

# A Blockchain and NLP Based Electronic Health Record System: Indian Subcontinent Context

Pranab Kumar Bharimalla  
School of Computer Engineering, KIIT University, India  
E-mail: pranab.bharimalla@gmail.com

Hammad Choudhury  
Infosys Ltd., USA  
E-mail: hchoudhury99@gmail.com

Shantipriya Parida  
Idiap Research Institute, Martigny, Switzerland  
E-mail: shantipriya.parida@idiap.ch

Debasish Kumar Mallick and Satya Ranjan Dash (Corresponding Author)  
School of Computer Applications, KIIT University, India  
E-mail: mdebasishkumar@gmail.com, sdashfca@kiit.ac.in

**Keywords:** blockchain, electronic health record, ResNet, CNN, natural language processing

**Received:** April 8, 2021

*The healthcare system in the Indian subcontinent is plagued with numerous issues related to the access, transfer, and storage of patient's medical records. The lack of infrastructure to properly communicate and track records between all key participants has allowed the distribution of counterfeit drugs, dependency on unsafe methods of communication, and lack of trust between patients and providers. During the global COVID-19 pandemic, the need for a robust communication and record tracking system has been further emphasized. To facilitate efficient communication and mitigate the mentioned issues, a nationwide EHR (electronic health record) system must be introduced to bring the healthcare system into digital space. To further enhance security, efficiency, and cost, the innovation of Blockchain is introduced. Blockchain is a decentralized data structure that allows secure transactions between untrusted parties without needing a central authority. In this paper, a Hyperledger fabric-based Blockchain Electronic Healthcare Record (EHR) system is proposed. The system is integrated with technologies such as NLP (Natural Language Processing), and Machine Learning to provide users with practical features.*

*Povzetek: Predstavljen je elektronski zdravstveni zapis na osnovi bločnih in NLP tehnologij v kontekstu Indije.*

## 1 Introduction

The SARS-CoV-2 or commonly known as the Coronavirus pandemic, has challenged healthcare systems worldwide and has exposed vulnerabilities even among the best prepared due to its uncertainty of transmission, the unavailability of a patient's proper medical history, and lack of adequate contact tracing, to name a few examples. Tracking a threat such as this pandemic requires dynamic adaptation of resource deployment to manage rapidly evolving care demands, ideally based on real-time data from a large population sample. Healthcare issues plague all nations regardless of development status due to environmental, economic, or societal conditions; according to The Institute of Medicine (IoM), over one hundred thousand people die each year from preventable medical errors in the US [2]. While healthcare systems in developed nations are in no way perfect as stated by IoM and proven by the pandemic,

developing or underdeveloped regions such as the Indian sub-continent remain more vulnerable. Some common issues that have arisen are due to the lack of communication and tracking infrastructures, such as the influx of counterfeit drugs in the market, dependability on handwritten prescriptions, especially in remote areas lacking any computer systems, and almost nonexistent integration between healthcare and insurance systems. The inadequate infrastructure facilitates the lack of accountability for healthcare providers and further damages relations with patients. Resolving the many health care issues faced in the Indian subcontinent is a formidable challenge but can be significantly improved with the implementation of an end-to-end integrated Electronic Health Record (EHR). In comparison to paper-based record keeping, a practice still utilized in the Indian subcontinent, EHR has clear and distinct advantages [7]. It is fair to say that EHR holds a lot of promise: lower morbidity and mortality rates, better continuity of

care, increased efficiencies, fewer adverse drug reactions, and, most importantly, lower healthcare costs. Paper-based records are more susceptible to human error due to fundamental factors such as legibility or loss of the physical item, causing a delay in treatment and possible fatality, which could have been prevented [18]. Also, the impact of EHR will undoubtedly be felt during the coming months as the global effort to distribute and administer the Coronavirus vaccine intensifies.

An ideal healthcare system should be open, affordable, innovative, and secure. Care costs should not hamper patients from receiving required treatments or buying medicines, and healthcare should be affordable to all regardless of wealth [19]. In the current scenario, an ideal healthcare system might seem far-fetched, but there is always room for improvement using the latest innovations [13]. As our digital infrastructure evolves, the need for robust privacy is increasing. Because of the sensitivity of an individual's healthcare information, a breach in the system could jeopardize the identity of a patient and the reputation of providers. According to [10], the patient's data contains information that is highly prized by cybercriminals. Few noticeable hacks of medical information such as AMCA Data Security Incident (approx. 25M records) and Anthem data breaches (approx. 80M records) caused enormous damage to the medical system [20]. Blockchain technology could help secure and protect sensitive patient information and is emerging as an alternative to the conventional way to log transactions and transfer data through a trusted intermediary to provide validity to the transaction [3].

So far, we have discussed various healthcare issues prevailing in the Indian subcontinent, such as the lack of complete end-to-end EHR interlinking between individuals, stand-alone hospital recording systems, extensive use of handwritten prescriptions, unsafe hospital databases vulnerable to data manipulation by hospital authorities without patient permission and lack of accommodation for caregivers. We also understood how a Blockchain-based EHR system can be a perfect solution to these problems. This article proposes a patient-centered blockchain-based EHR system that identifies patients using their national ID, grants individual control over their health records, protects against unauthorized data manipulation, and allows scalability. Cloud integration stores large CT scans and X-ray reports. An EHR patient-centered system must accurately identify individuals using unique identifiers. There is no end-to-end EHR system with national ID cards like Aadhar or PAN (Permanent Account Number). Integrating unique identifiers like national IDs is essential to successfully identify patients. The proposed scalable system will enable existing EHR and healthcare databases to be integrated along with other useful features, including a mobile app-based interface to convert paper prescriptions to text using Natural Language Processing (NLP) algorithms to bring old paper-based medical records into the new system.

The paper is organized as follows: Section 2 briefly de-

scribes some related work and their major contributions as well as research gaps. The proposed system architecture is presented in Section 3, followed by the proposed algorithms. Section 4 discusses the implementation strategy as well as the analysis of the results. Finally, Section 5 provides the conclusion and suggestions for future research.

