Management of Medical Data Access Using Blockchain Technology

Adeogun, D. Aminat, Ogunseye, E. Oluyemisi and Akinola, S. Olalekan

Department of Computer Science, University of Ibadan, Ibadan, Nigeria

Abstract

While there is still a considerable amount of debate around the superiority of electronic medical records over paper records, the research literature paints a more realistic picture of the benefits and downsides. The aim of this research work is to develop a blockchain model for enhancing trust and reliability in the medical care sector. The work used the immutable, transparent, secured and distributed properties of the blockchain to create a medical data access environment that allows patients and health administrators to have proper and secure access to their data, which is void of any form of tampering; hence leading to reduction in death rates among patients and quick response to users of the system. K-anonymity and K means algorithm were applied to the medical data obtained to cluster the medical data records based on the patient details to help provide management and monitoring options for both the patients and the health care institution. The result of this work showed a tamper-proof record creation and data access demonstrated by storing the patients' records into a blockchain and then the proposed k means algorithm was able to efficiently cluster the entire users into four cluster groups based on their medical records. The work was able to build a resilient system that prevents data loss as it is stored in a reliable system that aids maintenance of data integrity, assures individual of their records privacy, forms trusted partnerships and is immutable.

Key words: Blockchain technology, Medical patients' records, Medical data access, K-means algorithm, K-anonymity

1. Introduction

Medical Data Access Management System (MAMS) manages authentication, confidentiality, accountability and data sharing - crucial considerations when handling sensitive information for health care centers leveraging on unique blockchain properties. Medical Data Access Management System (MAMS) gives patients a comprehensive, immutable log and easy access to their medical information across providers and treatment site [1]. Patients benefit from a holistic, transparent picture of their medical history. This proves crucial in establishing trust and continued participation in the medical system, as patients that doubt the confidentiality of their records may abstain from full, honest disclosures or even avoid treatment. In the age of online banking and social media, patients are increasingly willing, able and desirous of managing their data on the web and on the go [8].

Record maintenance can prove quite challenging to initiate as patients are rarely encouraged and seldom enabled to review their full record [6]. Patients thus interact with records in a broken manner that reflects the nature of how these records are managed.

No single institution can hope to encompass a patient's entire record. Ideally, it should be possible to create or assemble each patient's personal health record so that it is accessible at all points of care within the health service and contains data from all institutions involved in that patient's care. Two main impediments stand in the way of this ideal: Firstly, most healthcare institutions do not provide effective access for patients to their own data and, despite technical feasibility, they show little willingness to share data with their competitors. Secondly, patients are becoming increasingly anxious about the privacy of their medical records. Such concerns seem justified when one considers that identifiable medical data are routinely shared with insurance companies, government, researchers, employers, state bureaus of vital statistics, pharmacy benefit managers (companies that track doctors' drug prescriptions), local retail pharmacies, attorneys, and others [8].

The aim of this project is to design and develop a medical data access management system using blockchain technology which will deliver a system-based record management solution.

The remaining part of this paper will be organized as follows. In Section 2 relevant existing works are presented. Section 3 is for the methodology used in this work. Presentation of results and discussion are done in Section 4 while conclusion is presented in Section 5.

2. Literature Review

Medical data describes the systematic documentation of a single patient's medical history and care across time within one particular health care provider's jurisdiction. The medical record includes a variety of types of "notes" entered over time by health care professionals, recording observations and administration of drugs and therapies, orders for the administration of drugs and therapies, test results, x-rays and reports. The maintenance of complete and accurate medical records is a requirement of health care providers and is generally enforced as a licensing or certification prerequisite [7]. Poor management of the patient’s medical data will eventually result to death of patient.

Blockchain technology could play a pivotal role in the healthcare industry with several applications in areas like public healthcare management, longitudinal healthcare records, automated health claims adjudication, online patient access, sharing patients’ medical data, user - oriented medical research, drug counterfeiting, clinical trial, and precision medicine. In particular, blockchain technology and the use of Smart Contracts could solve problems of scientific credibility of findings (missing data, endpoint switching, data dredging, and selective publication) in clinical trials as well as issues of patients’ informed consent. Managing patients’ Electronic Medical Records (EMRs) is probably the area with the highest potential growth. An EMR contains a patient’s short medical history, as part of her medical record, as well as data, predictions, and information of any kind relating to the conditions and the clinical progress of a patient throughout the course of a treatment.

A blockchain system for EMRs could be seen as a protocol through which users may access and maintain their health data that simultaneously guarantees security and privacy. The benefits of a blockchain-based system for EMRs are manifold: records are stored in a distributed way (they are public and easily verifiable across non-affiliated provider organisations), there is no centralised owner or hub for a hacker to corrupt or breach, data is updated and always available whereas data from disparate sources is brought together in a single and unified data repository [2].

The storage scheme of medical data uses blockchain technology and cloud storage technology to achieve safe storage and sharing. Medical institutions, patients and third-party agencies (such as medical information service platform, medical insurance company, etc.) are three main types of transaction bodies in the medical blockchain. Medical institutions are responsible for the diagnosis and treatment of patients and generating their medical records.

Patients can visit a doctor in different medical institutions and have ownership and control over their personal medical data. The third-party agencies can provide some services, such as medical institution recommendation and appointment registration. Different types of transaction bodies have different permissions. Data storage and access control are the main transactions in the medical blockchain. It would be optimal to be able to hold all medical data on the blockchain, but due to practical constraints such as cost, storage capacity, only index information of medical data and transaction records are recorded onto the blockchain. Large medical data should be encrypted and saved outside of the blockchain. These medical data are stored in cloud storage under the chain. Access control is determined by permission, and different transaction entities have different access control permissions. In the medical blockchain, the right to use personal medical data is entirely controlled by the patient, the patient may grant a subject access to the relevant data. The patient can also withdraw their authorization in time [3].

