to the pharmaceutical industry because it is hopeless for the company that owns them. A blockchain-based

* Corresponding author.

E-mail addresses: neetusharma85@gmail.com, neetu_2k19phdec25@dtu.ac.in (N. Sharma), rajesh@dce.ac.in (R. Rohilla).

framework can also be utilised for sharing research failures with smart contracts. Machine learning (ML) is applicable in many activities of drug development, like drug screening and molecular generation (Spjuth et al., 2021). ML can be employed in integration with blockchain to reduce the drug discovery period.

1.1. Challenges in existing drug discovery approaches

Existing drug development approaches face numerous challenges, including the following:

- There is no sharing environment among the various drug discovery institutes.
- Fears about knowledge contribution and its ownership (Radanović & Likić, 2018).
- The collection of genomic data in centralised systems is prone to single-point failure.
- Long and very costly clinical trial phases (Kamel Boulos et al., 2018).
- Public safety and privacy during trials.

1.2. Scope of blockchain and ML in drug discovery chain management

Drug discovery plays a significant role in healthcare outcomes. The entire development cycle is a very complex one, and outcomes are very dependent on how it is handled. The more readily we develop a drug, the sooner we get better health outcomes (Omidian & Omidi, 2022). This is possible with the help of blockchain technology in combination with ML. ML is a subset of artificial intelligence that permits software tools to analyze and visualise data. ML can be combined with blockchain technology to pre-process drug development data, support decision-making, and shorten drug development time.

Blockchain is a shared, tamper-proof ledger that governs the verifiable and permanent recording of transactions in a network of entities (Raj Kumar Reddy et al., 2021). Blockchain applications are being researched in many industries, including the life sciences and pharmaceutical domains. Along with being the most popular application for cryptocurrencies, blockchain can also be used to manage health records (Sharma & Rohilla, 2022), manage evidence in digital forensic investigations (Li et al., 2020), manage code copyright (Jing et al., 2021), manage transparent voting systems (Yang et al., 2020), improve the security of smart homes (Ren et al., 2020; Ammi et al., 2021), and more. Blockchain's benefits can also be used for fair exchange (Delgado-Mohatar et al., 2020), solving the password-forgetting problem (Huang & Zhang, 2020), and avoiding intruder manipulation as only authorised users have access rights (Bonna & Shiguang, 2020). Moreover, blockchain technology can be efficiently used for traceability in the supply chain (Agrawal et al., 2021). Tracking products from the manufacturing facility to the delivery location is possible with the help of organisations digitizing physical assets and keeping a decentralised, immutable record of all transactions. Blockchain-based supply chains, in contrast to conventional supply chains, automatically update the data transaction records whenever a change is made. Before the product is transferred to another entity, the concerned parties must fulfill a smart contract condition in order to validate the exchange of goods or services. The blockchain ledger is updated with transaction information after all participants have complied with their obligations and procedures. The blockchain ledger is leveraged to record the product's attributes, such as its nature, quantity, quality, location, and ownership. This can be used by the involved parties to view the complete transaction history of products from source to destination. As a result, the blockchain can be used to create a supply chain that is smarter, safer, and more transparent.

Blockchain features like integrity, safety, confidentiality, accountability, and decentralisation have great potential to benefit the

pharmaceutical industry (Liu et al., 2021). Blockchain technology has an intrinsic ability to eliminate issues related to the security of records (Maity et al., 2021). This study is one such attempt to investigate and demonstrate that blockchain technology can support and add value to the drug discovery process by leveraging the above-mentioned features. We are trying to eliminate the need for a central organisation that controls contributions to the drug discovery process. The suggested blockchain-based solution uses the cryptographic SHA-256 hashing algorithm to store data immutably. This uniform content update mechanism also prevents duplicate recording of contributions.

The following are the benefits of using blockchain in the proposed drug discovery process:

- Eliminates the issue of the integrity of knowledge contributions and their ownership.
- Offers verifiable and immutable databases for drug development, testing, and analytical purposes.
- Accelerates and strengthens the drug discovery procedure by giving secured data access to all bodies involved.
- Offers a secure and reliable environment for sensitive patient data during trials.
- With a transparent, decentralised ecosystem, the human engagement process is made trustworthy.

1.3. Motivation and objective

Protecting the drug development cycle is necessary to improve health outcomes. Most of the existing drug discovery management designs are centralised. There is only one existing blockchain-based study that we could find. This work lacked a design architecture, chaincode algorithms, validation mechanisms, and optimal scalability. Our aim is to create a novel framework for managing contributions in a drug discovery process using blockchain technology in combination with ML. This framework will help in safe, effective, and rapid drug development. It will ensure privacy, integrity, transparency, and ownership of research contributions. Validated information on the blockchain can be made accessible to all the authorised parties involved in the drug discovery process. This would reduce data transfer processing overhead and eventually provide a shorter time to market.

1.4. Contributions

The key contributions are listed below:

- We have designed a novel model for managing drug discovery contributions using blockchain and ML.
- We have only saved the *meta*-data and InChIKey of drug contributions on the blockchain ledger, and the rest is stored in off-chain storage for improved scalability.
- ML integration is done to pre-process and visualise data.
- We have created chaincode modules to upload, view, validate, and update the drug discovery contributions.
- A module has been added to issue a contribution certificate in order to prove the ownership of the contributions.
- We have successfully built an end-to-end decentralised drug discovery application, along with the front-end interface and test APIs (Application Program Interfaces).
- We have improved the performance parameters by increasing the number of endorsers and block sizes. The performance and comparative analysis show that the proposed design is more scalable and has advanced features.

2. Related work

This section covers existing drug discovery and blockchain-based schemes. Drug discovery involves the identification of target molecules, fusion, verification, optimization, etc. In (Deore et al., 2019), the author presented a review of the drug development process. This work lacks practical implementation. In (Olsson & Toorani, 2021), the drug discovery model was created using the Hyperledger framework. This work lacked design architecture, algorithms, and validation mechanisms. In (Andrews et al., 2015), the author proposed a collective drug development service in the form of a database. This was a centralised (single-custody) database, which is less trustworthy. In (Hariy et al., 2022), the authors focused on the clinical testing phase of drug discovery. This study has not practically proven the inferences. In (Lewis & McCallum, 2020), the authors discussed the issues of using emerging technologies for verifying smart pharmacovigilance devices. In (Spjuth et al., 2021), the authors focused on the applicability of cloud computing for ML modeling in drug discovery. Table 1 depicts a summary of existing schemes.

The usability of blockchain for complete drug cycle management was described in (Omidian & Omidi, 2022). The chemical identifiers that are used to identify the chemical structure of substances are represented as strings, also termed InCHI, in the drug development process (Heller et al., 2015). In (Ekins & Bunin, 2013), the authors presented a cloud-based solution for collaborating drug discovery researchers and sharing databases. This work was aimed at improving the drug development process. This scheme lacks integrity and security. In (Mubarakali, 2020), the author introduced the issue of information security in cloud computing. (Nguyen et al., 2019; Thakkar et al., 2018) are existing works that have measured the performance of the Hyperledger network under test using the Caliper tool. The performance analysis of both is performed on a smaller number of organisations. In (Küsters et al., 2020), the author highlighted the accountability feature of Hyperledger Fabric. In (Yu, 2020), the authors emphasised on converting research failures into value creation models using blockchain technology. This

Table 1
Summary of existing drug discovery and healthcare-based schemes.

Existing Scheme	Objective	Limitation
(Olsson & Toorani, 2021)	Created a collaborative drug discovery model using the Hyperledger framework	lacks design architecture, algorithms, validation mechanisms, and demonstration
(Andrews et al., 2015)	Proposed collective drug development service database	less trustworthy, centralised
(Hariy et al., 2022)	Focused on the clinical testing phase of drug discovery	Inferences based on academic journals
(Omidian & Omidi, 2022)	Described the usability of blockchain in drug cycle management	Only theoretical concepts
(Ekins & Bunin, 2013)	Presented a cloud-based solution to collaborate drug discovery researchers	lacks integrity and security
(Yu, 2020)	Emphasised on converting research failures into value creation model using blockchain technology	Design concept only not solution development
(Zhou et al., 2019)	Developed a blockchain based system for PHI management	Unavailability of design architecture
(Cao et al., 2019)	Proposed a cloud assisted EHR model	Scalability is not addressed
(Motohashi et al., 2019)	Suggested a scalable e-health management system using blockchain technology	Scalability testing for multiple clinical trials is not done
(Lee et al., 2020)	Designed a blockchain based EHR exchange system	Lacks design testing, and scalability is also not addressed

survey provided theoretical knowledge but lacked proper implementation.

The blockchain-based management system for information verification was discovered by the authors in (Zhou et al., 2019). The architectural design for this work was unavailable. A cloud-based EHR framework based on blockchain was proposed in (Cao et al., 2019). In this work, scalability is not addressed. In (Motohashi et al., 2019), a framework for a scalable e-health management system using blockchain is suggested. The scalability testing for multiple clinical trials is not done in this work. A blockchain-based, cyber-free model for sharing EHRs was presented in (Lee et al., 2020). This work lacks design testing and scalability. In (Hussien et al., 2019; Mayer et al., 2019; Nguyen et al., 2019), the use of blockchain in the safe sharing of electronic health records (EHRs) was reviewed. This review demonstrates the value of adopting blockchain for managing drug discovery contributions.

3. Proposed drug discovery chain management design

3.1. System overview

Current drug development depends on research institutes and clinical trials. Distributed ledger technology can be used to secure drug discovery chain contributions. This technology has the potential to prevent ransomware attacks on drug discovery data. Generally, multiple organisations need to work together in the drug development process. Blockchain can make it easier to manage the process of uploading, sharing, updating, and validating contributions.

The various stages of the drug discovery chain are genomic data collection, target finding, drug development, animal testing, and clinical trials (Phase I, Phase II, and Phase III). In Stage 1, genomic data is collected and sequenced. In stage 2, the targeted proteins or nucleic acids are detected according to a specific disease. Stage 3 involves all activities related to drug development, such as forming antibodies to compensate for missing proteins. In stage 4, pre-clinical testing of the drug is performed on animals. In stages 5, 6, and 7, phases I, II, and III of clinical trials are conducted to acquire enough evidence for a drug to be effective. In the proposed drug discovery decentralised application (Dapp), the data generated at each stage has been added as a contribution to the blockchain network through a web-based user interface (UI).

3.2. Hyperledger-fabric fundamentals

Drug discovery can not be carried out in a reasonable manner without the privacy and security of drug contributions. Data privacy measures assure the clinical trial participants that their information is secure and safe. It is critical to protect the confidentiality of all contributions in order to foster trust among research participants and uphold the ethical standards of the research process. In the proposed system, Hyperledger Fabric blockchain technology is used. We have used Hyperledger, which is a natively private blockchain, to maintain accessibility only for authorised users. The ledgers are used in this blockchain to maintain records of transactions immutably. These ledgers are accessible only to permissioned users. Multiple organisations can collaborate in the Hyperledger Fabric environment to form a consortium blockchain network. This blockchain framework works on the execute-order-validate transaction model to speed up transaction processing. The SHA-256 algorithm is used for linking the transaction blocks cryptographically, and enhancing security.