## 2 Related work

With the advent of different Blockchain platforms like Hyperledger Fabric, Ethereum, and Azure Blockchain Workbench, many patients centric, permission-based Blockchain EHR schemes have been proposed in the literature. [17] proposed a Permission-based EHR sharing system intending to enhance security and privacy. They also proposed a design access control policy algorithm with a smart contract and formulated a performance optimization mechanism of the system. However, this work ignores the reusability of existing healthcare records sitting in individual hospital databases. [8] proposed a Blockchain-based permissioned EHR system with the capability of data integration of local standalone EHR systems that house in different hospitals or clinics. The framework proposed to store metadata and access only in Blockchain whereas actual health records in the cloud. This is a novel concept, but actual patient-sensitive EHR data resides in the cloud and does not enjoy the immutability that Blockchain offers. [15] tried to enhance the framework proposed in [17]; and added few new modules like a chemist, insurance, and doctor's appointment. Even though the authors formulated a comprehensive approach, the work ignores the scalability of data and interoperability of existing EHR and healthcare databases. [6] proposed a similar Blockchain-based searchable encryption scheme for EHR that not only brings convenience to patients, healthcare providers but also to researchers. In the proposed system only, indexes are added to the Blockchain, whereas actual patient information is stored in an encrypted format in the cloud. There are different ways of encrypting data in the cloud and achieving privacy preservations. [9] defined a novel way of splitting EHR records into sub-messages and finally construct shares of EHR to store in different computer nodes locally and upload the indexes in healthcare Blockchain. [3, 11] evaluated the performance for common public and private/consortium Blockchain-based healthcare systems using metrics such as memory consumption, disk write and read performance, network data utilization, transaction execution per unit time, and CPU usage with consortium-based systems yielding the best performance results. [4] proposed a framework called Blockchain-Based Deep-Learning as a-Service. The framework shares EHR records among multiple healthcare users and operates in two phases to prevent collusion attacks through authentication and predicts possible future conditions for patients through deep learning. [5] proposed a Blockchain-based architecture that allows access to the database based on user

roles and enhances the traditional encryption system employing the Quantum blind signature to protect the system from quantum attacks using hyperledger fabric.

On the other hand, [14] proposed a system that uses AES cryptography to perform the cryptic operation and block chaining it through the hash keys. In addition to that, their proposed healthcare ecosystem includes a prediction model to diagnose the disease of the patient with the deep learning algorithm. In addition to the related works mentioned above, we also performed literature surveys on the ANN-based text extraction model, which could contribute to our work. [1] have implemented the Artificial Neural Network (ANN) approach for text extraction from 64 different types of prescriptions with 98% accuracy. [12] have proposed a CNN common approach for extracting numeric text using handwritten numbers commonly found in India, including Odiya, Telugu, Devnagari, Bangla, and English. The Bangla characters were 95% accurate, Devanagari characters were 98.54% accurate, Odia characters were 97.2% accurate, Telugu characters were 96.5% accurate, and English characters were 99.10% accurate.

### 3 Proposed model

The proposed system prototype is based on Hyperledger Fabric, an open-source distributed ledger technology built to meet enterprise requirements. It is a widely used private Blockchain option. Being a permissioned platform with improved configurability and modularity using pluggable consensus protocols, it is ideal for a range of industries requiring a level of trust between known participants in a governance model. For a transaction to occur, the user must be admitted with the organization's certificate authority, received the means for network authentication, the chain code must be deployed to the channel and installed on the peers, and both parties must have agreed to the endorsement policy. A transaction proposal using an SDK is constructed along with a signature to call a chain code function with parameters to update the ledger. After approvals from peers are received, the chain code is executed against the current database, and the response, read set, and write set are received; these values and the signature are sent back to the SDK to be parsed and consumed by the application. Once the application has validated the responses, the proposal and response are bounded in a transaction message for the ordering service; the ordering service creates blocks of transactions per channel. The transactions are sent to all peers in the channel, the peers complete a final validation, the ledger is updated, and the peer reports to the client about transaction validation or invalidation.

#### 3.1 System architecture

In the proposed patient-centric, Blockchain-based healthcare system, there are seven key participants, the patients or the public, doctors or caregivers, pharmacies, labs, insurance company staffs, government institutions, and the

admin user. Assistance from the government is necessary to implement a functional Blockchain-based EHR system. The government provides credibility to insurance companies, labs, pharmacies, doctors, and even patients through national identification numbers. The proposed system provides patients substantial control over their medical records, including the right to read, write, authorize, and revoke records in the Hyperledger Fabric Blockchain network. Doctors work closely with patients to diagnose conditions, plan treatments, and prescribe medication. Pharmacies work in parallel with doctors to distribute medications to patients. Due to the many factors related to an accurate diagnosis, labs are specialized in detecting distinct conditions with the help of specific tools and trained professionals. Insurance companies help share risk between a large population, making the cost of healthcare affordable for the public, especially during unexpected events such as accidents. Doctors, pharmacies, labs, and insurance companies can read and update the patient's medical record in the Hyperledger Fabric Blockchain network if access has been provided. The admin is critical for system maintenance, and they have unrestricted access to the system, including the right to read, write, update, remove and grant access to participants in the Hyperledger Fabric Blockchain network. The admin's enrollment certificate is obtained from the certification authority. Implementing a national health portal or EHR system is not possible without the government's involvement in India due to the sizable population of 1.4 billion. In the proposed system, the government institution acts as a founder organization or a trusted anchor that can provide credibility to hospitals, pharmacies, insurance companies, labs, and other institutions and provide them with trusted roles to play. The participants with trusted roles can create and issue credential schema and definition to the public or patients

The participants could register through a client application or SDK and request an enrollment certificate from a Membership Service Provider (MSP) to the certificate authority. An MSP allows peers to validate incoming transactions and sign off endorsements. After receiving the enrollment request, the certificate authority issues the certificate and private key with a new ID to enroll the participant. The Hyperledger Fabric Blockchain network distributes all transactions. Participants such as doctors, pharmacies, labs, insurance companies, etc., have different roles in the system and are only granted access when authorized. In the proposed system, an individual's identity could be verified against a national identification number such as Aadhar and be structured to contain identity information like names, date of birth, gender, and other identifying information which could be fetched from the Aadhar database. Patients can use the client application to update details like blood group, allergies, medications, insurance details, etc. Once the transaction is submitted, it will be broadcasted to the network. Endorsing peers will verify the transaction and authenticate using their certification and private key. The transaction will next go to the Orderer through

the SDK client. The Orderer creates a block and sorts the blocks based on different ordering algorithms (viz crash fault-tolerant) and broadcasts to the network peers. All the committing peers validate the blocks once again and check if it is from the correct Orderer and validate conflicts before committing. MS is the body that manages the network identities of organizations and users; however, it does not have access to medical records on the Blockchain network. The MS verifies participants based on TIN/PAN before enrolling them within the network. CouchDB, a NoSQL database that stores data in JSON-based format, is a popular database option used alongside Hyperledger Fabric. A network is comprised of peer groups that hold ledgers and smart contracts used to encapsulate shared network processes and information. In Hyperledger Fabric, transactions produced by smart contracts are contained in a chain code. The key components and processes of the proposed system are highlighted in the below architecture Figure 1. The algorithm 1 explains enrollment of patients whereas algorithm 2 refers to hospital, pharmacy, etc. addition to the network. Table 1 details the abbreviations used in both the algorithms.