Frameworks and trials such as the one at the Sweden Land Registry aim to demonstrate the effectiveness of the blockchain at speeding land sale deals. The Republic of Georgia is piloting a blockchain-based property registry. The Government of India is fighting land fraud with the help of a blockchain. In the first half of 2018, an experiment was conducted on the use of blockchain technology to monitor the reliability of the Unified State Real Estate Register (USRER) data in the territory of Moscow [9].

Blockchain integrity verification applications store information and transactions related to the creation and lifetime of products or services. Mature solutions like Ascribe and Mediachain use Bitcoin blockchain to link digital content with their creators. Ascribe uses it to transfer ownership and loan digital assets, while Mediachain tries to store metadata on the blockchain to allow media recovery and querying [2].

For several years e-voting has been considered a promising and inevitable development which could speed up voting processes, simplify and reduce the cost of elections, and the development of stronger democracies. Decentralised voting systems such as BitCongress and Liquid Democracy propose frameworks to enforce distributed decision making. In general, blockchain technology offers an open-source, peer-to-peer, decentralised and independently verifiable network to gain the confidence required by voters and election organisers while being consistent with domestic legislation [2].

Blockchain applications appear to offer considerable performance enhancement and commercialisation opportunities, improving credibility in e-commerce and enabling IoT companies to optimise their operations while saving time and cost. Blockchain-based applications could serve as decentralised business process management systems for several enterprises [2].

Blockchain technology can be used for any transaction or information exchange that takes place in which the government is involved. Governance plays a role in blockchain as well. Governance of the blockchain technology determines how the technology operates and how the users can engage with it. All too often there might be a few experts who dictate the rules in which the application governs the users, whereas policymakers should play a prominent role to ensure that public values and societal needs are fulfilled and taken into account in the design and governance of Blockchain architectures and applications. Close cooperation between experts and policy-makers is needed to develop governance by BC on the one hand, and to ensure compliance with public values and societal needs for BC applications developed by other parties on the other hand [11].

Financial Institutions Joining Forces Barclays, Credit Suisse, Canadian Imperial Bank of Commerce, HSBC, MUFG and State Street have joined ranks on the Unity Settlement Coin (USC), a digital currency created by Switzerland's UBS bank in conjunction with UK-based Clearmatics. By exploiting the benefits of blockchain technology, USC partners expect to make it faster and safer for central banks to settle [5].

Blockchain may reduce costs and enable new business models and marketplaces, can better manage complexity, data security, and ownership along grids, can engage prosumers in the energy market acting as enabler for the creation of energy communities, can enhance the transparency and trust of the energy market system, can guarantee accountability while preserving privacy requirements, can enhance direct peer-to-peer trading to support the smooth operation of the power grid, and can better handle demand response and provide a framework for more efficient utility billing processes and transactive energy operations [2].

Teachers add blocks into the blockchain storing the learning achievements of students. Educational certificate management can also be enhanced by blockchain improving data security and trust in digital infrastructures, and for credit [2].

Whether we like it or not, online companies know all about us. Some companies whom we purchase from sell our identity details to advertisers who send people their ads. The blockchain blocks by creating a protected data point where one encrypts only the information that he/she wants relevant people to know at certain times. Identity such as Passports; Birth, wedding, and death certificates and Personal Identification are examples of means of identification used always. [12].

Wearables and mobile apps today support fitness, health education, symptom tracking, and collaborative disease management and care coordination. All those platform analytics can raise the relevancy of data interpretations, reducing the amount of time that end users spend piecing together data outputs. Insights gained from big data analysis will drive the digital disruption of the healthcare world, business processes and real-time decision-making. A new category of "personalised preventative health coaches" (Digital Health Advisors) will emerge through IoT. These workers as mentioned above will possess the skills and the ability to interpret and understand health and well-being data. They will help their clients avoid chronic and diet-related illness, improve cognitive functions, achieve improved mental health and achieve improved lifestyles overall [4].

K-means clustering is used because it is simple and has relatively low computational complexity. In addition, it is suitable for biomedical image segmentation as the number of clusters (K) is usually known for images of particular regions of human anatomy. Magnetic Resonance (MR) Image of the head generally consists of regions representing the bone, soft tissue, fat and background. Hence we select K to be 4. **I**nitial cluster centers are chosen in a first pass of the data. The dataset is partitioned into K clusters and the data points are randomly assigned to the clusters resulting in clusters that have roughly the same number of data points. For each data point, we calculate the Euclidean distance from the data point to the mean of each cluster. If the data point is not closest to its own cluster, it will have to be shifted into the closest cluster. If the data point is already closest to its own cluster, we will not shift it. The process continues until cluster means do not shift more than a given cut-off value or the iteration limit is reached.

Electronic Health Record (EHR) databases contain vast amounts of information about patients. Machine learning techniques such as Boosting and support vector machine (SVM) can potentially identify patients at high risk for serious conditions, such as heart disease, from EHR data. With objectives of modelling detection of heart failure more than 6 months before the actual date of clinical diagnosis using machine learning techniques applied to EHR data and compare the performance of logistic regression, SVM, and Boosting, along with various variable selection methods in heart failure prediction. [14].

3. Methodology

This section presents the method and the approach used in carrying out this research work. It shows the underlying factors to consider for the proper implementation of our system and further gives a detailed explanation of the work in the subsequent sections. The end goal of this work is to build a model that will establish trust and reliability in the access to patients' medical data collected at different points of their life and by different health care practitioners; thus avoiding any form of data theft and human manipulation. Also, the model should be able to develop a learning model to find trending patterns of patient illness based on patients' medical data which will aid in the reduction of patients' death and faster response to patients.

Discussed in this section is the operation of our proposed model which is geared towards establishing trust and reliability in the access to patients' medical records. Figure 1 depicts the overview of the entire blockchain model formulated to process the incoming medical data filled in by a health care administrator. It also shows the step by step approach for the clustering of anonymized patients' medical data based on their illnesses using K means algorithm.