The roles of different entities in organisations are defined as user, peer, or admin. Smart contracts that form application logic are known as “chaincode” in Fabric. The simple use-case diagram of the proposed fabric-based drug discovery chain system for managing drug discovery contributions is shown in Fig. 1. The graphical description represents the use-case scenario with limited information, and the detailed description is available in the below sub- sections.

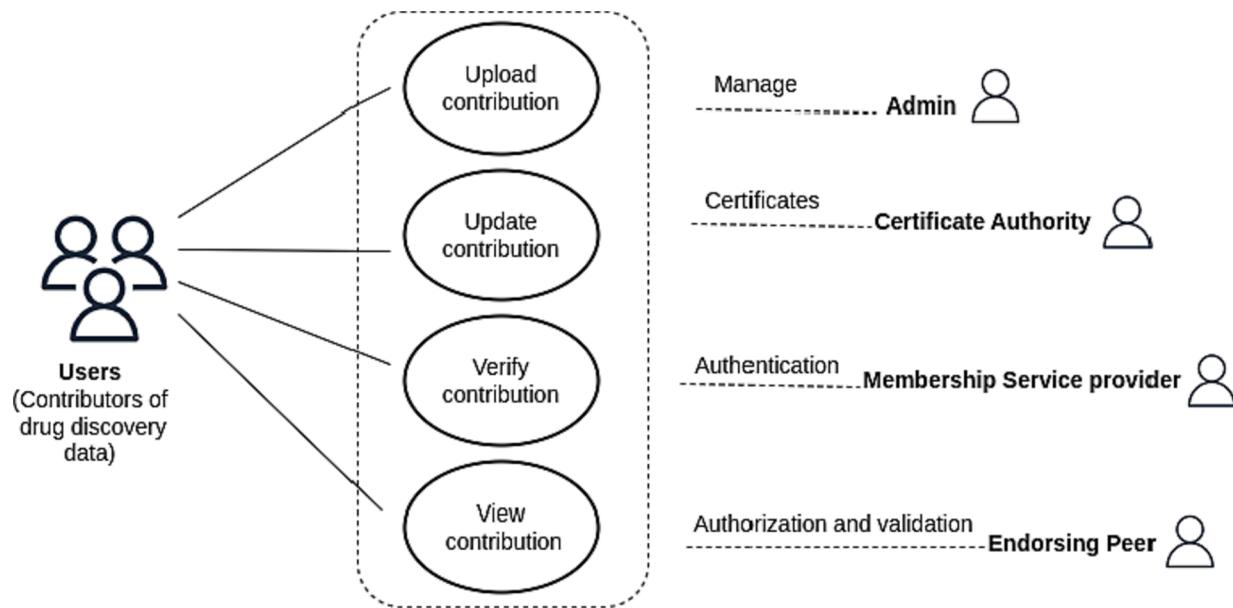


Fig. 1. Use-case diagram of the proposed fabric-based drug discovery chain system.

In this proposed scenario, the web application is used by users (contributors of drug discovery data) to submit requests to be onboarded. Onboarded users can perform various actions, such as uploading, updating, verifying, or viewing drug contributions. The admin manages user activities and access rights in the blockchain network. The endorsement policies define access rights to ensure that users only perform actions for which they are authorised. The X.509 digital certificates and public-private keys are issued by the certificate authorities of the various organisations within the blockchain network. Users can only use their digital identities to access the application. The admin provides these certificates to users in order for them to be identified by

the network. The users receive the certificates in response to the onboarding request. After onboarding, users can invoke transactions related to drug discovery data contributions for which they are authorised. The membership service provider verifies each user's identity at the back-end with the help of certificates issued by CA. Only the user presenting a valid certificate is authenticated to invoke the transactions; otherwise, an attempt to initiate the transaction fails.

After a successful authentication check, the endorsing peer validates the user's authorization (access rights) to invoke this transaction. Endorsing peers then simulate the requested transactions against the transaction logic already defined in the chaincode. Transaction blocks

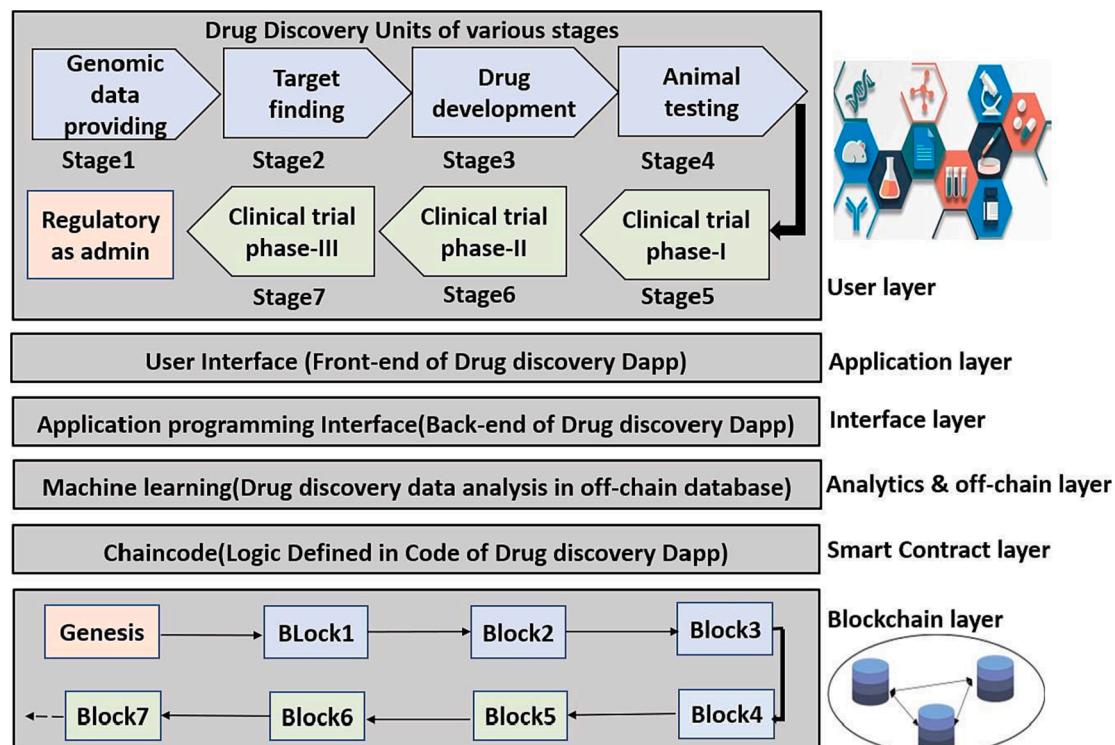


Fig. 2. Layered architecture of the proposed drug discovery chain management system.

are created by the ordering service following receipt of the responses from the endorsing peers. The blocks are then distributed to all peers on the channel within the Hyperledger Fabric network. Finally, users receive the network's request response via the web application after each peer validates the transaction block. In this manner, this proposed blockchain-based system maintains the privacy and security of contributions.

3.3. Layered architecture

The proposed model enables research organisations to upload drug discovery contributions onto the blockchain network and empowers supporting organisations to access and validate them. Fig. 2 illustrates the layered architecture of the proposed drug discovery chain management system, which has six layers, which are as follows:

- **User layer:** The users of the proposed scheme are associated with various drug discovery units such as genomic data, target finding, drug development, animal testing, and clinical trials (phases 1, 2, and 3). The regulatory authority serves as the admin responsible for managing the entire system network.
- **Application layer:** It serves as the front-end of the proposed drug discovery Dapp. It is to empower organisations to upload, update, validate, and view contributions on the blockchain ledger. This layer also facilitates the regulatory authority to issue contribution certificates.
- **Interface layer:** It is the application programming interface. It forms the back-end of a Dapp to connect the web front-end to the blockchain layer and off-chain storage.
- **Analytics layer:** It forms the preprocessing and visualisation mechanism using ML. It collects data uploaded by the genomic data provider unit and analyses the data within off-chain storage.
- **Smart contract layer:** It forms the chaincode layer that defines the logic of the proposed Dapp.
- **Blockchain layer:** It forms the chain of continuously growing blocks (block 1, block 2, etc.) of drug discovery data. The Genesis block is the first block that contains network configuration data.

In the proposed system, the ChEMBL bioactivity data has been used to create ML models. The ChEMBL database has genomic data on more than 2 million compounds. We have searched out the seven target proteins from the genomic database. Then we visualized their quality scores using ML. Fig. 3 shows the quality scores of searched targets using ML.

For plotting quality scores of targets, the Matplotlib library has been used. We have used an off-chain database with a blockchain ledger to

store the complete drug discovery contribution data and stored only the contribution metadata and InCHKey on the blockchain network for validation. This improves the performance of the proposed Dapp.

4. Proposed system implementation

4.1. Development environment

The process of implementation of the proposed scheme is set up on an Ubuntu 64-bit environment (version 20.04.2 LTS) with 12.00 GB of RAM, which is depicted in Table 2.

The Hyperledger Fabric environment has been set up using Docker (v19.03.15), and for the configuration of the Docker containers, docker-compose (v1.24.0) has been used. We have used Fabric (v2.2.0) and Node (v10.19.0) for the fabric-SDK-node setup. To investigate the performance, the Hyperledger Caliper Tool 0.4.2 version has been used. ML code has been written in Python (v3.7.2) and is executed on the Jupyter notebook.

4.2. System assets

The assets of the proposed system are drug discovery contribution records and contribution certificates. The research organisations can create or update contribution assets. Each contribution asset has the following fields:

- Contribution ID: Unique identity for contributions.
- Organisation name: Name of the contributing organisation.
- Contribution Data: Contribution Name.

The research authority can create contribution certificates as an asset. Each contribution certificate asset has the following fields:

Table 2
Details of development environment.

Name	Version
Ubuntu	20.04.2 LTS
Docker-engine	19.03.15
Docker-compose	1.24.0
Hyperledger-fabric	2.2.0
Node	10.19.0
Caliper	0.4.2
Python	3.7.2

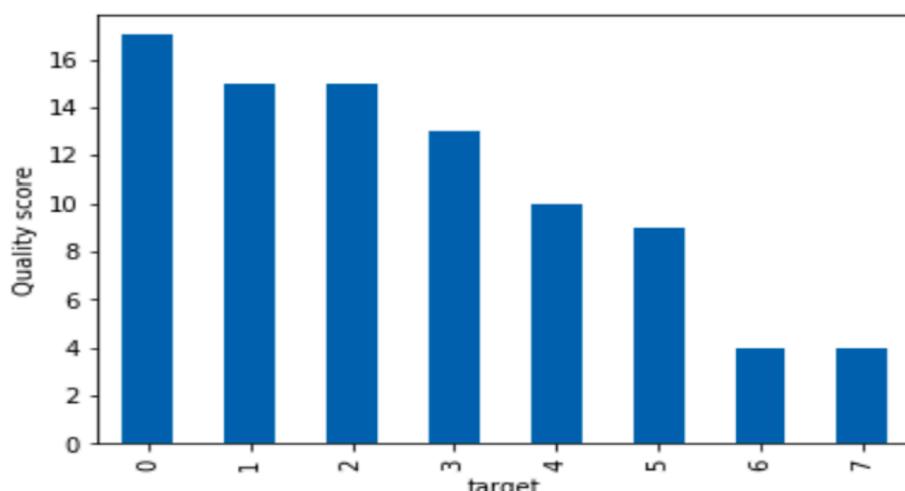


Fig. 3. Illustrates the quality scores of searched targets using ML.