### 3.2 Data pulling and sharing

In a permissioned blockchain system, the patients have the authority to decide who can read or update their records; however, the admin reserves the right to grant access to an institution in case of an emergency. Individual hospitals will integrate their EHR system with a Blockchain node and a web API that has full access to their local EHR to convert any existing SQL records to No SQL format for storage within Blockchain. A hybrid data management approach is utilized to facilitate EHR data scalability where all key patient information, including demographics, allergies, medications, and access controls, are stored in Blockchain, and sensitive medical files such as x-ray and scanning reports are stored in private cloud storage using encryption in Figure2.

In the proposed system, a patient can provide access to institutions and participants, including hospitals, doctors, insurance companies, and pharmacies, through the user interface using the web or mobile app. The patient will need to identify the participant requiring access, the category of data to be shared, and the period until the data is accessible. If a patient has visited a particular institution in the past and there are medical records contained in the local EHR system not found on the Blockchain network, they can initiate a pull request using web apps. Once the pull request is approved by the institution's admin user, the web apps will connect to the local EHR system to fetch relevant data, insert patient information into the Blockchain network and upload large files to the private cloud in an encrypted format.

### 3.3 Patient data management at hospital

During a routine or emergency hospital visit, the patient provides the hospital access to their medical records on the Blockchain network to check and amend their records based on the latest assessment. The Hospital must have a valid node on the Blockchain network and request keys from the network admin to permit the login. The patient will select the category of data to be shared and how long the Hospital will have access to the records. Once the Hospital has been provided access, they can read and update the records for an individual using their Aadhar identification card described in Figure3.

### 3.4 Patient data management at pharmacy, patho lab and insurance firm

During a pharmacy visit to fulfill a prescription, the patient provides the pharmacy access to their medical records on the Blockchain network, assuming the pharmacy has a valid node on the network and has requested keys from the network admin to enable login. The patient can provide itemized permissions where the pharmacy will only have access to select prescriptions through private datasets and control how long the pharmacy has access to the records. Once the pharmacy has been provided access, they can read and update the records for an individual using their Aadhar identification card.

During a lab visit at a specialized facility to assess specific conditions, the patient provides the lab access to their medical records on the Blockchain network. The lab requires a valid node on the network and must have requested keys from the network admin to allow login. The patient controls what data is to be shared and how long the lab will have access. Once the lab has been provided access, they can read and update the records for an individual using their Aadhar identification card. Large lab files such as x-ray reports can be encrypted and uploaded to private cloud storage. Similarly, Insurance firms need to update a patient's insurance and policy information for policy procurement or medical expense claims. The patient provides the insurance firm access to their medical records on the Blockchain network if the institution has a valid node on the network and has requested keys from the network admin to authorize a login. The patient will select the type of data to be shared and how long the insurance firm will have access to the record. Once access has been provided, they can read and update the records for an individual using their Aadhar identification card.

### 3.5 Patient uploads old handwritten/ printed prescriptions and bills

A core proposal of this paper is to convert paper prescriptions to text using Natural Language Processing (NLP) algorithms to bring old paper-based medical records into the new system through a mobile app-based interface. In the

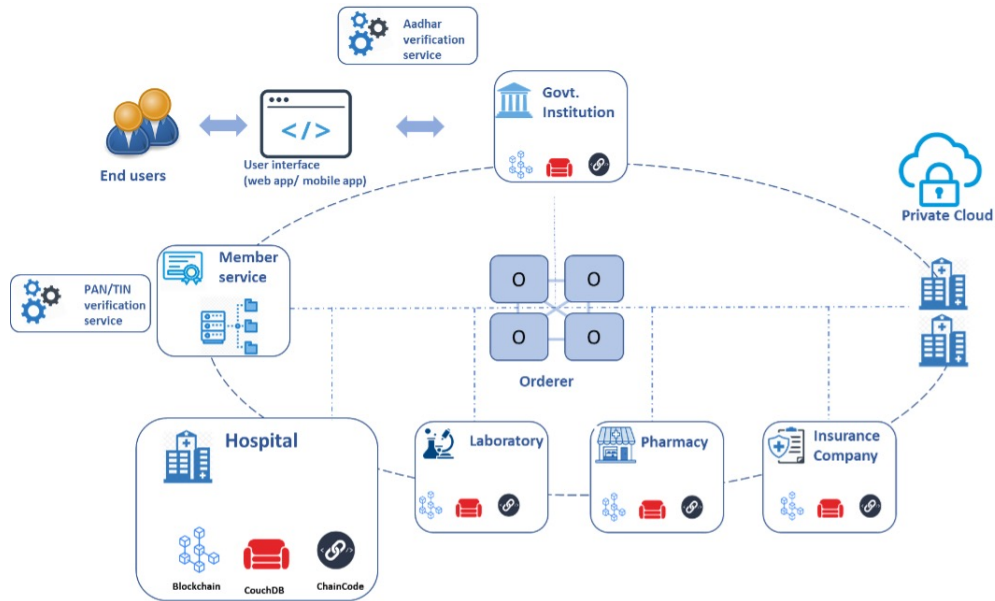


Figure 1: Proposed architectural Framework for Blockchain-based Healthcare System. It depicts key participants, components and transaction processes involved.

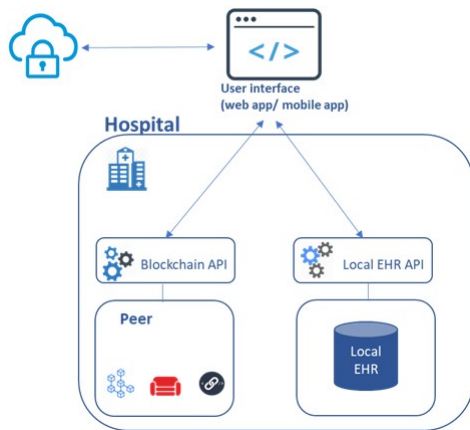


Figure 2: Secured Hospital data management and data pulling process from local EHR.

below module in Figure 4, the system flow for NLP-based data extraction is presented, highlighting the key components.