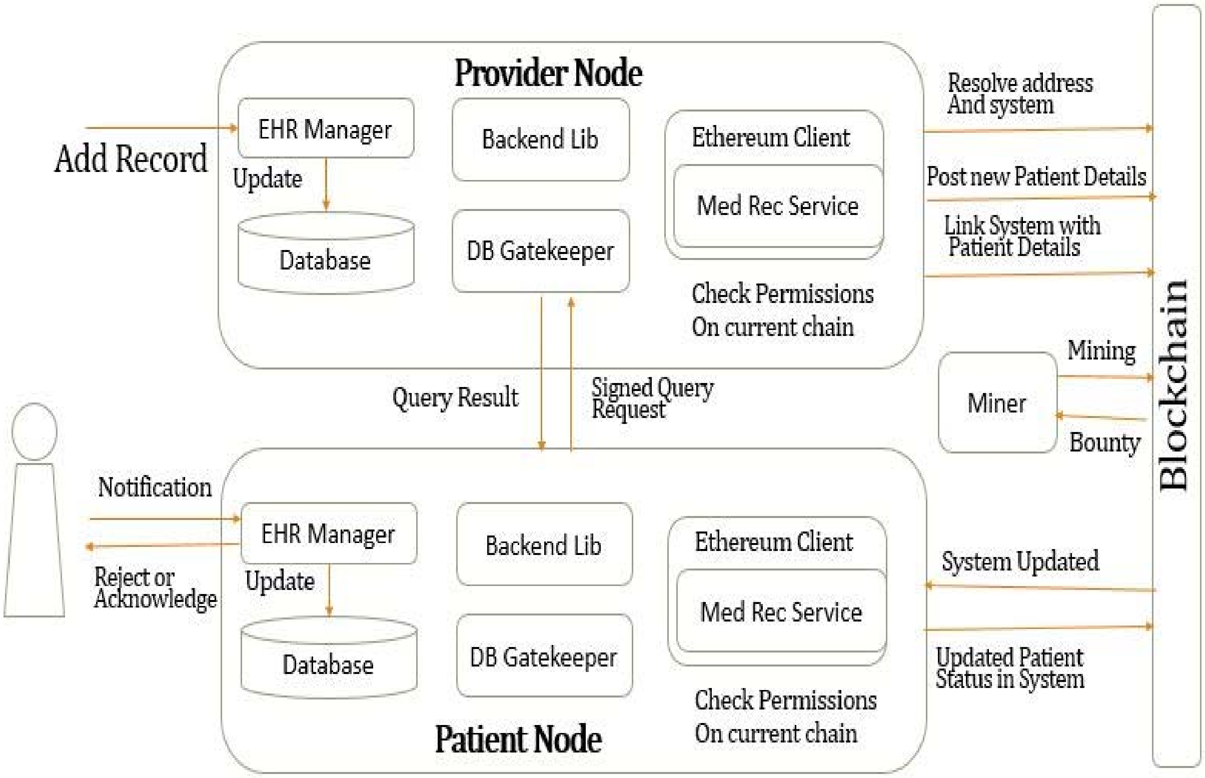


Figure 1: The overview of the entire blockchain model

3.1 Overview of the Proposed Model

In Figure 1, the model consists of four software components which can be executed on servers, combined to create a coherent, distributed system. The layers are explained in details below:

1) Backend Library: Here, a provider adds a record for a new patient. Using the Admin contract on the blockchain, the patient's identity information is first resolved to their matching Ethereum address and the corresponding summary contract is located. Next, the provider uploads a new patient details to the blockchain, indicating their stewardship of the data owned by the patient's Ethereum address. The provider node then crafts a query to reference this data and updates the patient details accordingly. Finally, the node sends a transaction which links the new patient details to the patient's summary contract, allowing the patient node to later locate it on the blockchain.

2) Ethereum Client: This component implements the full functionality required to join and participate in the Ethereum blockchain network. It handles a broad set of tasks, such as connecting to the peer-to-peer network, encoding and sending transactions and keeping a verified local copy of the blockchain. This service runs continuously within the client to monitor real-time changes to the system update. In the event of an update, the service signals the EMR Manager to issue a user notification. The patient's modified Ethereum client continuously monitors her system update. Once a new block is mined with the newly linked patient details, the client issues a signal which results in a user notification. The user can then acknowledge or decline her communication with the provider, updating the Summary Contract accordingly. If the communication is accepted, the prototype implementation automatically issues a query request to obtain the new medical data. It uses the information in the new patient details to locate the provider on the network and connect to its Database Gatekeeper server.

3) Database Gatekeeper: The Database Gatekeeper implements an off-chain, access interface to the node's local database, governed by permissions stored on the blockchain. The Gatekeeper runs a server listening to query requests from clients on the network. A request contains a query string, as well as a reference to the blockchain patient details that warrants permissions to run it. The request is cryptographically signed by the issuer, allowing the gatekeeper to confirm identities. Once the issuer's signature is certified, the gatekeeper checks the blockchain contracts to verify if the address issuing the request is allowed access to the query. If the address checks out, it runs the query on the node's local database and returns the result over to the client.

4) Electronic Medical Record Manager: All the software components previously mentioned were tie together with the Electronic Health Record management and user interface application. The application renders data from local SQLite database for viewing and presents the users with update notifications, data sharing and retrieval options.

5) Mining: A mining model brings medical researchers and health care authorities to mine in the network. In return the network beneficiaries, i.e. providers and patients, release access to aggregate, anonymized medical data as mining rewards. It requires care providers to attach a bounty query to any transaction they send updating the patient details. When the block containing the transaction is mined, the function automatically appends the block's miner as the owner of the bounty. The miner can then collect it by simply issuing a request for this bounty to the provider's Database gatekeeper. Because it is signed by the provider as part of the transaction, the bounty query is safe from malicious alterations.