- Certificate ID: A unique identity for contribution certificates. The certificate ID is a composite key that has been made by combining contribution ID and contribution status.
- Contribution status- Contributions approval.
- Owner- Owner of the contribution.
- InChiKey: The unique hash identifier of the contribution. An InChiKey is a fixed 27-character digital representation of the input InChi string, which is obtained by applying the SHA-256 hashing algorithm.

4.3. Chaincode modules

In the proposed model, the drug discovery contribution and certificate assets are created in JSON format, and the chaincodes are written in NodeJS. The proposed system has various chaincode modules to upload, retrieve, verify, and update contributions, issue contribution certificates, and view contribution histories. Contributions include genomic data, target finding data, drug development data, animal testing results, and clinical trial outcomes. Algorithms 1–6 show the workings of these chaincode modules.

Algorithm 1. Algorithm of the chaincode module- "Upload contribution".

```

1: INPUT: Contribution ID, organisation name, contribution data
2: function uploadContribution(ContributionID, organisationName,
   contributionData)
3: if ContributionID! == undefined then
4:   register contributionID
5: else
6:   ContributionID already exist
7: end if
8: end function
9: OUTPUT: Contribution asset
end

```

Algorithm 2. Algorithm of the chaincode module- "Get contribution".

```

1: INPUT: Contribution ID
2: function getContribution(ContributionID)
3: return contribution data
4: end function
5: OUTPUT: Contribution details
End

```

Algorithm 3. Algorithm of the chaincode module- "Issue contribution certificate".

```

1: INPUT: Contribution ID, contribution status, owner, InChiKey
2: function issueContributionCertificate(ContributionID, contributionStatus, owner,
   InChiKey)
3: if valid contributionID then
4:   Issue contribution certificate
5: else
6:   Invalid contributionID
7: end if
8: end function
9: OUTPUT: contribution certificate asset
End

```

Algorithm 4. Algorithm of the chaincode module- "Verify contribution certificate".

```

1: INPUT: Contribution ID, contribution status, current InChiKey
2: function verifyContributionCertificate(ContributionID, contributionStatus,
   currentInChiKey)
3: fetch client identity and save it as verifier identity
4: gets certificate ID details through contributionID and contributionStatus.
5: if certificateID!== undefined || certificate.InChiKey ==currentInChiKey then
6:   Valid contribution
7: else
8:   Invalid Contribution
9: end if

```

(continued)

Algorithm 4. Algorithm of the chaincode module- "Verify contribution certificate".

```

10: end function,
11: OUTPUT: Contribution validation result and verifier identity
End

```

Algorithm 5. Algorithm of the chaincode module- "Update contribution".

```

1: INPUT: Contribution ID, new owner, new contribution data
2: function updateContribution(ContributionID, newOwner, newContributionData)
3: fetch contribution details through contributionID
4: if valid contributionID then
5:   add newOwner field and new Contribution field to the existing contribution
   asset
6: else
7:   Invalid contributionID
8: end if
9: end function
10: OUTPUT- Update the owner and contribution data fields.
end

```

Algorithm 6. Algorithm of the chaincode module- "View contribution history".

```

1: INPUT: Contribution ID, contribution status, current InChiKey
2: function ViewContributionHistory(Contribution ID)
3: return entire history related to contribution ID
4: end function
5: OUTPUT: Contribution ID history
End

```

The functions of the chaincode modules are as follows:

- **uploadContribution()**– This function can be invoked by involved entities to upload a new contribution and store it on the ledger. It consists of the following fields: contribution ID, organisation name, and contribution data.
- **getContribution()**– This function can be used by permissioned entities to retrieve the contribution details. It consists of the following fields: contribution ID.
- **issueContributionCertificate()**– This function can only be invoked by the regulatory authority to issue contribution certificates. It consists of the following fields: contribution ID, contribution status, owner, and InChiKey.
- **verifyContribution()**– This function can be invoked by any entity registered on the network and is used to validate the contributions uploaded by different organisations. It consists of the following fields: contribution ID, contribution status, and InChiKey.
- **updateContribution()**– This function can be called by the current owner of the contribution to update the contribution data or its ownership. It consists of the following fields: contribution ID, new owner, and new contribution data.
- **viewContributionHistory()**– This function can be called by entities to view the entire history of a contribution. It consists of the following fields: contribution ID.

Fig. 4 illustrates the workflow of the proposed drug discovery chain. In the proposed scheme, the genomic data providing unit collects the genomic data, the target finding unit figures out the target molecules, the drug development unit creates a new drug molecule, the animal testing unit performs the task of pre-clinical testing, and the clinical trial units drive various testing phases. The various drug discovery units upload their new contributions to the ledger using uploadContribution(). The regulatory authority can fetch contributions made by various organisations using the getContribution() function. The regulatory authority uses issueContributionCertificate() to generate a unique hash identity for the uploaded drug contributions and issue contribution certificates.

The organisations can use verifyContribution() to check the integrity of an existing contribution before updating it. Contributors can update

(continued on next column)

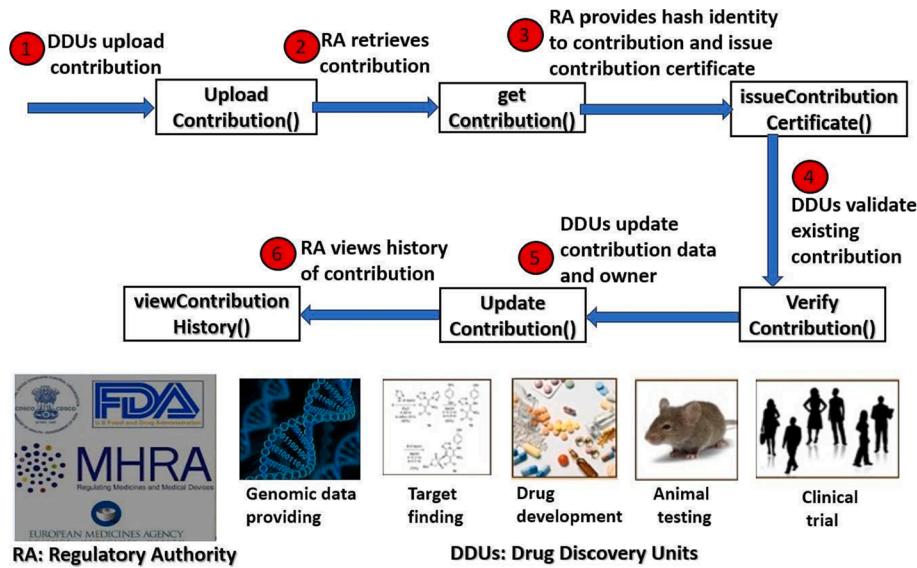


Fig. 4. Workflow of proposed drug discovery chain.

the contribution data field of existing contribution assets by using the `updateContribution()` function. Organisations can also change the owner field in `updateContribution()` to transfer ownership of their contributions. Authorised users can view the entire transaction history of stored contributions on the ledger by calling `viewContributionHistory()`. The regulatory authority can also validate the integrity of the new drug molecule using this function.

4.4. Proposed system architecture in the Hyperledger-Fabric environment

We have used the Hyperledger Fabric blockchain framework for the proposed drug discovery Dapp. By invoking chaincode's upload contribution API, research organisations can add drug discovery contributions at various stages across blockchain networks. The architecture of a proposed Hyperledger-Fabric-based drug discovery network is depicted in Fig. 5. The proposed architecture consists of one channel named "drug discovery channel," one orderer, and three organisations.

The three organisations are named "regulatory," "research and development," and "testing and trial." Each organisation has two peers and one CA. Each peer has a chaincode (CC) named "drug discovery net"

deployed. The various steps needed to submit a transaction proposal are as follows:

- To connect the user with the Fabric blockchain network, the system back-end uses the Fabric-SDK-node tool. Fabric-node-SDK facilitates invoking or querying chaincode transactions.
- To access the functions defined inside the chaincode modules from the web UI, first the node modules for various application functions are converted into APIs. These APIs are the endpoints that the client application accesses using an HTTP request.
- The "Express" library of Node.js is used to create an HTTP node server. A URL has been linked to correspond to every node module API on the server. This is to execute APIs and get a response back. The chaincode APIs have been used by the node modules to perform their respective logic.
- Using this instance of the `FileSystemWallet` class, the certificates needed by the user to identify themselves on the network have been stored in a local file system directory

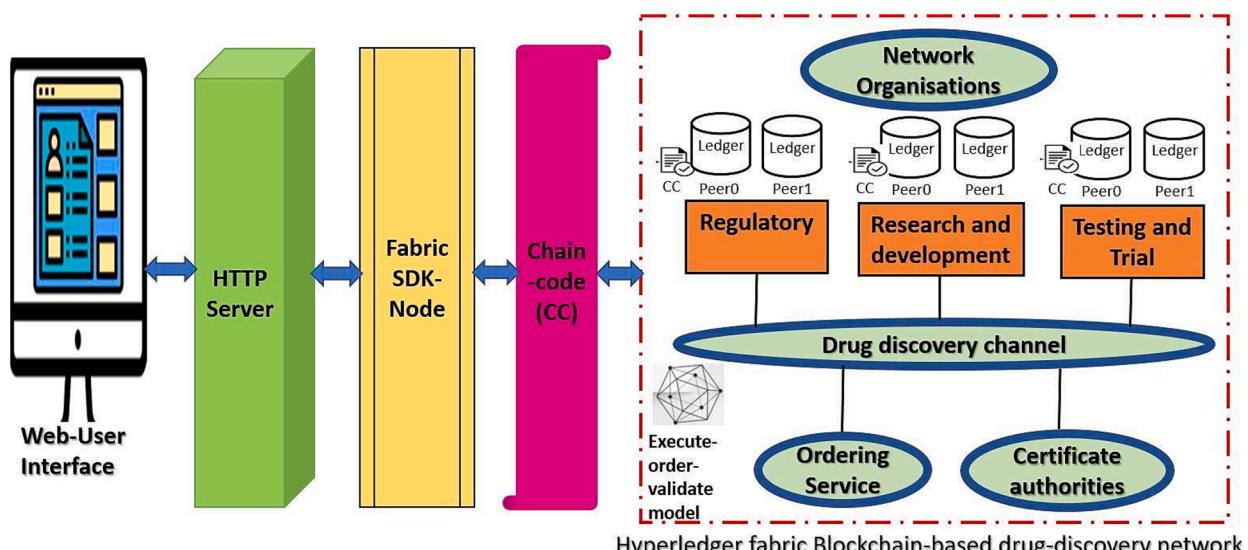


Fig. 5. Architecture of a Hyperledger-Fabric-based drug discovery network.