**3.5.1 Handwritten prescription data extraction**

**Convolutional Neural Network (CNN)** CNN is a Neural Network (NN) that performs convolutional operations instead of simple matrix multiplication operations; it is one of the layers. The structure of CNN consists of a Convolution layer, pooling layer, and fully connected layer. Feature extraction operations are done in the convolution layer, and the output of it is passed to the activation function. The size of output reduces by pooling layers and gives robust learn-

ing results for input data. By performing the convolution layer and pooling layer multiple times, global features can be obtained. In the end, extracted features are passed to the fully connected layer for regression and classification.

**Residual Network (ResNet)** According to the image processing research, the number of layers or depth of a network is crucial for the performance of a model, but the greater number of layers is responsible for the degradation. Much research shows those types of degradation are not caused by overfitting but due to the matter of optimization. The ResNet can solve the degradation problem by introducing a residual framework.

**Long Short-Term Memory (LSTM)** RNN is also a Neural Network that is specially designed for the processing of sequential data. In time-series data, the output  $t - 1$ -time step affects the decision of future time step  $t$ . So, RNN is not able to solve long sequence data. i.e., is called a vanishing or exploding problem. LSTM was designed to resolve this issue of vanishing or exploding problems. LSTM has an internal memory cell called a cell state. This gets the previous output and determines which element should be updated, erased, and maintained in the internal state vector. These processes are handled by four gates, forget gate  $f_t$ , output gate  $o_t$ , input gate  $g_t$ . which are shown in Figure 5.

The proposed model consists of Resnet-50 and three LSTM layers. To do non-linearity, Rectified Linear Unit (RELU) activation function is used in every convolutional layer. The images are divided into 28 sub-windows, so the image height is equal to the height of text-line images. The vector map will be produced by the last layer of convolution. The output of the last convolution fed into the first

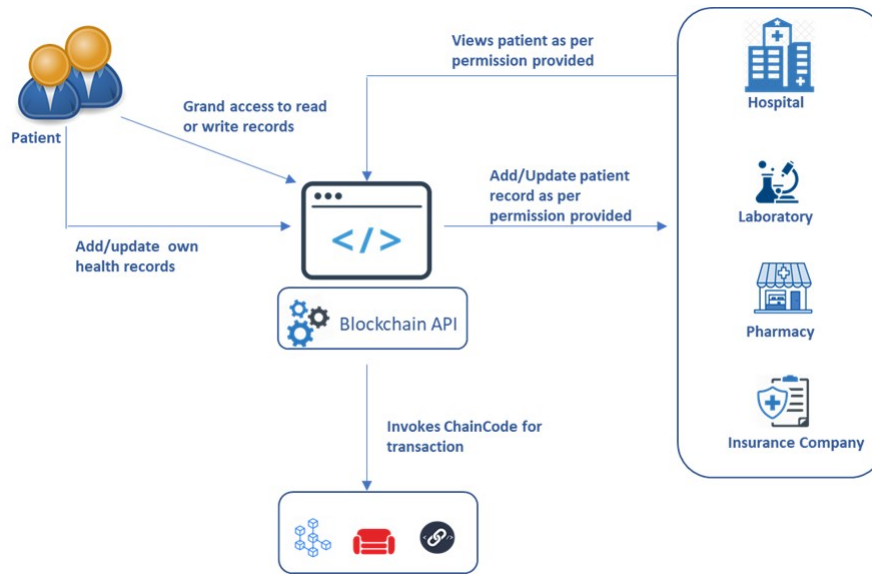


Figure 3: High-level access flow diagram: The patient can add own records to the Blockchain or grants access to other participants like Hospital, Laboratory, Pharmacy or Insurance company to add/update records.

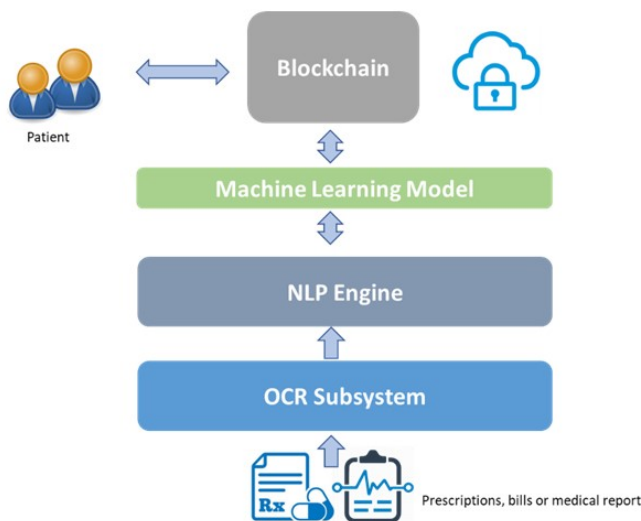


Figure 4: NLP based data extraction flow showing the components.

LSTM layer. By doing LSTM operations, the weights are optimized. Initially, 0.001 is used as the rate of learning.

### 3.5.2 Printed prescription data extraction

For extracting printed prescription data, we have used the Google tesseract model,[16] a pre-trained character made by Google. For the extraction process, we have done some preprocesses. For a more efficient model, first, we convert the image into grayscale. After the greyscale operation, the Otsu thresholding is next, where pixels are converted into zeros and ones. During thresholding, some of the pixels

#### Algorithm 1: Patient Enrollment.

---

**Input:** Enrollment Certificate ( $E_C$ ) from Certification Authority ( $C_A$ )  
**Output:** Successful registration of patient  
**Initialization:**  $N_{Admin}$  should be valid node.  
 $N_{Admin}$  can Write/Read/Remove/Update patients;  
**while** (*true*) **do**  
    **if**  $P_{AId}$  is valid **and**  $FetchAadharRecord(P_{AId})$   
        **not null then**  
             $P_{Rec} \leftarrow FetchAadharRecord(P_{AId})$   
            Add Patient ( $B_{NT}, P_{AId}$ )  
            Grant Access( $P_{AId}$ )  
            Create Record ( $P_{AId}, P_{Rec}, B_{NT}$ )  
        **else**  
            Invalid( $P_{AId}$ )  
        **end**  
    bool  $chk \leftarrow (0 : malicious, 1 : genuine)$   
    **if**  $!(behaviour(chk))$  **then**  
        | Remove Update ( $P_{AId}$ )  
    **else**  
        | Add Update ( $P_{AId}$ )  
    **end**  
**end**

---

may be lost. To restore those pixels, Erosion and Dilation operations are performed where Erosion expands some pixels, and Dilation shrinks some pixels as shown in Figure 6a, Figure 6b and Figure 7.