3.2 Methodological Approach to Medical based Blockchain System

Figure 2 shows how the record keeping can be enforced to ensure a secured tamper proof data. The registration phase requires that the basic information of the users are registered for identity management. The phase will also be responsible for every record of addition that is to be stored in the system. The data to be stored are collected from a web interface and then stored into the Ethereum blockchain. The Ethereum framework makes it possible for Decentralized apps (Dapps) to be created for storing data into the blockchain network. Figure 3 shows the pictorial representation of the methodology cycle.

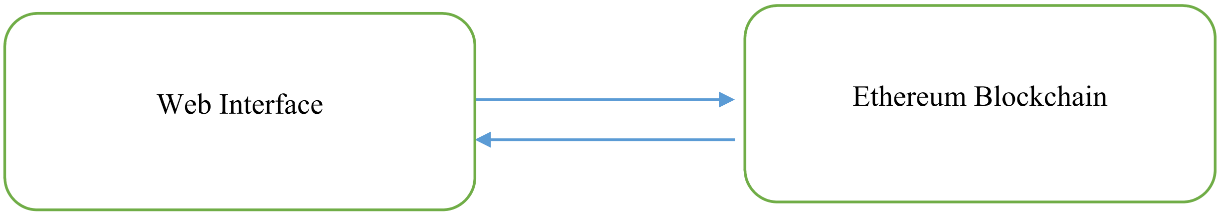


Figure 2: Model for secure medical data creation and access into the blockchain

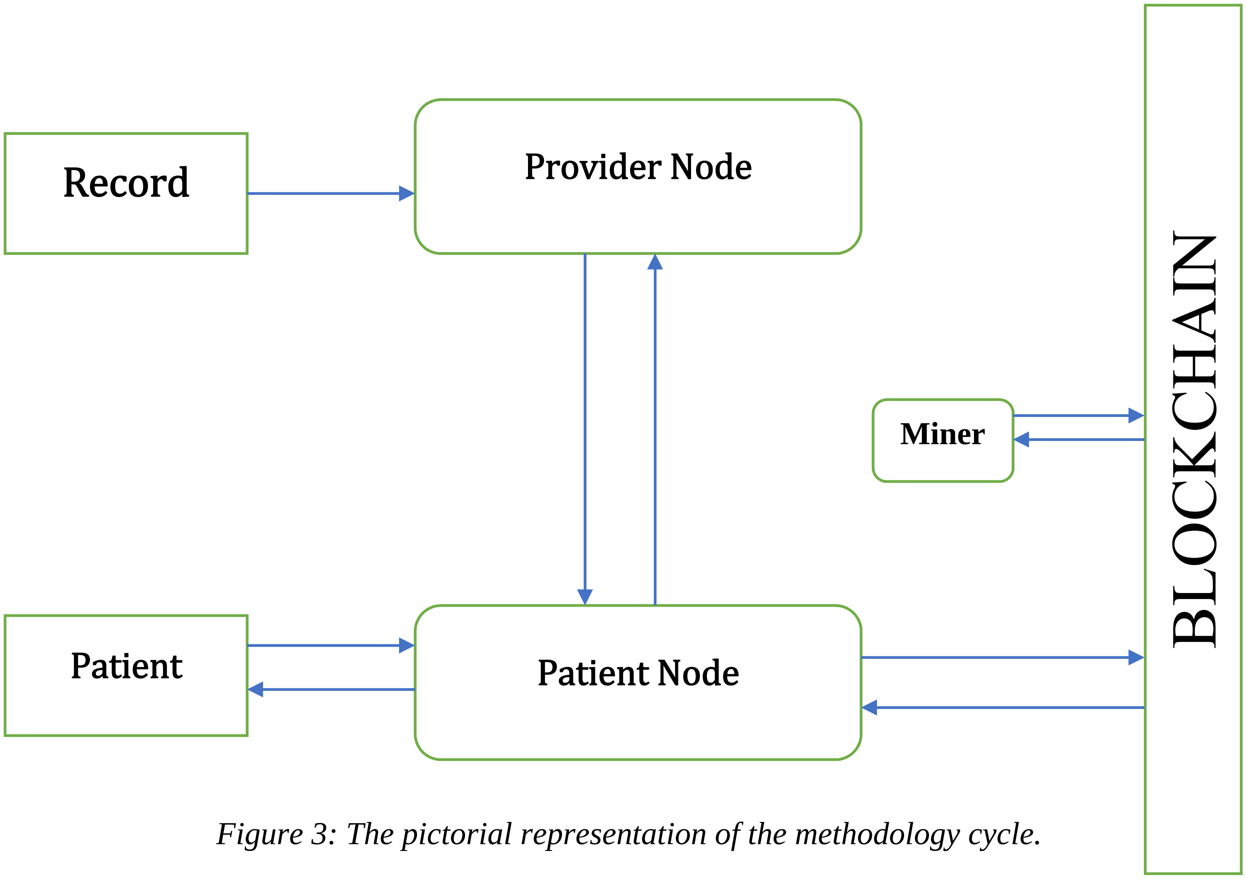


Figure 3: The pictorial representation of the methodology cycle.

For the processing of the patients' information, it takes the first name, last name, phone number, email, contact address and other data required and creates a block representing that transaction. This block is then broadcasted to all the nodes on the network and then the validation of the block commences using the proof of stake algorithm which the Ethereum blockchain uses. Once the nodes in the network validate this block, it gets added into the chain and the transaction gets verified and executed. While processing every incoming transaction, the smart contract module is always listening, and will get triggered enforcing the action of revoking access to the internet whenever a user tries a malicious attempt by tampering with the smart meter reading.

3.3 Applying Data Mining to Cluster anonymized patients' data

The Blockchain model will ensure that the collection of data into the network is secured and cannot be illegally altered. Once the data has been stored, the second phase of the model involves learning from the data to obtain clusters. The method applied is described in the following sections.

3.4 Data Extraction & Pre-Processing

The model will perform data extraction from the blockchain network database when new patients' records want to be obtained for very accurate learning results. A web interface is provided for data about the patients' medical details stored on the blockchain to be downloaded from.