- The connection parameters are defined inside a single file named the “common connection profile” (CCP), which is a simple YAML file for clients to be able to connect with the fabric network.
- Then an instance of the gateway class is connected to a network using the identity of the user and the CCP.
- After that, the user connects to the drug discovery network using the connect() method of the gateway class.
- Next, the user accesses the drug discovery channel inside the network using the getNetwork() method of the gateway class. Then the user accesses the chaincodes using the getContract() method.
- The next step is to submit the transactions using the createTransaction() method. The response from the chaincode to the server is JSON, and the server sends the HTTP response back to the client.
- Next, the orderer performs the task of inserting the user transaction proposal into the block
- Then the block is forwarded to all peers on the channel within the Hyperledger Fabric network. Each peer validates the received block and updates their ledger with the most recent block.
- Finally, the user receives the request response from the network through the web application. In this way, the proposed Dapp records the data related to drug discovery contributions in the blockchain ledger.

5. Simulation results and performance analysis

This section details the testing outcomes for the proposed drug discovery network configuration, chaincode deployment, application back-end, and web interface. Furthermore, we include performance and comparative analysis.

5.1. Proposed system test results

In the proposed fabric-based drug discovery Dapp, all required entities, including peers, orderers, and certificate authorities, are defined inside the crypto-config.yaml file. Network configurations are defined in the configtx.yaml file. This file contains the capabilities and rights of different services like channels, orderers, and applications. The first configuration file is the genesis block file, the second is the channel configuration file, and the next three are to update the anchor peers in the drug discovery network. We have first generated the crypto-material (X.509 certificates) for all entities using the cryptogen tool. Then, we used the fabric configtxgen tool to create the channel artifacts (genesis block, channel, and anchor peer configuration files). Fig. 6 illustrates the generation of certificates and channel configurations. The genesis block

initiates the ordering service. The different properties defined are orderer type, addresses, batch size, etc. These properties are defined under the orderer profile configuration. Block size is the maximum size limit of a block in the network. Under the genesis block, parameters like batch timeout, maximum transaction count, and absolute and preferred block sizes are defined.

The research and development, regulatory, and testing and trial industries are the peer organisations of the proposed system. We have set up and run the fabric components of the drug discovery network on a local system using Docker. Fig. 7 shows the Docker containers of the proposed drug discovery network with container IDs, image names, time of creation, status, and port numbers. We have created a total of 12 containers with the domain “drug-discovery-network.com”. There are six containers for peer organisations, three for certificate authorities, one for the CLI, one for chaincode, and one for the orderer.

Next, we have created a channel, made all peers join the drug discovery channel, and updated the anchor peers of drug discovery organisations. Then, we logged into the CLI container and installed a chaincode on endorsing peers for testing APIs. Fig. 8 shows the deployment of the drug discovery chaincode.

The proposed Dapp’s front-end interface has been designed to activate the various APIs. The first API that has been used is to add a user’s identity to the wallet. The user can log in using wallet identity to access the various functionalities of the fabric chaincode, such as uploading organisation contributions, viewing contributions, etc. Fig. 9 shows the UI for logging into the drug discovery Dapp.

To access the developed Dapp account, the user must enter the path for the certificate and private key in the input field and click the “login” button. After successfully logging in, the user can create or update network contribution assets.

Fig. 10 depicts the web interface for creating a contribution asset on the blockchain. This figure represents the uploading of genomic data as a new contribution asset. To create a contribution asset, the user must provide details about the contribution, such as the contribution ID, organisation name, and contribution data. Once the user enters the details and clicks the “create contribution account” button, the system sends a contribution creation request to the blockchain network. This activates the “upload contribution” chaincode API on the blockchain. Upon successful response, a new contribution asset is created on the UI with the alert message “Contribution account created.” When this asset is successfully added, the contribution details are displayed on the right side of the UI. The filled-out details are verified on the blockchain network in the back-end of the web application. If all of the contribution fields are not filled out or contain invalid details, the system throws an

```
neetu@neetu-ubuntu:~/worknew/Drug_Discovery_Chain/network$ sudo ./fabricNetwork.sh generate
Generating certs and genesis block for channel 'drugdiscoverychannel' with CLI timeout of '15' seconds and CLI delay of '5' seconds and chaincode version '1.1'
Continue? [Y/n] y
proceeding ...
/home/neetu/worknew/Drug_Discovery_Chain/network/bin/cryptogen
#####
##### Generate certificates using cryptogen tool #####
#####
+ cryptogen generate --config=./crypto-config.yaml
researchanddevelopment.drug-discovery-network.com
regulatory.drug-discovery-network.com
testingandtrial.drug-discovery-network.com
+ res=0
+ set +x
/home/neetu/worknew/Drug_Discovery_Chain/network/bin/configtxgen
#####
##### Generating Orderer Genesis block #####
#####
+ configtxgen -profile OrdererGenesis -channelID testingandtrial-sys-channel -outputBlock ./channel-artifacts/genesis.block
2022-04-15 15:04:33.606 IST [common.tools.configtxgen] main -> INFO 001 Loading configuration
2022-04-15 15:04:33.616 IST [common.tools.configtxgen.localconfig] completeInitialization -> INFO 002 orderer type: solo
2022-04-15 15:04:33.616 IST [common.tools.configtxgen.localconfig] Load -> INFO 003 Loaded configuration: /home/neetu/worknew/Drug_Discovery_Chain/network/configtx.yaml
```

Fig. 6. Illustrates the generation of certificates and channel configurations.

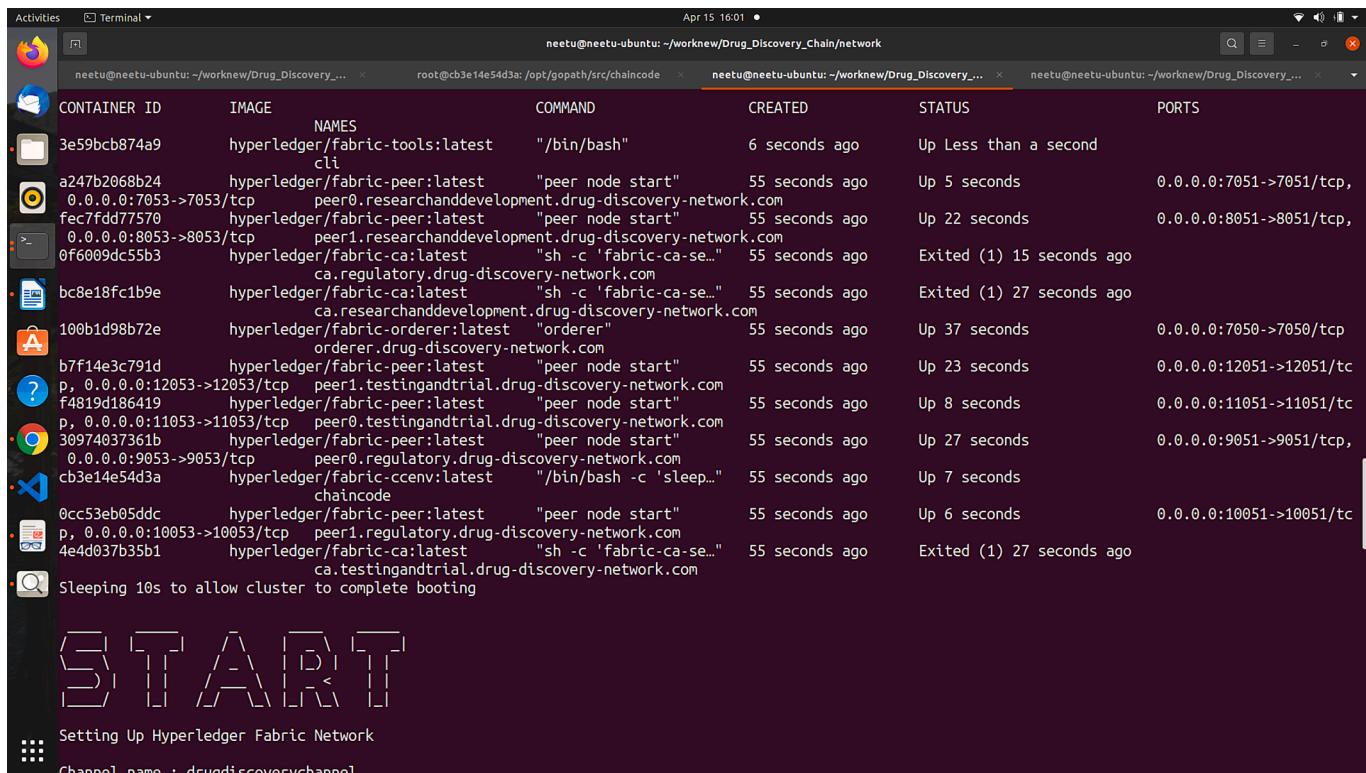


Fig. 7. Docker containers of the proposed drug discovery network.

```

Deploying Chaincode drugdiscoverynet On Drug Discovery Network

Channel name : drugdiscoverychannel
Installing chaincode on peer0.researchanddevelopment.drug-discovery-network.com ...
+ peer chaincode install -n drugdiscoverynet -v 1.1 -l node -p /opt/gopath/src/github.com/hyperledger/fabric/peer/chaincode/
+ res=0
+ set +x
2022-04-15 09:40:22.377 UTC [chaincodeCmd] checkChaincodeCmdParams -> INFO 001 Using default escc
2022-04-15 09:40:22.377 UTC [chaincodeCmd] checkChaincodeCmdParams -> INFO 002 Using default vscc
2022-04-15 09:40:22.449 UTC [chaincodeCmd] install -> INFO 003 Installed remotely response:<status:200 payload:"OK" >
===== Chaincode is installed on peer0.researchanddevelopment =====

Installing chaincode on peer0.regulatory.drug-discovery-network.com ...
+ peer chaincode install -n drugdiscoverynet -v 1.1 -l node -p /opt/gopath/src/github.com/hyperledger/fabric/peer/chaincode/
+ res=0
+ set +x
2022-04-15 09:40:22.557 UTC [chaincodeCmd] checkChaincodeCmdParams -> INFO 001 Using default escc
2022-04-15 09:40:22.557 UTC [chaincodeCmd] checkChaincodeCmdParams -> INFO 002 Using default vscc
2022-04-15 09:40:22.629 UTC [chaincodeCmd] install -> INFO 003 Installed remotely response:<status:200 payload:"OK" >
===== Chaincode is installed on peer0.regulatory =====

```

Fig. 8. Deployment of the drug discovery chaincode.

exception with an error message.

The drug discovery Dapp's back-end, which generates the contribution asset as a response, is depicted in Fig. 11.