After these images complete preprocessing, the image will be sent to Google-Tesseract. The tesseract will extract all text from the image and send it to the network.

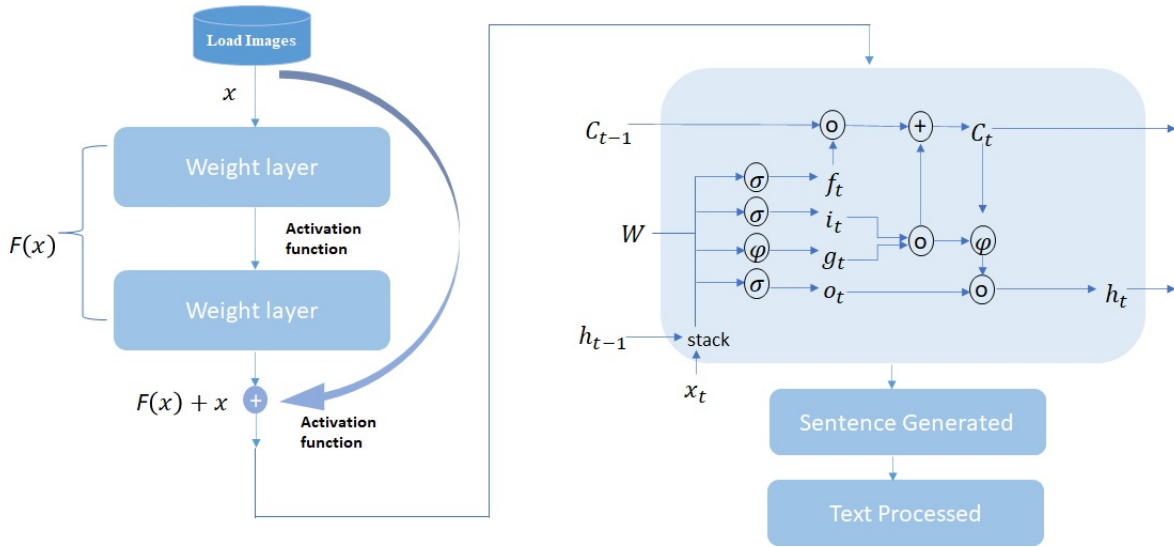


Figure 5: Resnet-LSTM approach for Text Extractions.

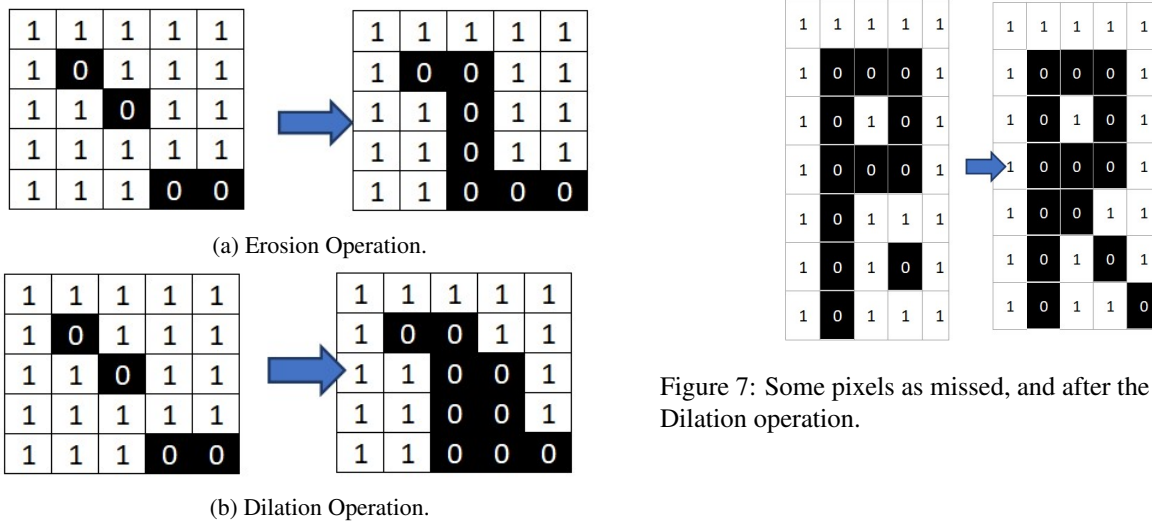


Figure 6: Pixel restoration using Erosion and Dilation operations.

## 4 Implementation and result analysis

### 4.1 Blockchain network setup

To realize the proposed architecture, hyperledger fabric and Sandbox are utilized. Hyperledger is authentication and distributed ledger-based platform. It is an open-source technology used to implement different smart contracts with constraints and logic over the network for applications. The smart contracts are implemented over the network using the sandbox module. In Sandbox, the participants are known, and the Blockchain is in the permissioned consortium mode, making it a secure and trusted Blockchain. The proposed architecture is not limited to

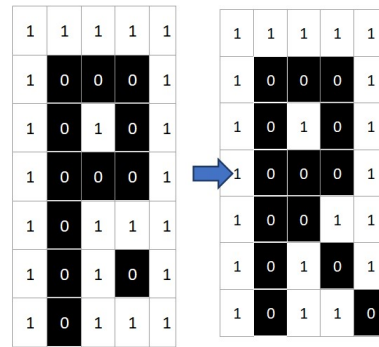


Figure 7: Some pixels as missed, and after the Erosion and Dilation operation.

the healthcare domain. Programming languages such as Node.js, Java, Go, etc., are used for contract and business network development. Docker is used for setting up and initialization when working with hyperledger fabric and composer. Docker is an operating system-level container used by developers, system administrators, etc., for creating, deploying, and running business networks or hyperledger-based applications in a container, enabling the dependencies and functionalities to be packaged together. The hyperledger fabric and composer network can run inside a container using Docker.