The data pre-processing will help get rid of unwanted information that could be extracted as features and reduces the processing time and work to be done. Data reduction and transformation can also be performed if the dataset extracted is too large for processing at once. However, the pre-processing stage will not be ignored. This will involve performing data cleaning by recognizing outliers, smoothing noise data or correcting any inconsistent data. This is because extraction of nonrelevant features could lead to misclassification and reduce the accuracy of our system.

3.5 Data Anonymization

Patient sensitive data can be divided into three categories: explicit identifiers, quasi-identifiers, and privacy attributes. Explicit identifier can uniquely indicate a patient, such as an ID number, name, and cell phone number. A combination of quasi-identifiers can also uniquely indicate a patient, such as age, birth data, and address. Privacy information refers to sensitive attributes of a patient, including illness and income. It is necessary to ensure that the individual attributes of the new dataset are properly processed, so as to protect the patient's privacy. At present, random perturbation technology and data anonymous technology are usually used to solve these issues such as k-anonymity, l-diversity, and confidence bounding. In particular, the traditional k-anonymity is widely applied [13].

3.6 Application of Data Anonymization

Many clustering algorithms can be applied in data anonymization for k-anonymous data. In the context of longitudinal data, the challenge is to define a distance metric for trajectories. Figure 4 provides an overview of the longitudinal data anonymization process. The Maximum Distance to Average Vector (MDAV) algorithm, an efficient heuristic for k-anonymity is used to develop clustering algorithm. MDAV iteratively selects the most frequent trajectory in a longitudinal dataset and forms a cluster of at least k records around the latter.

In addition, they define the distance between two trajectories as the cost of their anonymization. The MDAV can generate anonymized data that permit effective biomedical analysis, using heuristics inspired from sequence alignment and clustering methods. A clustering method based K-anonymity algorithm was introduced as the building block of privacy preserving for medical wearable devices. The clustering K-anonymity would assign similar records into the same equivalent set, while the similarity among these records make it harder to discriminate different identities than before. Then, they unify the quasi-identifiers in the same clusters by generalizing and suppressing operations. The output of this MDAV algorithm is a table that satisfies the principle of K-anonymity. All the records in the same equivalent set are similar to each other. In this way, it would be harder to recognize the users' identities in one equivalent set, and the privacy of these subjects would be securer [13].

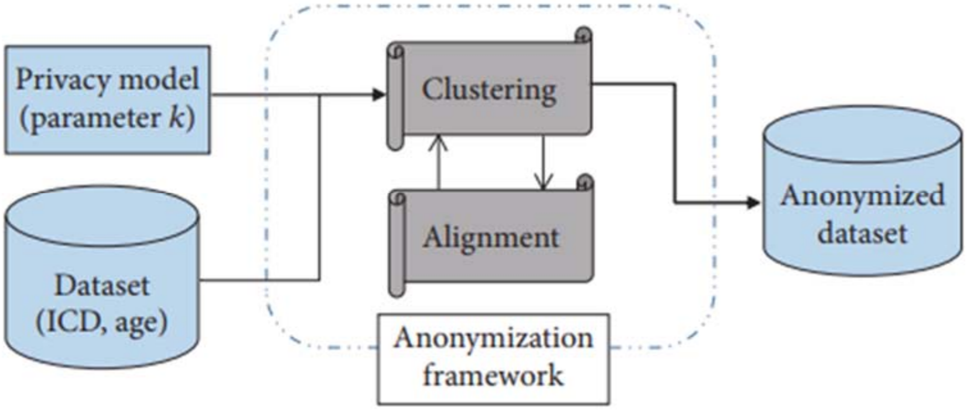


Figure 4: A general architecture of the longitudinal data anonymization process.

3.7 K-Means Clustering

K-Means Clustering is a fairly simple clustering algorithm that partitions the dataset into several clusters k. The algorithm is fairly easy to implement and run, relatively fast, easy to customize and widely used. This method is one of the non-hierarchical data clustering methods that group data in the form of one or more clusters / groups. The data that have the same characteristics are grouped in one cluster / group and the data having different characteristics are grouped with other clusters /groups so that the data in one cluster / group has small variation level. The purpose of this data clustering is to minimize the objective function set in the clustering process, which generally tries to minimize the variation within a cluster and maximize the variation between clusters. In general, K-Means Clustering method is done with the following steps:

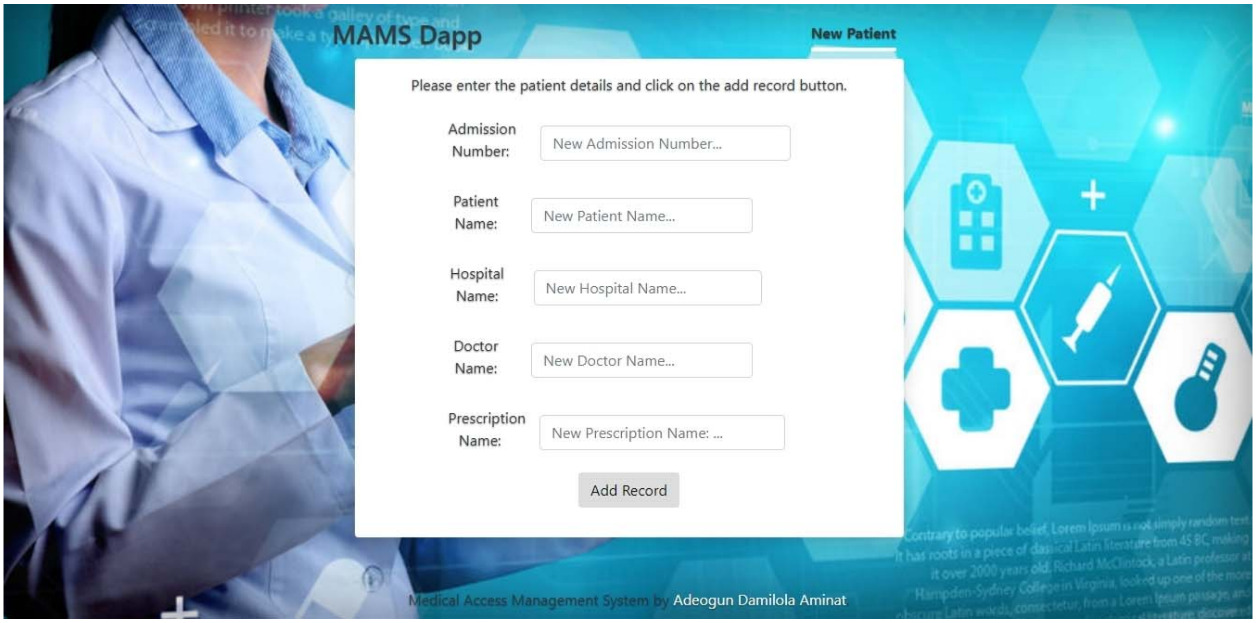
- Select the number of clusters k.
- Initialization of cluster center k. In general, cluster centers are given initial values with random numbers.
- Allocate all data / objects to the nearest cluster. The proximity of two objects is determined by the distance of the two objects. Likewise, the proximity of a data to a particular cluster is determined by the distance between the data and the cluster center.
- Recalculate cluster center with current cluster membership. The cluster center is the average of all data / objects in a particular cluster.