The figure demonstrates the creation of the contribution asset. This asset has the following attributes: contribution ID, organisation name, contribution, organisation identity, created at, and updated at. On the back-end of the Dapp, the response from the blockchain-based drug discovery network is in JSON format. Other contributions, such as target molecules, drug molecules, pre-clinical test results, and clinical trial results, can be uploaded in the same way to generate new contribution assets.

Following that, an API for issuing contribution certificates has been

tested, which regulatory authorities can use to issue contribution certificates. Fig. 12 shows the web interface for creating a contribution certificate asset on the blockchain.

The regulatory authority can issue a certificate for the contribution by clicking on the icon in the "Action" column. The system sends a request for a contribution certificate to the blockchain-based drug discovery network. Upon a successful response, a new contribution certificate asset is created on the UI with the message "Certificate issued to contribution". The contribution certificate details are verified on the blockchain network in the application back-end. If the details are invalid, the system throws an exception with an error message. Fig. 13 shows the back-end of the proposed Dapp that creates the response in

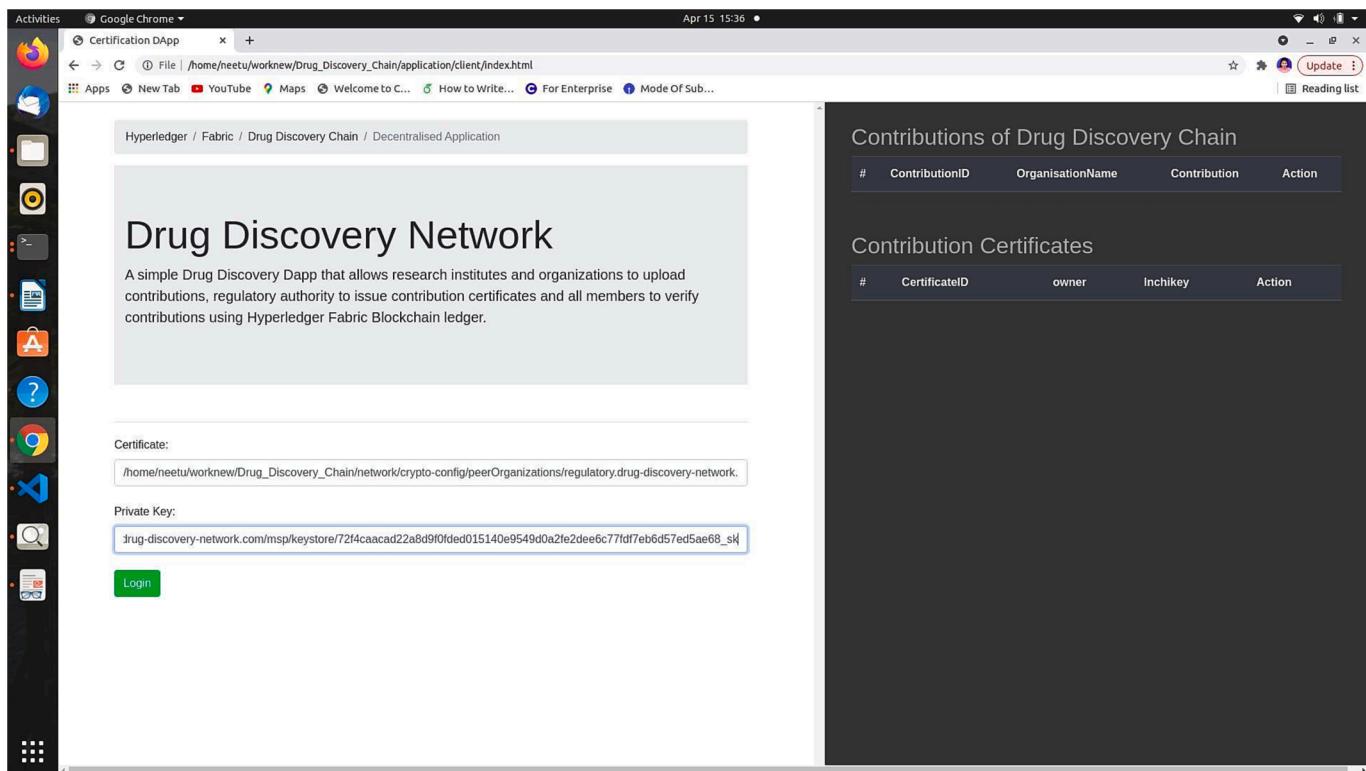


Fig. 9. User-interface to login into the drug discovery Dapp.

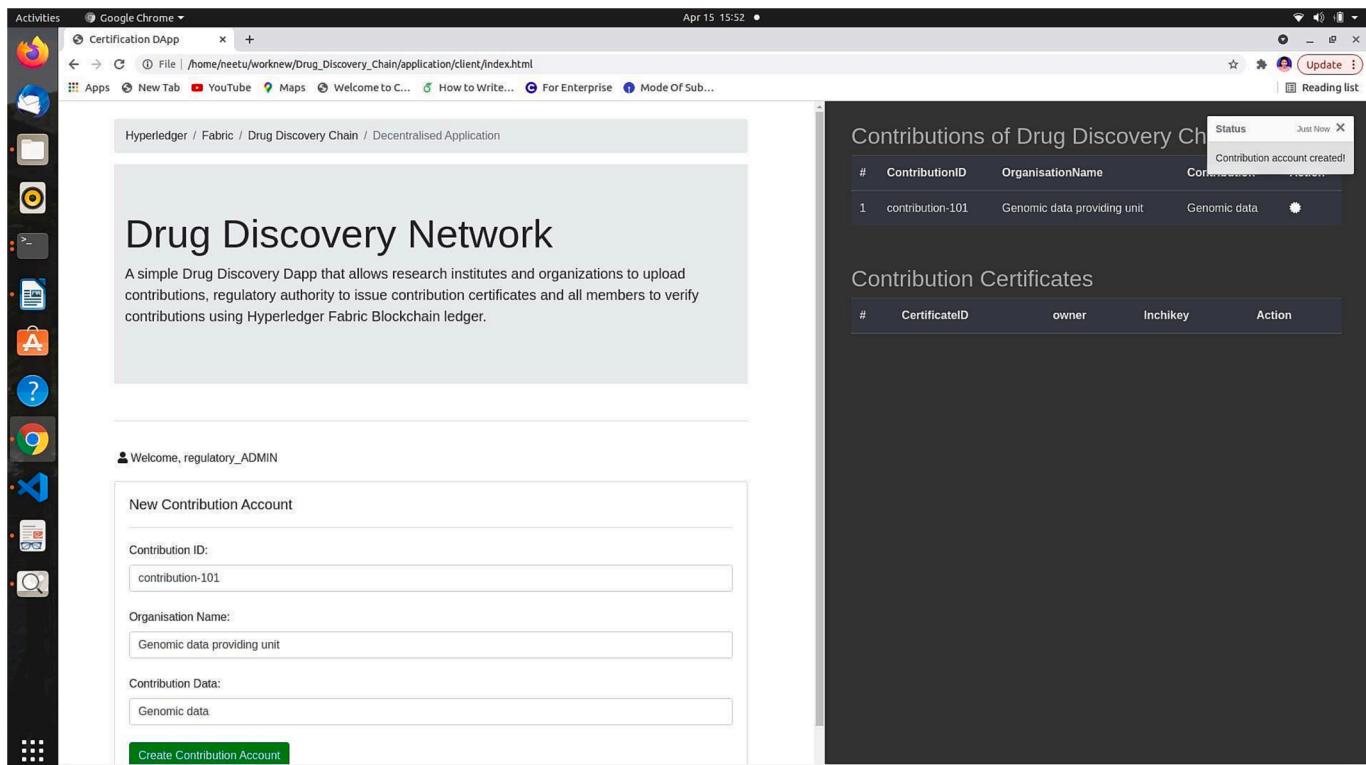


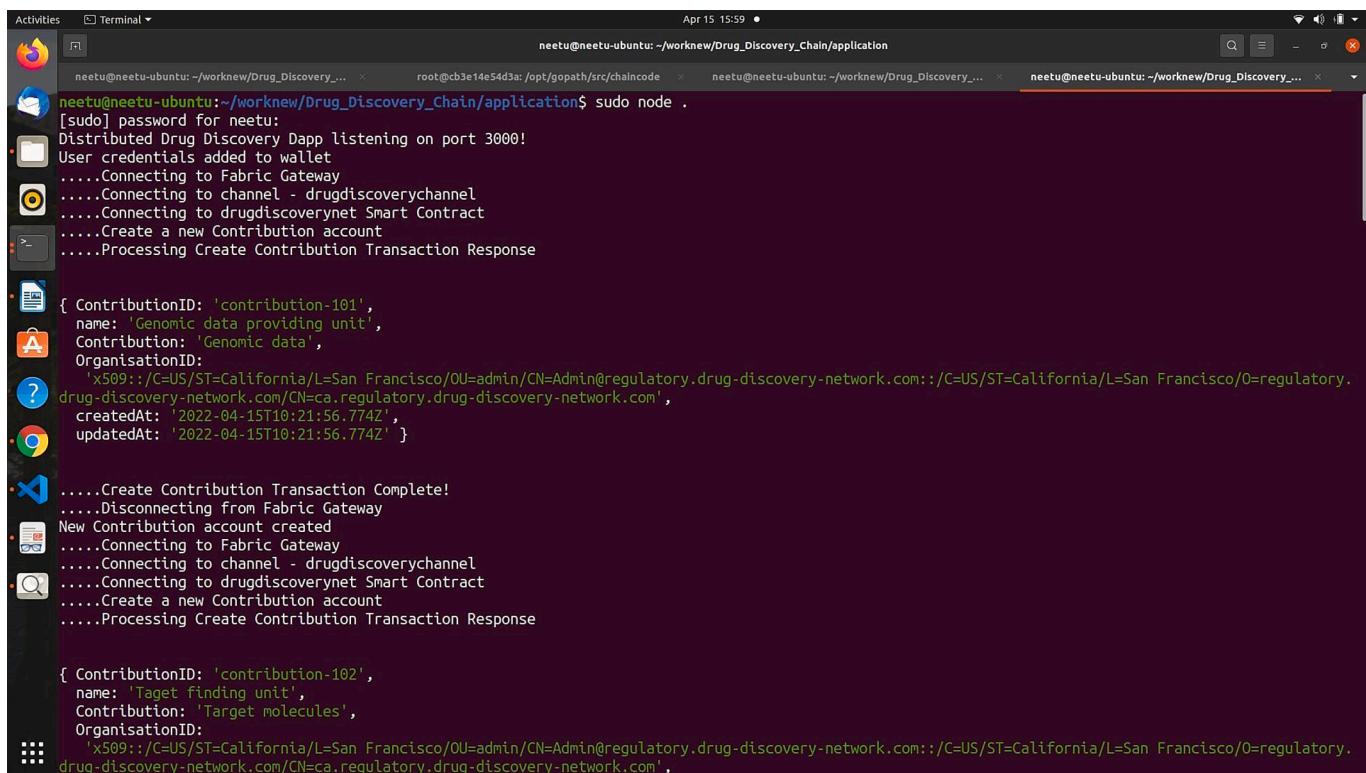
Fig. 10. Web interface to create a contribution asset on the blockchain.

the form of a contribution certificate asset.