In our simulation phase, we used a network model of 3 organizations with 2 peers each and one Orderer. The experiment is carried out with basic writing transactions at various rates, with 1000 transactions per round at 50, 100, 150, 200, and 250 transactions per second. The experiment is done for 1 Org 1 Peer, 2 Org 2 Peer, and 3 Org 3 Peer with different performances of transactions. The results are calculated over five rounds, with each round consisting of 1000 transactions at various transaction rates per second



---

**Algorithm 2:** Hospital, Pharmacy, Patho Lab and Insurance firm Enrollment.
 

---

**Input:** Enrollment Certificate ( $E_C$ ) from Certification Authority ( $C_A$ )

**Output:** Access to all nodes  $H_N, P_N, L_N, I_N$

**Initialization:**  $N_{Admin}$  should be valid node.  
 $N_{Admin}$  can Read/Write/Update/Remove participants  $H_N, P_N, L_N, I_N$  ;

**while** (*true*) **do**

**if**  $H_N$  is valid **and**  $H_{TIN}$  is valid **then**

Add Node ( $B_{NT}, H_N$ )

Grant Access( $H_N$ )

**else**

Invalid( $H_N$ )

**end**

**if**  $P_N$  is valid **and**  $P_{TIN}$  is valid **then**

Add Node ( $B_{NT}, P_N$ )

Grant Access( $P_N$ )

**else**

Invalid( $H_N$ )

**end**

**if**  $L_N$  is valid **and**  $L_{TIN}$  is valid **then**

Add Node ( $B_{NT}, L_N$ )

Grant Access( $L_N$ )

**else**

Invalid( $L_N$ )

**end**

**if**  $I_N$  is valid **and**  $I_{TIN}$  is valid **then**

Add Node ( $B_{NT}, I_N$ )

Grant Access( $I_N$ )

**else**

Invalid( $I_N$ )

**end**

**end**

bool  $chk \leftarrow (0 : \text{malicious}, 1 : \text{genuine})$

**if**  $!(\text{behaviour}(chk))$  **then**

Remove Update ( $H_N, P_N, L_N, I_N$ )

**else**

Add Update ( $H_N, P_N, L_N, I_N$ )

**end**

---

(tps).

The graphs in figure 8a and 8b highlight the average latency and throughput for varying transaction rates along with the number of transactions completed per minute per three network models. The network model using 1 Org and 1 Peer had the lowest average latency and the highest throughput per transaction rate while completing the highest number of tasks per minute. In contrast, the network model using 3 Org and 3 Peer had the highest average latency and the lowest throughput per transaction rate while completing the lowest number of tasks per minute. The network model using 2 Org and 2 Peer fell in between the above results. (i.e  $3org3peer > 2org2peer > 1org1peer$ ). So, we can conclude from figure 8a, that latency increased as the system scaled up with more organizations and more peers. Throughput of 1 org 1 peer is measured to be highest of 190 whereas it keeps decreasing with the number of organization and peer increases. For 2 org 2 peer, it was found to be 182, and the same for 3 org 3 peer was 180 in Figure 8b. Figure 8c shows successfully completed transactions per minute. 1 org 1 peer completed 5000 transactions in around 4 minutes whereas the 2 org 2 peers completed 4500 transactions and 3 org 3 peers completed 4000 transactions at the same time. As a result, transaction time has been observed to increase in perfect sync with the organization's and peers' growth. Figure 8d on the other hand, highlights the CPU consumption per network model. For different rates of transactions, the network models resulted in varying average CPU usage. We observed that among all peers, peer1.org1.example.com touched the highest CPU utilization at a transaction rate of 200 per sec. whereas peer0.org1.example.com recorded the lowest CPU utilization at a transaction rate of 100 per sec. Table 2, details other resource consumption parameters. With these experiment results, we move forward to our next set of experiments related to data extraction from prescriptions.

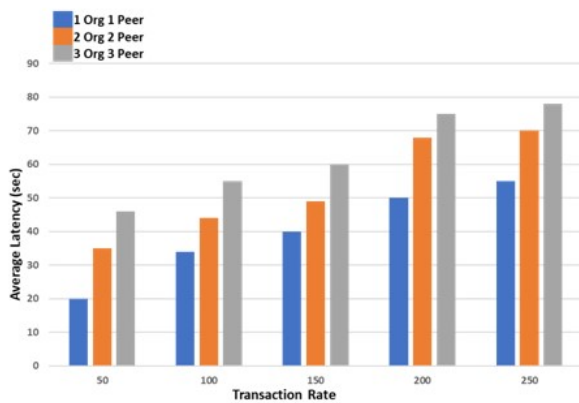
## 4.2 Handwritten prescription data extraction- training and validation

In regions dominated by paper prescription usage, providing the ability to transfer vast amounts of existing data into the digital space is essential. Allowing users to upload prescriptions simplifies the transition to an electronic system and populates the system with useful patient data given in Figure 9a, 9b and 10. The following experiments are related to training and validation for data extraction from handwritten prescriptions. The first step of training requires the number of inputs, hidden layers, and output layers. Twenty handwritten prescriptions of different classes are taken, including numeric characters, alphabetical characters, spaces, and punctuation. To improve image accuracy or legibility, the prescription image may be taken by section by users, causing the model to be confused and less accurate. To overcome this issue, images are converted into small segments. For analysis of images 64x64 pixels in size, 16 feature vectors are extracted from the feature

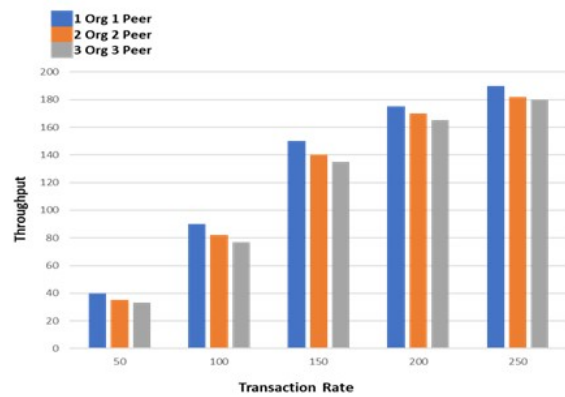


Abbreviation	Explanation
$P_{Rec}$	Patient’s Record
$B_{NT}$	Blockchain Network
$N_{Admin}$	Admin Node
$P_N, H_N, L_N, I_N$	Pharmacy, Hospital, Patho lab, Insurance firm nodes respectively
$P_{TIN}, H_{TIN}, L_{TIN}, I_{TIN}$	Pharmacy, Hospital, Patho lab, Insurance firm tax Id’s respectively

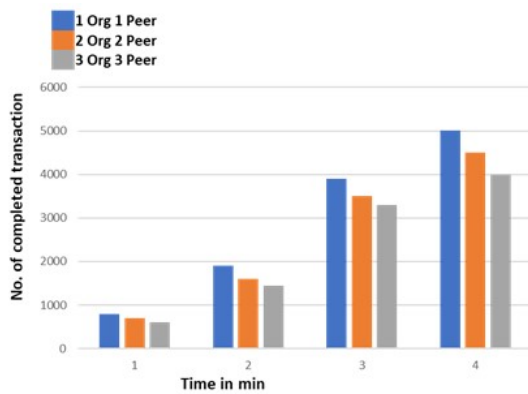
Table 1: Abbreviation used in the algorithm and its explanations.