- Check each new cluster center user object. If the cluster center does not change again then the clustering process is complete. Or, go back to step 3 until the center of the cluster does not change anymore.

4. Results

4.1 Data Creation and Access

To ensure a reliable and immutable record keeping of the whole data generated from the hospital or health care center, we implemented a Decentralized Application built on using the Ethereum framework to store and retrieve information into the blockchain network. The proof of stake consensus algorithm was used to validate the blocks of transaction before it was finally added into the chain. Figure 5 shows the user interface for user to input their details which would be saved in the block in the blockchain. Figure 6 and Figure 7 show how the patient medical record are being stored into the blockchain respectively. Once the records are added to the chain successfully, it becomes impossible for data to be tampered with and the smart contract is executed in the event of attempts by the parties. Apart from the identity management of the users i.e., the patient and the health care administrator and the prescriptions, the system also provides management with the user's information for privacy reasons. Figure 8 shows the details of every block mined into the network as records are saved to the network. As each successive blocks keep getting added, it becomes computationally impossible for the records to be tampered with.



Please enter the patient details and click on the add record button.

Admission Number: New Admission Number...

Patient Name: New Patient Name...

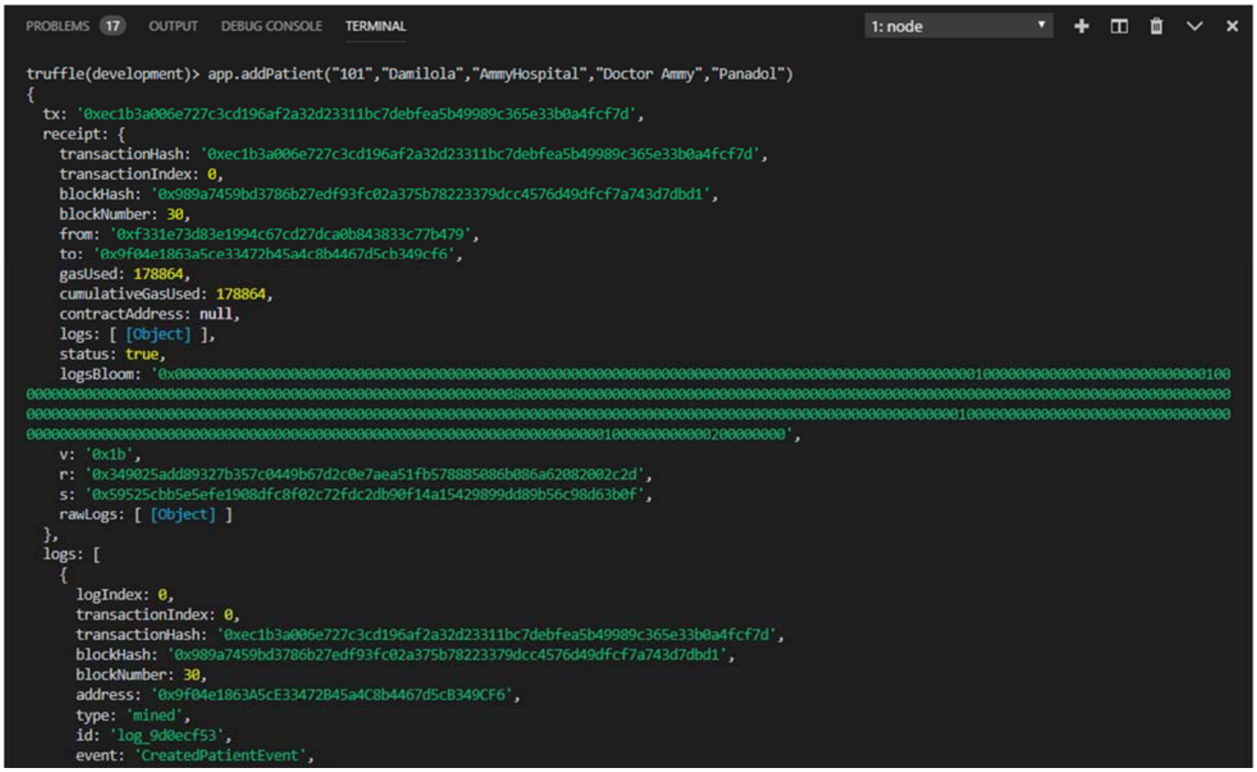
Hospital Name: New Hospital Name...

Doctor Name: New Doctor Name...

Prescription Name: New Prescription Name: ...