This asset has the following attributes: contribution ID, contribution status, regulatory authority, certificate ID, InChiKey, owner, created at, and updated at. After the contribution certificate is issued, it can be

verified using the “verify contribution” chain code API. Fig. 14 depicts the web interface for validating a contribution asset on the blockchain.

In the input field next to the InChiKey field of the certificate, the verifier can enter the current InChiKey of the contribution and test



```

Activities Terminal Apr 15 15:59
neetu@neetu-ubuntu: ~/worknew/Drug_Discovery_... root@cb3e14e54d3a:/opt/gopath/src/chaincode neetu@neetu-ubuntu: ~/worknew/Drug_Discovery_... neetu@neetu-ubuntu: ~/worknew/Drug_Discovery_...
[sudo] password for neetu:
Distributed Drug Discovery Dapp listening on port 3000!
User credentials added to wallet
.....Connecting to Fabric Gateway
.....Connecting to channel - drugdiscoverychannel
.....Connecting to drugdiscoverynet Smart Contract
.....Create a new Contribution account
.....Processing Create Contribution Transaction Response

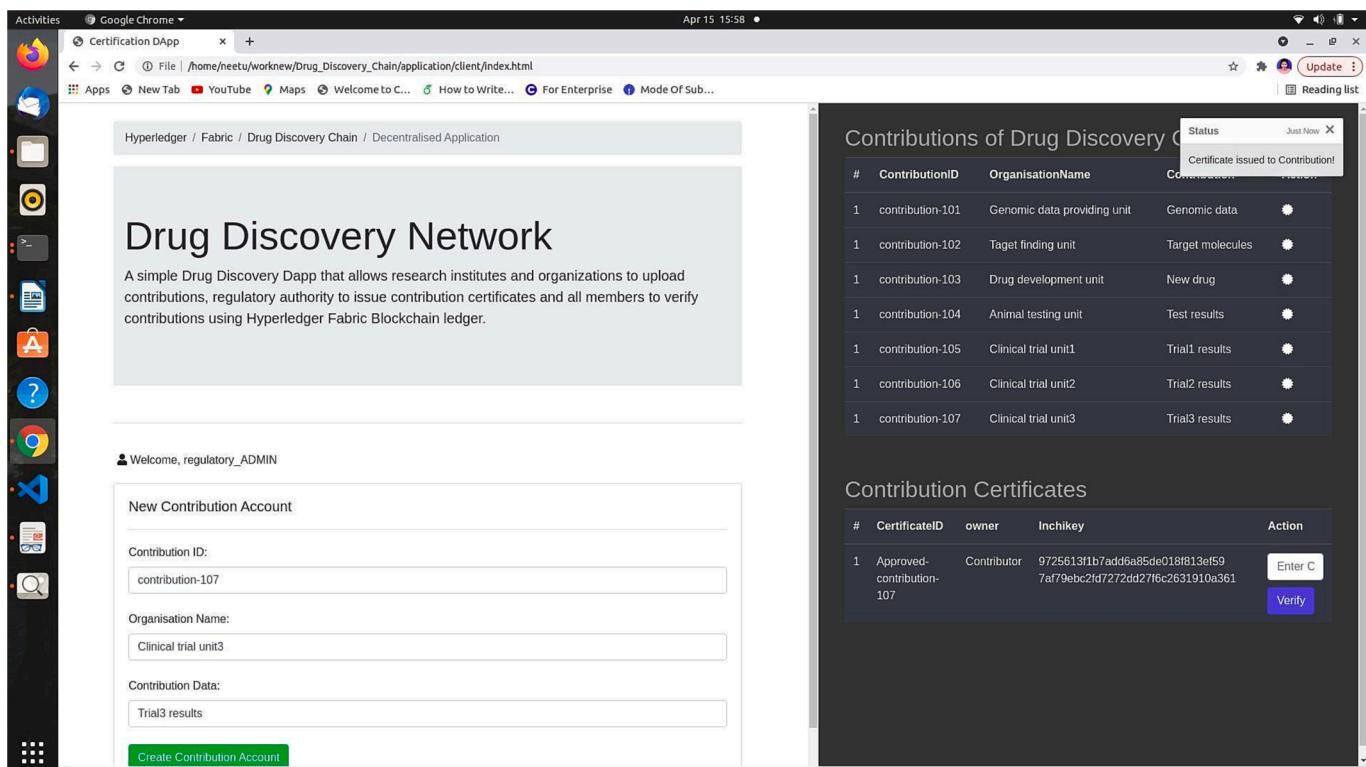
{ ContributionID: 'contribution-101',
  name: 'Genomic data providing unit',
  Contribution: 'Genomic data',
  OrganisationID:
    'x509::/C=US/ST=California/L=San Francisco/OU=admin/CN=Admin@regulatory.drug-discovery-network.com:/C=US/ST=California/L=San Francisco/0=regulatory.drug-discovery-network.com/CN=ca.regulatory.drug-discovery-network.com',
  createdAt: '2022-04-15T10:21:56.774Z',
  updatedAt: '2022-04-15T10:21:56.774Z' }

.....Create Contribution Transaction Complete!
.....Disconnecting from Fabric Gateway
New Contribution account created
.....Connecting to Fabric Gateway
.....Connecting to channel - drugdiscoverychannel
.....Connecting to drugdiscoverynet Smart Contract
.....Create a new Contribution account
.....Processing Create Contribution Transaction Response

{ ContributionID: 'contribution-102',
  name: 'Taget finding unit',
  Contribution: 'Target molecules',
  OrganisationID:
    'x509::/C=US/ST=California/L=San Francisco/OU=admin/CN=Admin@regulatory.drug-discovery-network.com:/C=US/ST=California/L=San Francisco/0=regulatory.drug-discovery-network.com/CN=ca.regulatory.drug-discovery-network.com',
  createdAt: '2022-04-15T10:21:56.774Z',
  updatedAt: '2022-04-15T10:21:56.774Z' }

```

Fig. 11. Back-end of the proposed Dapp that creates a contribution asset.



#	ContributionID	OrganisationName	Action
1	contribution-101	Genomic data providing unit	Genomic data
1	contribution-102	Taget finding unit	Target molecules
1	contribution-103	Drug development unit	New drug
1	contribution-104	Animal testing unit	Test results
1	contribution-105	Clinical trial unit1	Trial1 results
1	contribution-106	Clinical trial unit2	Trial2 results
1	contribution-107	Clinical trial unit3	Trial3 results

#	CertificateID	owner	Inchikey	Action
1	Approved-contribution-107	Contributor	9725613f1b7add6a85de018f813ef597af79ebc2fd7272d27f6c2631910a361	Enter C Verify

Fig. 12. Web interface to create a contribution certificate asset on the blockchain.

whether it matches the InChIKey of this already stored contribution. An alert message of “Invalid” appears when a user enters the incorrect InChIKey into the input field and clicks the “verify” button. When a user enters the correct InChIKey and clicks on the Verify button, an alert

appears indicating that the certificate is valid. The contribution details are verified from the blockchain network in the web application’s back-end, and if the details are invalid, the system throws an exception with an error message. Fig. 15 depicts the back-end of the proposed Dapp that

```

Activities Terminal Apr 15 16:00
neetu@neetu-ubuntu: ~/worknew/Drug_Discovery_... root@cb3e14e54d3a: /opt/gopath/src/chaincode neetu@neetu-ubuntu: ~/worknew/Drug_Discovery_... neetu@neetu-ubuntu: ~/worknew/Drug_Discovery_...
{ ContributionID: 'contribution-107',
  name: 'Clinical trial unit3',
  Contribution: 'Trial3 results',
  OrganisationID:
    'x509:::/C=US/ST=California/L=San Francisco/OU=admin/CN=Admin@regulatory.drug-discovery-network.com:/C=US/ST=California/L=San Francisco/O=regulatory.drug-discovery-network.com/CN=ca.regulatory.drug-discovery-network.com',
  createdAt: '2022-04-15T10:27:42.577Z',
  updatedAt: '2022-04-15T10:27:42.577Z' }

.....Create Contribution Transaction Complete!
.....Disconnecting from Fabric Gateway
New Contribution account created
.....Connecting to Fabric Gateway
.....Connecting to channel - drugdiscoverychannel
.....Connecting to drugdiscoverynet Smart Contract
.....Issue Contribution Certificate
.....Processing Issue contribution Certificate Transaction Response

{ ContributionID: 'contribution-107',
  ContributionStatus: 'Approved',
  RegulatoryAuthority:
    'x509:::/C=US/ST=California/L=San Francisco/OU=admin/CN=Admin@regulatory.drug-discovery-network.com:/C=US/ST=California/L=San Francisco/O=regulatory.drug-discovery-network.com/CN=ca.regulatory.drug-discovery-network.com',
  certId: 'Approved-contribution-107',
  Inchikey:
    '9725613f1b7add6a85de018f813ef59 7af79ebc2fd7272dd27f6c2631910a361',
  owner: 'Contributor',
  createdAt: '2022-04-15T10:28:05.648Z',
  updatedAt: '2022-04-15T10:28:05.648Z' }

.....Issue contribution Certificate Transaction Complete!
.....Disconnecting from Fabric Gateway
New certificate issued to Contribution
.....Connecting to Fabric Gateway

```

Fig. 13. Back-end of the proposed Dapp that creates the contribution certificate asset.

#	ContributionID	OrganisationName	Contrib...
1	contribution-101	Genomic data providing unit	Genomic data
1	contribution-102	Target finding unit	Target molecules
1	contribution-103	Drug development unit	New drug
1	contribution-104	Animal testing unit	Test results
1	contribution-105	Clinical trial unit1	Trial1 results
1	contribution-106	Clinical trial unit2	Trial2 results
1	contribution-107	Clinical trial unit3	Trial3 results

#	CertificateID	owner	Inchikey	Action
1	Approved-contribution-107	Contributor	9725613f1b7add6a85de018f813ef59 7af79ebc2fd7272dd27f6c2631910a361	972561 Verify

Fig. 14. Web interface to validate contribution assets on the blockchain.

generates the response of the “verify contribution” API.