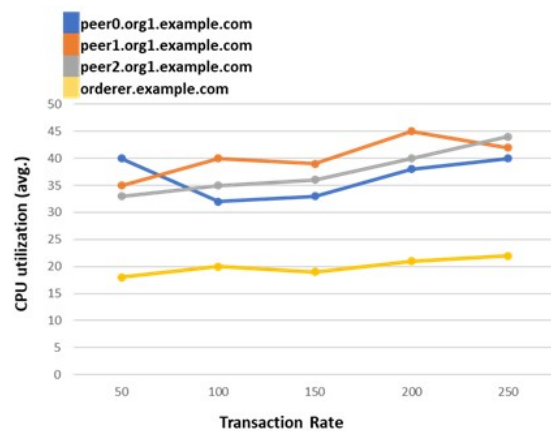
(a) Average Latency with Varying Transaction Rate.



(b) Throughput with Varying Transaction Rate.



(c) No. of Completed Transactions with Time.



(d) Resource consumption.

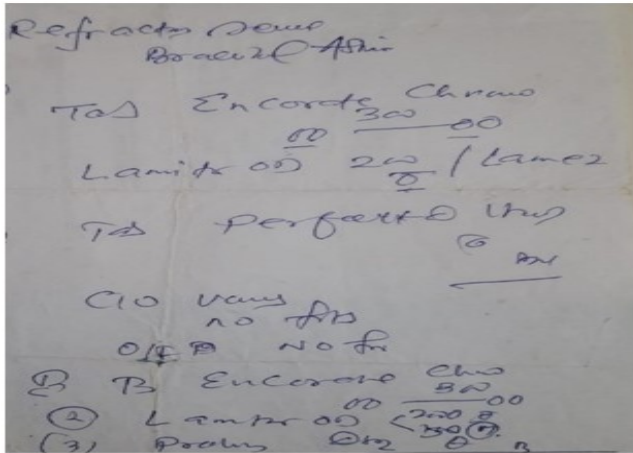
Figure 8: Measurement of different performance parameters. The results are calculated over five rounds, with each round consisting of 1000 transactions at various transaction rates per second (tps).

Type	Name	CPU(avg)	Memory(avg)	Traffic-In	Traffic-out	disc Write
Docker	peer0.org1.example.com	36.6	284.5MB	10.4MB	4.5MB	4.2MB
Docker	peer0.org2.example.com	28.4	280.0MB	10.5MB	5.6MB	4.2MB
Docker	peer0.org3.example.com	25.1	275.5MB	9.8MB	9.8MB	4.2MB
Docker	Orderer.example.com	2.34	50.0MB	2.5MB	1.2MB	1.2MB

Table 2: Resource consumption of various parameters.

map. The feature vectors are produced by the convolution layer and are extracted from each sliding. The model is

trained by 11,000 samples, with the epoch starting from 0 to 10000. The learning rate is 0.0005 when the size of the batch is 10. For this experiment, 2000 images are validated.



(a) Sample handwritten Prescription.

```
(base) debasish@ThinkPad-Yoga-460:~/NLP/Dustbin$ python Code.py
Refraction new
Br
Tas Encords Chrjkl
Lamitor on 200
Tas perfect
bjhvsdfgvhnfvjhj
O/F No For
B Encorete
Lamtor on 200 sjfgh
300
Prosdfbngngjf jhf df
(base) debasish@ThinkPad-Yoga-460:~/NLP/Dustbin$
```

(b) Sample output.

Figure 9: Handwritten prescription and sample output.

	CNN-LSTM	ResNet-LSTM
Train	89.1	90.3
Test	81.6	88.3

Table 3: Model Accuracy for Handwritten Data extraction approach.

According to the above results, the training accuracy was improved after the 30th epoch. Finally, an 88.3% accuracy result was accomplished using Test Dataset given in table 3. The generate model will take an input and generate a text output which is pre-processed by string operations. After the string operations are completed, the produced output will be sent to the network.

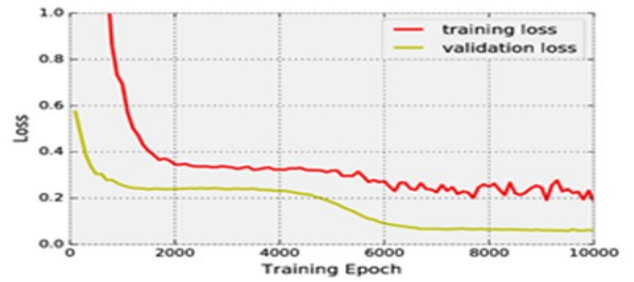


Figure 10: Training graph sample of Resnet-LSTM with output.

### 4.3 Printed prescription data extraction-training and validation

The Tesseract-OCR is a pre-trained model created by Google.

For improved image processing dilation, erosion methods are applied to Otsu’s thresholding. This approach also provides better accuracy as mentioned by the Google tesseract research. [16]. In our case, the test data is resulting in 99% accuracy. The illustration of the image is shown in 11.

*Dr Sanghamitra Kanungo*

Patient Name: SANBHAV DA                      Hospital : Kar Viaison Eye 2  
 Age : 9    Patient Id: KV20325  
 Gender : Male                                        Regd. Date: 13-Jul-2019 1:40 PM

(a) Printed Prescription.

```
Dr Sanghamitra Kanungo
Patient Name SANBHAV DA Hospital 2 Kar Viaion Bye
age 9 Patient ID Kv20325
Gender Mate Regd. Date 13-gul-2019 1540 #4
```

(b) Sample output.

Figure 11: Output of Tesseract OCR text extraction.

## 5 Conclusion and future work

EHR systems will be of considerable importance to advance the digital medical space of developing regions such as the Indian subcontinent. With the advancement of Blockchain technology, its potential has been recognized to significantly impact the future of EHR systems due to the superiority of Blockchain-based systems over traditional systems and paper-based record keeping. Blockchain-based EHR systems improve security, efficiency, and cost, making it an excellent option for the Indian subcontinent. In this paper, we have highlighted some common

issues that have arisen due to the lack of communication and tracking infrastructure, such as the influx of counterfeit drugs, dependability on handwritten prescriptions, and lack of integration between healthcare and insurance systems. We have discussed solutions to these issues using Blockchain, NLP, Hyperledger Fabric and Docker Containers, etc. The proposed scalable system will allow integration of existing EHR and healthcare databases, national identification, cloud technology to store large files with encryption, and a mobile app-based interface to convert paper prescriptions to text using OCR and deep learning techniques, then to bring old paper-based medical records into the new system. Our future research will include addressing implementation challenges at the grassroots level and also collect more samples for training, to increase the accuracy during converting the handwritten prescriptions into text.