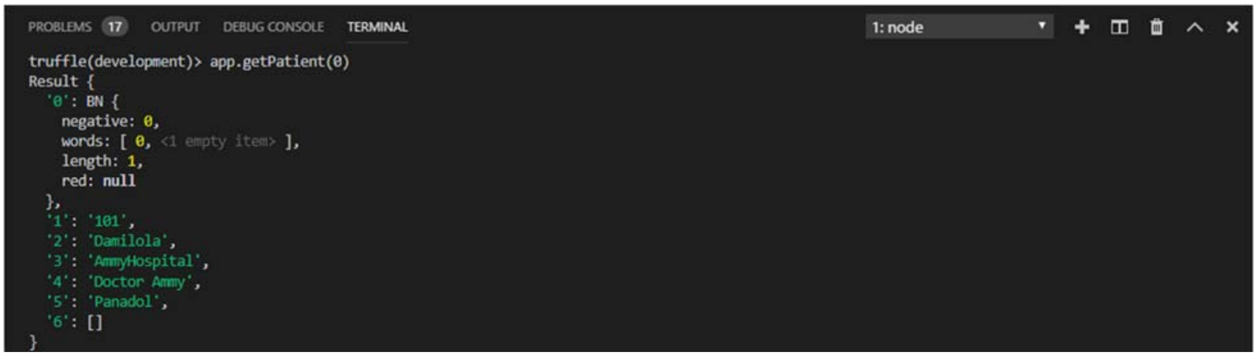
Medical Access Management System by Adeogun Damilola Aminat

Figure 5: A screen-shot showing user interface for user to input their details which would be saved in the block in the blockchain



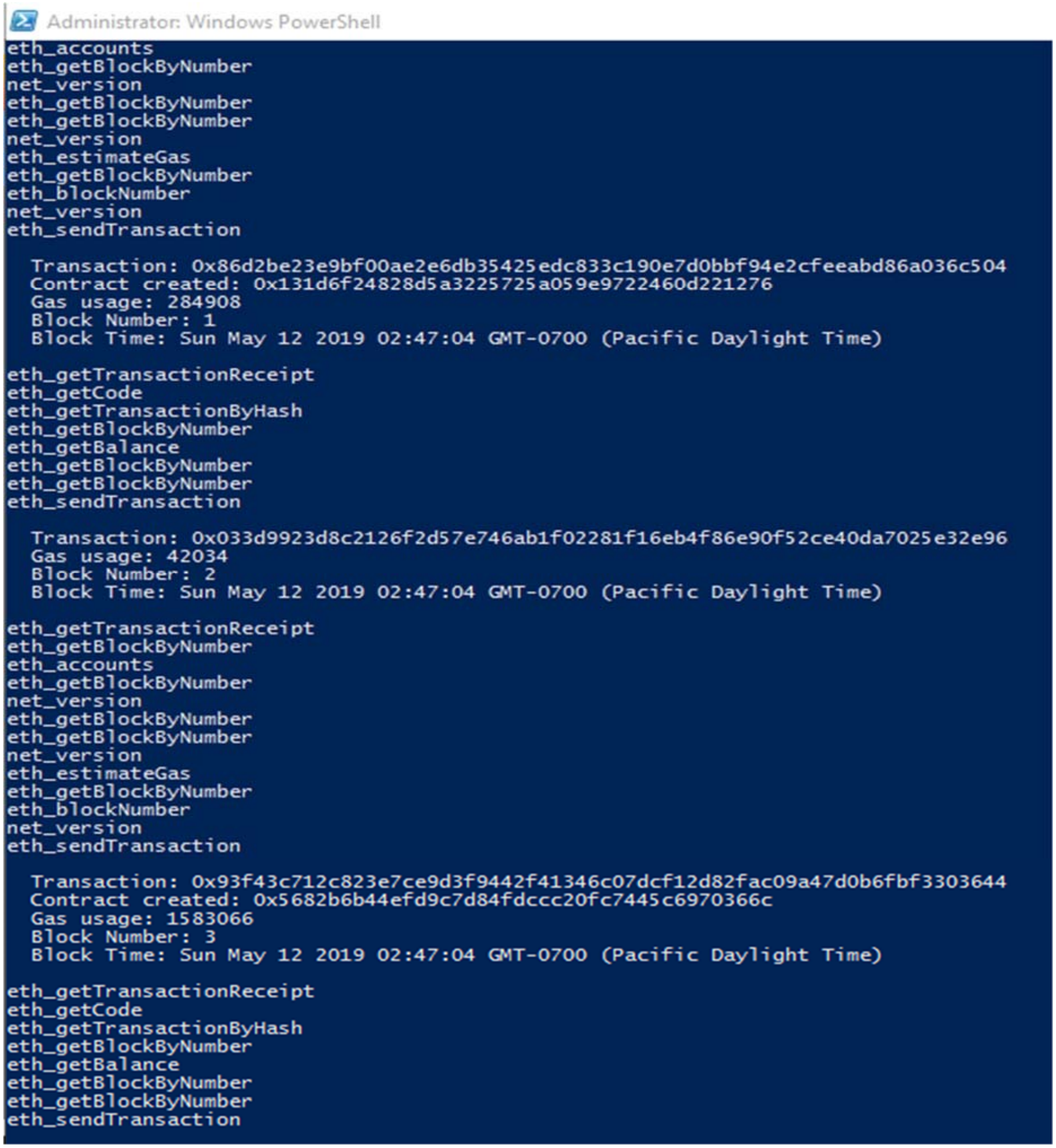
```
truffle(development)> app.addPatient("101","Damilola","AmmyHospital","Doctor Ammy","Panadol")
{
  tx: '0xec1b3a006e727c3cd196af2a32d23311bc7debfea5b49989c365e33b0a4fcf7d',
  receipt: {
    transactionHash: '0xec1b3a006e727c3cd196af2a32d23311bc7debfea5b49989c365e33b0a4fcf7d',
    transactionIndex: 0,
    blockHash: '0x989a7459bd3786b27edf93fc02a375b78223379dcc4576d49dfcf7a743d7dbd1',
    blockNumber: 30,
    from: '0xf331e73d83e1994c67cd27dca0b843833c77b479',
    to: '0x9f04e1863a5ce33472b45a4c8b4467d5cb349cf6',
    gasUsed: 178864,
    cumulativeGasUsed: 178864,
    contractAddress: null,
    logs: [ [Object] ],
    status: true,
    logsBloom: '0x0000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000100000000000000000000000000000100
00000000000000000000000000000000000000000000000000000000000800000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
00000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000100000000000000000000000000000000000
00000000000000000000000000000000000000000000000000000000000000000000001000000000000200000000',
    v: '0x1b',
    r: '0x349025add89327b357c0449b67d2c0e7aea51fb578885086b086a62082002c2d',
    s: '0x59525cbb5e5efe1908dfc8f02c72fdc2db90f14a15429899dd89b56c98d63b0f',
    rawLogs: [ [Object] ]
  },
  logs: [
    {
      logIndex: 0,
      transactionIndex: 0,
      transactionHash: '0xec1b3a006e727c3cd196af2a32d23311bc7debfea5b49989c365e33b0a4fcf7d',
      blockHash: '0x989a7459bd3786b27edf93fc02a375b78223379dcc4576d49dfcf7a743d7dbd1',
      blockNumber: 30,
      address: '0x9f04e1863A5cE33472B45a4C8b4467d5cB349CF6',
      type: 'mined',
      id: 'log_9d0ecf53',
      event: 'CreatedPatientEvent',
```