The response of any requested API has the following attributes: certificate, contribution data, verifier identity, result, and verification date. We have successfully built an end-to-end application with a front-

end interface and tested APIs.

```

Activities Terminal • Apr 15 16:00 • neetu@neetu-ubuntu: ~/worknew/Drug_Discovery_... root@cb3e14e54d3a: /opt/gopath/src/chaincode neetu@neetu-ubuntu: ~/worknew/Drug_Discovery_... neetu@neetu-ubuntu: ~/worknew/Drug_Discovery_...
drug-discovery-network.com/CN=ca.regulatory.drug-discovery-network.com', certId: 'Approved-contribution-107', Inchikey: '9725613f1b7add6a85de018f813ef59 7af79ebc2fd7272dd27f6c2631910a361', owner: 'Contributor', createdAt: '2022-04-15T10:28:05.648Z', updatedAt: '2022-04-15T10:28:05.648Z' }

.....Issue contribution Certificate Transaction Complete!
.....Disconnecting from Fabric Gateway
New certificate issued to Contribution
.....Connecting to Fabric Gateway
.....Connecting to channel - drugdiscoverychannel
.....Connecting to drugdiscoverynet Smart Contract
.....Verify Contribution

*** NEW EVENT ***
{ chaincode_id: 'drugdiscoverynet',
  tx_id:
    'e8cb0ea5b0af27ab7cea9d0b16553a13a5e27d57862b11f6195d275e51651f9e',
  event_name: 'verifyContribution',
  payload:
    { certificate: 'Approved-contribution-107',
      Contribution: 'contribution-107',
      verifier:
        'x509::/C=US/ST=California/L=San Francisco/OU=admin/CN=Admin@regulatory.drug-discovery-network.com::/C=US/ST=California/L=San Francisco/O=regulatory.drug-discovery-network.com/CN=ca.regulatory.drug-discovery-network.com',
      result: 'VALID',
      verifiedOn: '2022-04-15T10:28:34.948Z' } }

.....Disconnecting from Fabric Gateway
Validation result available

```

Fig. 15. Back-end of the proposed Dapp that creates the response of the verify contribution API.

5.2. Performance analysis

To investigate the performance of a proposed fabric-based design, the Hyperledger Caliper tool is used. To evaluate the performance of the chaincode, we have measured metrics such as throughput (number of successful transactions per second), latency (transaction confirmation time in seconds), and resource consumption (e.g., CPU, memory, traffic-in, traffic-out). The following are the various strategies used to increase scalability and improve the performance parameters of the proposed system:

- The timeout for a batch is set to 1 s in the network configuration file.
- A high throughput has been achieved using the event-sourcing technique within chaincode. In this technique, the only difference is the asset state stored on the ledger.
- Throughput is further improved by increasing the number of endorsers and block size.

The proposed drug discovery fabric network's performance has been evaluated in terms of throughput, latency, and resource statistics across 1, 6, 11, 16, 21, and 26 organisations. We have measured the performance parameters with 1–2 endorsers to investigate the performance improvement with the number of endorsers. Furthermore, in order to evaluate the performance improvement with the configuration block

parameters, “Max Message Count” (the maximum number of transactions in a block) has been set to 250 to 500, “Absolute Max Bytes” (the absolute block size) to 5 MB to 10 MB, and “Preferred Max Bytes” (the preferred block size) to 1 MB to 2 MB. The “uploadContribution” chaincode API is used for the investigations. The transaction rate is set at 1 TPS (transactions per second), and the transaction duration is set at 30 s.

Fig. 16 illustrates the test results obtained using the caliper tool for invoking the upload contribution API for 21 organisations.

The configuration block parameter “Max Message Count” has been set to 250, “Absolute Max Bytes” to 5 MB, and “Preferred Max Bytes” to 1 MB for this result. The number of successful transactions obtained is 6494, while the number of failed transactions is zero. The obtained send rate is 249.9 TPS, and the obtained throughput is 249.8 TPS, the maximum latency is 0.93 s, the minimum latency is 0.01 s, and the average latency is 0.06 s. The throughput and latency measurements for up to 26 organisations at different block sizes are shown in **Fig. 17**.

The graphs in this figure demonstrate that increasing block size can result in high throughput and low latency. The block size to measure the chaincode performance has been set at 5 MB and 10 MB. The graphs show that throughput increased quickly for six organisations before gradually decreasing for up to 26 organisations. The throughput for a 10 MB block size is significantly higher than for a 5 MB block size. The average latency decreased gradually for six organisations before

Name	Succ	Fail	Send Rate (TPS)	Max Latency (s)	Min Latency (s)	Avg Latency (s)	Throughput (TPS)
uploadOrganisationContribution	6494	0	249.9	0.93	0.01	0.06	249.8

Fig. 16. Illustrates a test result using the caliper tool.

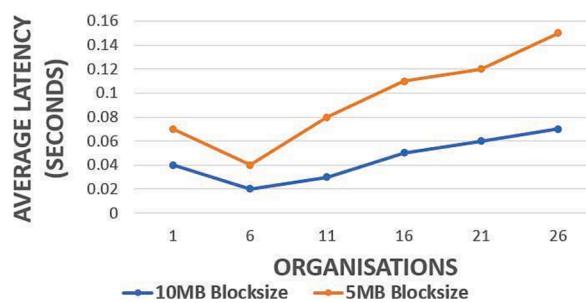


Fig. 17. Throughput and latency at different block sizes.

increasing for up to 26 organisations. The average latency is very low for a 10 MB block size compared to a 5 MB block size.

The throughput and latency measurements with up to 26 organisations at different numbers of endorsers are shown in Fig. 18. This figure demonstrates that increasing the number of endorsers can result in high throughput and low latency. This figure also shows that throughput increased rapidly for six organisations before gradually decreasing for up to 26 organisations. The average latency decreased gradually for six organisations before increasing for up to 26 organisations. The highest throughput that has been achieved is 365.2 TPS at six organisations and two endorsers.

The proposed drug discovery network's resource usage has been assessed in terms of CPU and memory. The resource consumption of the developed system has been measured in terms of CPU-max, CPU-avg, memory-max, and memory-min which are shown in Table 3.

The regulatory organisation has utilised an average CPU resource of 5.84% and a maximum of 16.34%. The maximum and average memory resource statistics measured are 58.8 and 58.7 MB, respectively. The maximum CPU resource stat of the orderer node is 1.54%, and the average CPU resource stat is 0.49%. The maximum value of the memory resource stat of the orderer is 0.817 MB, and the average value of the memory resource stat is 0.131 MB. Hence, the proposed system has low resource consumption.

5.3. Key findings and comparative analysis

The key findings from the performance analysis are as follows:

- Any blockchain framework's performance is deemed satisfactory when throughput is high and latency is low. With six organisations and two endorsers, the proposed system's highest throughput is 365.2 TPS.
- This peak performance is achieved by increasing the maximum number of transactions in a block to 500 and the block size to 2 MB.
- Results show that high throughput and low latency can also be achieved by increasing the number of endorsers.

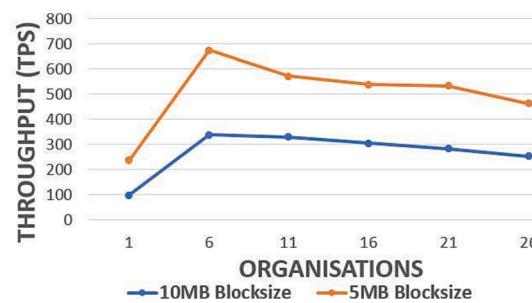


Table 3
Resource statistics of the proposed system.

Name	CPU% (max)	CPU% (avg)	Memory (max)[MB]	Memory (avg)[MB]
peer0.regulatory.drug-discovery-network.com	16.34	5.84	58.8	58.7
orderer.drug-discovery-network.com	1.54	0.49	0.817	0.131

- The performance analysis of (Olsson & Toorani, 2021) shows that this system failed to perform well with more than 16 organisations. The throughput was below 2 and the latency was above 0.04 s for 21 organisations and 1 orderer when the transaction duration was set at 30 s and the transaction rate at 1 TPS. Using the same parameters, the proposed design produced a throughput of 300 TPS and a latency of 0.05 s.
- The throughput and latency of the proposed scheme are much better than (Olsson & Toorani, 2021) with more than 16 organisations confirming that the proposed drug discovery scheme is more scalable.
- The CPU and memory consumption in the developed system have been measured using the caliper tool. The resource statistics show that it has low resource consumption.

Table 4 shows the comparative analysis of the proposed scheme with prior works. There is only one existing implementation paper, and the rest are theoretical. In (Olsson & Toorani, 2021), the collaborative drug discovery model was created using the Hyperledger framework. This design lacked design architecture, chaincode algorithms, analytics, a validation mechanism, and demonstration through a UI. In previous works (Andrews et al., 2015; Ekins & Bunin, 2013), authors proposed architectures for collective drug development and performed some analyses. These designs were not blockchain-based. These works also lacked design algorithms, validation mechanisms, performance testing, and demonstrations. We have designed the architecture, presented algorithms, and created a validation mechanism.

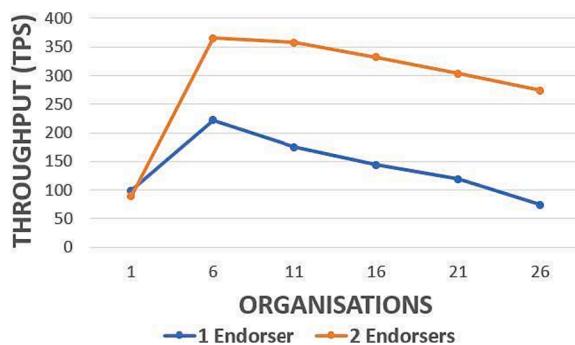
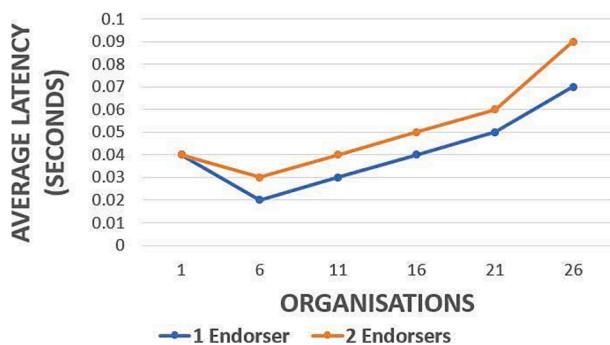


Fig. 18. Throughput and latency at different numbers of endorsers.

Table 4

Comparative result analysis of the proposed work with prior works.

Qualitative metrics	(Olsson & Toorani, 2021)	(Andrews et al., 2015)	(Ekins & Bunin, 2013)	Proposed work
Blockchain-based	✓	✗	✗	✓
Analytics	✗	✓	✓	✓
Architecture	✗	✓	✓	✓
Algorithm	✗	✗	✗	✓
Validation mechanism	✗	✗	✗	✓
Web-user interface	✗	✗	✗	✓
Performance analysis	✓	✗	✗	✓
scalability	✓	✗	✗	✓

✗: Unavailable, ✓: Available.

We have demonstrated the chaincode test results on the front-end web UI and the back-end server. By increasing the number of endorsers and block sizes, we have improved the performance parameters. Scalability is a big challenge to blockchain adoption, but most of the previous works have not addressed it. To improve scalability and reduce storage costs, we have stored only the *meta-data* on the blockchain ledger, and the rest of the data is stored off-chain.

6. Conclusion

Distributed ledger technology has the potential to play a critical role in preserving the integrity and security of drug discovery contributions. This study suggests a drug discovery chain management design using blockchain technology in integration with ML. The proposed decentralised application empowers research organisations to upload drug contributions. It also enables the regulatory authority to assign an identifier to each contribution asset and issue a certificate to prove ownership. The blockchain ledger has been used to store the *meta-data* and InChiKey of drug contributions, and actual contributions are stored in off-chain storage. We have successfully built an end-to-end decentralised drug discovery application, along with the front-end interface and test APIs. The end-to-end application is not available in any previous work. A simple user interface is developed for contributors to invoke transactions easily without having technical knowledge of blockchain. We have also demonstrated the chaincode algorithms for all required functionalities.