## References

- [1] R. Achkar, K. Ghayad, R. Haidar, S. Saleh, and R. Al Hajj. Medical handwritten prescription recognition using crnn. In *2019 International Conference on Computer, Information and Telecommunication Systems (CITS)*, pages 1–5. IEEE, 2019. <https://doi.org/10.1109/CITS.2019.8862004>.
- [2] A. Baker. *Crossing the quality chasm: a new health system for the 21st century*, volume 323. British Medical Journal Publishing Group, 2001. <https://doi.org/10.1136/bmj.323.7322.1192>.
- [3] P. K. Bharimalla, S. Praharaj, and S. R. Dash. Ann based block chain security threat mechanism. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 8 (10), 2019. <https://doi.org/10.35940/ijitee.J9442.0881019>.
- [4] P. Bhattacharya, S. Tanwar, U. Bodke, S. Tyagi, and N. Kumar. Bindaas: Blockchain-based deep-learning as-a-service in healthcare 4.0 applications. *IEEE Transactions on Network Science and Engineering*, 2019. <https://doi.org/10.1109/TNSE.2019.2961932>.
- [5] M. Bhavin, S. Tanwar, N. Sharma, S. Tyagi, and N. Kumar. Blockchain and quantum blind signature-based hybrid scheme for healthcare 5.0 applications. *Journal of Information Security and Applications*, 56: 102673, 2021. <https://doi.org/10.1016/j.jisa.2020.102673>.
- [6] L. Chen, W.-K. Lee, C.-C. Chang, K.-K. R. Choo, and N. Zhang. Blockchain based searchable encryption for electronic health record sharing. *Future Generation Computer Systems*, 95:420–429, 2019. <https://doi.org/10.1016/j.future.2019.01.018>.
- [7] B. Devkota and A. Devkota. Electronic health records: advantages of use and barriers to adoption. *Health Renaissance*, 11(3):181–184, 2013. <https://doi.org/10.3126/hren.v11i3.9629>.
- [8] A. Dubovitskaya, F. Baig, Z. Xu, R. Shukla, P. S. Zambani, A. Swaminathan, M. M. Jahangir, K. Chowdhry, R. Lachhani, and N. Idnani. Actionehr: Patient-centric blockchain-based electronic health record data management for cancer care. *Journal of medical Internet research*, 22(8):e13598, 2020. <https://www.jmir.org/2020/8/e13598/>.
- [9] J. Fu, N. Wang, and Y. Cai. Privacy-preserving in healthcare blockchain systems based on lightweight message sharing. *Sensors*, 20(7):1898, 2020. <https://doi.org/10.3390/s20071898>.
- [10] G. Gavrilov, O. Simov, and V. Trajkovik. Blockchain-based model for authentication, authorization, and immutability of healthcare data in the referrals process. 2020. <http://hdl.handle.net/20.500.12188/8179>.
- [11] C. Kombe, A. Sam, M. Ally, and A. Finne. Blockchain technology in sub-saharan africa: Where does it fit in healthcare systems: A case of tanzania. *Journal of Health Informatics in Developing Countries*, 13(2), 2019. <https://www.jhidc.org/index.php/jhidc/article/view/24>.
- [12] D. S. Maitra, U. Bhattacharya, and S. K. Parui. Cnn based common approach to handwritten character recognition of multiple scripts. In *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, pages 1021–1025. IEEE, 2015. <https://doi.org/10.1109/ICDAR.2015.7333916>.
- [13] B. Mounia and C. Habiba. Big data privacy in healthcare moroccan context. *Procedia Computer Science*, 63:575–580, 2015. <https://doi.org/10.1016/j.procs.2015.08.387>.
- [14] R. Shanthapriya and V. Vaithianathan. Blockhealthnet: security based healthcare system using block-chain technology. *Security Journal*, pages 1–19, 2020. <https://doi.org/10.1057/s41284-020-00265-z>.
- [15] A. P. Singh, N. R. Pradhan, S. Agnihotri, N. Jhanjhi, S. Verma, U. Ghosh, D. Roy, et al. A novel patient-centric architectural framework for blockchain-enabled healthcare applications. *IEEE Transactions on Industrial Informatics*, 2020. <https://doi.org/10.1109/TII.2020.3037889>.
- [16] A. P. Tafti, A. Baghaie, M. Assefi, H. R. Arabnia, Z. Yu, and P. Peissig. Ocr as a service: an experimental evaluation of google docs ocr, tesseract, abbyy finereader, and transym. In *International*

- Symposium on Visual Computing*, pages 735–746. Springer, 2016. [https://doi.org/10.1007/978-3-319-50835-1\\_66](https://doi.org/10.1007/978-3-319-50835-1_66).
- [17] S. Tanwar, K. Parekh, and R. Evans. Blockchain-based electronic healthcare record system for healthcare 4.0 applications. *Journal of Information Security and Applications*, 50:102407, 2020. <https://doi.org/10.1016/j.jisa.2019.102407>.
- [18] M. Thakkar and D. C. Davis. Risks, barriers, and benefits of ehr systems: a comparative study based on size of hospital. *Perspectives in Health Information Management/AHIMA, American Health Information Management Association*, 3, 2006. <https://pubmed.ncbi.nlm.nih.gov/18066363/>.
- [19] D. Thompson, F. Velasco, D. Classen, and R. J. Raddemann. Reducing clinical costs with an ehr: investments in performance management are essential to realizing the full benefits of an ehr system—including reduced costs and improved quality of care. *Healthcare Financial Management*, 64(10):106–112, 2010. <https://link.gale.com/apps/doc/A243277528/AONE?u=anon~b7bb66fd&sid=googleScholar&xid=4b596be9>.
- [20] V. Varadharajan, D. Bansal, S. J. Nair, et al. Blockchain reinventing the healthcare industry: Use cases and applications. In *Industry Use Cases on Blockchain Technology Applications in IoT and the Financial Sector*, pages 309–328. IGI Global, 2021. <https://doi.org/10.4018/978-1-7998-6650-3.ch013>.