The CPU and memory consumption in the developed system have been measured using the caliper tool. The resource statistics show that it consumes low resources. We have used the same tool to investigate the performance metrics in terms of throughput, and latency for up to 26 organisations. With six organisations and two endorsers, the proposed system's highest throughput is 365.2 TPS. This peak performance is achieved by increasing the maximum number of transactions in a block to 500 and the block size to 2 MB.

The performance analysis shows that the existing system failed to perform well with more than 16 organisations. The throughput was below 2 TPS and the latency was above 0.04 s for 21 organisations and 1 orderer when the transaction duration was set at 30 s and the transaction rate at 1 TPS. Using the same parameters, the proposed design produced a throughput of 300 TPS and a latency of 0.05 s. The throughput and latency of the proposed scheme are much better than the existing scheme, with more than 16 organisations confirming that the proposed drug discovery scheme is more scalable.

The performance parameters have been improved and optimised by increasing block sizes and the number of endorsers. This study shows that increased scalability can be achieved by optimising configuration block parameters. Performance and comparative analysis also prove that the proposed blockchain-based design is promising for drug discovery

chain management.

7. Future scope

Future work includes a critical analysis of the proposed solution over a larger number of organisations to further test scalability. The platform should support a large number of participants while maintaining acceptable throughput and latency. Since drug discovery data is highly valuable and the development of new drugs takes a huge amount of time, it is essential to ensure the scalability of the blockchain-based solution. Moreover, within the blockchain, it is not possible to modify data once it has been recorded, and rewriting the data in all of the blocks to make necessary corrections is time-consuming and costly. Hence, the preprocessing techniques need to be improved to ensure that no further data correction is required. Further, the lack of standardisation in blockchain versions creates compatibility issues for developers that need to be addressed.

CRediT authorship contribution statement

Neetu Sharma: Conceptualization, Methodology, Software, Writing – original draft, Visualization, Investigation. **Rajesh Rohilla:** Supervision, Validation, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

- Daizadeh, I. (2021). Has the Covid-19 crisis affected the growth of food and drug administration drug approvals? The answer is not yet! A time series (forecasting) study. *Therapeutic Innovation & Regulatory Science*, 55(3), 553–557.
- Wouters, O. J., McKee, M., & Luyten, J. (2020). Estimated research and development investment needed to bring a new medicine to market, 2009–2018. *Journal of the American Medical Association*, 323(9), 844–853.
- Deore, A. B., Dhumane, J. R., Wagh, R., & Sonawane, R. (2019). The stages of drug discovery and development process. *Asian Journal of Pharmaceutical Research and Development*, 7(6), 62–67.
- Olsson, C., & Toorani, M. (2021). A permissioned blockchain-based system for collaborative drug discovery. *ICISSP*, 121–132.
- Andrews, D. M., Degorce, S. L., Drake, D. J., Gustafsson, M., Higgins, K. M., & Winter, J. J. (2015). Compound passport service: Supporting corporate collection owners in open innovation. *Drug Discovery Today*, 20(10), 1250–1255.
- Hariy, R. E., Barenji, R. V., & Paradkar, A. (2022). Towards pharma 4.0 in clinical trials: A future-orientated perspective. *Drug Discovery Today*, 27(1), 315–325. <https://doi.org/10.1016/j.drudis.2021.09.002>
- Lewis, D. J., & McCallum, J. F. (2020). Utilizing advanced technologies to augment pharmacovigilance systems: Challenges and opportunities. *Therapeutic Innovation & Regulatory Science*, 54(4), 888–899.
- Sharma, N., & Rohilla, R. (2020). Blockchain based approach for managing medical practitioner record: A secured design, in. *International Advanced Computing Conference*, Springer, 73–82. https://doi.org/10.1007/978-981-16-0404-1_6
- Spjuth, O., Frid, J., & Hellander, A. (2021). The machine learning life cycle and the cloud: Implications for drug discovery. *Expert Opinion on Drug Discovery*, 16(9), 1071–1079.
- Radanović, I., & Likić, R. (2018). Opportunities for use of blockchain technology in medicine. *Applied Health Economics and Health Policy*, 16(5), 583–590.
- M. N. Kamel Boulos, J. T. Wilson, K. A. Clauson, Geospatial blockchain: promises, challenges, and scenarios in health and healthcare (2018).
- Omidian, H., & Omidi, Y. (2022). Blockchain in pharmaceutical life cycle management. *Drug Discovery Today*, 27(4), 935–938. <https://doi.org/10.1016/j.drudis.2022.01.018>
- Raj Kumar Reddy, K., Gunasekaran, A., Kalpana, P., Raja Sreedharan, V., & Arvind Kumar, S. (2021). Developing a blockchain framework for the automotive supply chain: A systematic review, 107334 *Computers & Industrial Engineering*, 157. <https://doi.org/10.1016/j.cie.2021.107334>.

- Sharma, N., & Rohilla, R. (2022). Blockchain based electronic health record management system for data integrity. In *Proceedings of International Conference on Computational Intelligence* (pp. 289–297). https://doi.org/10.1007/978-981-16-3802-2_24
- Li, M., Lal, C., Conti, M., & Hu, D. (2020). Lechain: A blockchain-based lawful evidence management scheme for digital forensics. *Future Generation Computer Systems*.
- Jing, N., Liu, Q., & Sugumaran, V. (2021). A blockchain-based code copyright management system, 102518 *Information Processing & Management*, 58(3). <https://doi.org/10.1016/j.ipm.2021.102518>.
- Yang, X., Yi, X., Nepal, S., Kelarev, A., & Han, F. (2020). Blockchain voting: Publicly verifiable online voting protocol without trusted tallying authorities. *Future Generation Computer Systems*, 112, 859–874.
- Ren, Y., Leng, Y., Qi, J., Sharma, P. K., Wang, J., Almakhdem, Z., & Tolba, A. (2020). Multiple cloud storage mechanism based on blockchain in smart homes. *Future Generation Computer Systems*.
- Ammi, M., Alarabi, S., & Benkhelifa, E. (2021). Customized blockchain-based architecture for secure smart home for lightweight IoT, 102482 *Information Processing & Management*, 58(3). <https://doi.org/10.1016/j.ipm.2020.102482>.
- Delgado-Mohatar, O., Sierra-Cámaras, J. M., & Anguiano, E. (2020). Blockchain-based semi-autonomous ransomware. *Future Generation Computer Systems*.
- Huang, X., & Zhang, Y. (2020). Indistinguishability and unextractability of password-based authentication in blockchain. *Future Generation Computer Systems*.
- Bonnah, E., & Shiguang, J. (2020). Decchain: A decentralized security approach in edge computing based on blockchain. *Future Generation Computer Systems*, 113, 363–379.
- Agrawal, T. K., Kumar, V., Pal, R., Wang, L., & Chen, Y. (2021). Blockchain-based framework for supply chain traceability: A case example of textile and clothing industry, 107130 *Computers and Industrial Engineering*, 154. <https://doi.org/10.1016/j.cie.2021.107130>.
- Li, X., Barenji, A. V., Li, Z., Montreuil, B., & Huang, G. Q. (2021). Blockchain-based smart tracking and tracing platform for drug supply chain, 107669 *Computers & Industrial Engineering*, 161. <https://doi.org/10.1016/j.cie.2021.107669>.
- Maity, M., Tolooie, A., Sinha, A. K., & Tiwari, M. K. (2021). Stochastic batch dispersion model to optimize traceability and enhance transparency using blockchain, 107134 *Computers & Industrial Engineering*, 154. <https://doi.org/10.1016/j.cie.2021.107134>.
- Heller, S. R., McNaught, A., Pletnev, I., Stein, S., & Tchekhovskoi, D. (2015). INCHI, the IUPAC international chemical identifier. *Journal of Cheminformatics*, 7(1), 1–34.
- S. Ekins, B. A. Bunin, The collaborative drug discovery (cdd) database, in: *In silico models for drug discovery*, Springer, 2013, pp. 139–154.
- Mubarakali, A. (2020). Healthcare services monitoring in cloud using secure and robust healthcare-based blockchain (SRHB) approach. *Mobile Networks & Applications*.
- Nguyen, T. S. L., Jourjon, G., Potop-Butucaru, M., & Thai, K. L. (2019). Impact of network delays on hyperledger fabric. In *IEEE INFOCOM 2019-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)* (pp. 222–227).
- P. Thakkar, S. Nathan, B. Viswanathan, Performance benchmarking and optimizing hyperledger fabric blockchain platform, in: 2018 IEEE 26th International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS), IEEE, 2018, pp. 264–276.
- Küsters, R., Rausch, D., & Simon, M. (2020). Accountability in a permissioned blockchain: Formal analysis of hyperledger fabric. *Cryptography ePrint Archive*. <https://eprint.iacr.org/2020/386>.
- Yu, H. (2020). Leveraging research failures to accelerate drug discovery and development. *Therapeutic Innovation & Regulatory Science*, 54(4), 788–792.
- Zhou, T., Li, X., & Zhao, H. (2019). Med-ppphis: Blockchain-based personal healthcare information system for national physique monitoring and scientific exercise guiding. *Journal of Medical Systems*, 43(9), 305.
- Cao, S., Zhang, G., Liu, P., Zhang, X., & Neri, F. (2019). Cloud-assisted secure ehealth systems for tamper-proofing ehr via blockchain. *Information Sciences*, 485, 427–440.
- Motohashi, T., Hirano, T., Okumura, K., Kashiyama, M., Ichikawa, D., & Ueno, T. (2019). Secure and scalable mhealth data management using blockchain combined with client hashchain: System design and validation. *Journal of Medical Internet Research*, 21(5), e13385.
- Lee, H.-A., Kung, H.-H., Udayasankaran, J. G., Kijasanayotin, B., Marcelo, A. B., Chao, L. R., & Hsu, C.-Y. (2020). An architecture and management platform for blockchain-based personal health record exchange: Development and usability study. *Journal of Medical Internet Research*, 22(6), e16748.
- Hussien, H. M., Yasin, S. M., Udzir, S., Zaidan, A. A., & Zaidan, B. B. (2019). A systematic review for enabling to develop a blockchain technology in healthcare application: Taxonomy, substantially analysis, motivations, challenges, recommendations and future direction. *Journal of Medical Systems*, 43(10), 320.
- A. H. Mayer, C. A. da Costa, R. d. R. Righi, Electronic health records in a blockchain: A systematic review, *Health Informatics Journal* (2019) 1460458219866350.
- Nguyen, D. C., Pathirana, P. N., Ding, M., & Seneviratne, A. (2019). Blockchain for secure ehrs sharing of mobile cloud based e-health systems. *IEEE Access*, 7, 66792–66806.