Figure 6: A screen-shot showing how the patient record is stored into the blockchain



```
truffle(development)> app.getPatient(0)
Result {
  '0': BN {
    negative: 0,
    words: [ 0, <1 empty item> ],
    length: 1,
    red: null
  },
  '1': '101',
  '2': 'Damilola',
  '3': 'AmmyHospital',
  '4': 'Doctor Ammy',
  '5': 'Panadol',
  '6': []
}
```

Figure 7: A screen-shot showing how the patient's record is retrieved from the blockchain



```
eth_accounts
eth_getBlockByNumber
net_version
eth_getBlockByNumber
eth_getBlockByNumber
net_version
eth_estimateGas
eth_getBlockByNumber
eth_blockNumber
net_version
eth_sendTransaction

  Transaction: 0x86d2be23e9bf00ae2e6db35425edc833c190e7d0bbf94e2cfeeabd86a036c504
  Contract created: 0x131d6f24828d5a3225725a059e9722460d221276
  Gas usage: 284908
  Block Number: 1
  Block Time: Sun May 12 2019 02:47:04 GMT-0700 (Pacific Daylight Time)

eth_getTransactionReceipt
eth_getCode
eth_getTransactionByHash
eth_getBlockByNumber
eth_getBalance
eth_getBlockByNumber
eth_getBlockByNumber
eth_sendTransaction

  Transaction: 0x033d9923d8c2126f2d57e746ab1f02281f16eb4f86e90f52ce40da7025e32e96
  Gas usage: 42034
  Block Number: 2
  Block Time: Sun May 12 2019 02:47:04 GMT-0700 (Pacific Daylight Time)

eth_getTransactionReceipt
eth_getBlockByNumber
eth_accounts
eth_getBlockByNumber
net_version
eth_getBlockByNumber
eth_getBlockByNumber
net_version
eth_estimateGas
eth_getBlockByNumber
eth_blockNumber
net_version
eth_sendTransaction

  Transaction: 0x93f43c712c823e7ce9d3f9442f41346c07dcf12d82fac09a47d0b6fbf3303644
  Contract created: 0x5682b6b44efd9c7d84fdccc20fc7445c6970366c
  Gas usage: 1583066
  Block Number: 3
  Block Time: Sun May 12 2019 02:47:04 GMT-0700 (Pacific Daylight Time)

eth_getTransactionReceipt
eth_getCode
eth_getTransactionByHash
eth_getBlockByNumber
eth_getBalance
eth_getBlockByNumber
eth_getBlockByNumber
eth_sendTransaction
```

Figure 8: The details of every block mined into the network as records are saved to the network. As each successive blocks keep getting added, it becomes computationally impossible for the records to be tampered with.

4.2 Data Collection and Preprocessing

For the implementation of this project, patients medical records were collected, the data were collected via registration during hospital visit. This data was pre-processed by splitting the data per patient and missing data were handled. Figures 9 and 10 show the results after the pre-processing and missing values has been handled.

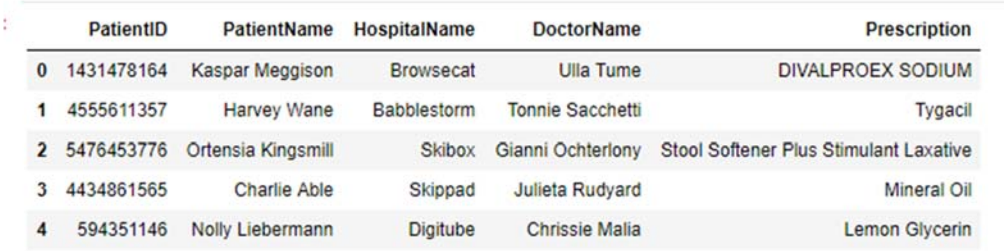
The data used contained 1000 records including patient's ID and name with their various doctor's name and hospital where the doctor's practice and also prescription given to the patient. The records were pre-processed using the following methods: import all important libraries needed, importing the dataset as shown in Figure 9, identifying and handling all values missing from the data set as shown in Figure 10 and finally taking care of categorical features.

4.2.1 Correlation

To properly understand our data and attempt to get the best results when we apply the K means algorithm, the StandardScaler from the sklearn package was used to fit transform the data before calculating the correlation between each pair of attributes obtained from the segmentation. Figure 11 shows the obtained statistical data of the first 26 medical data after segmentation.

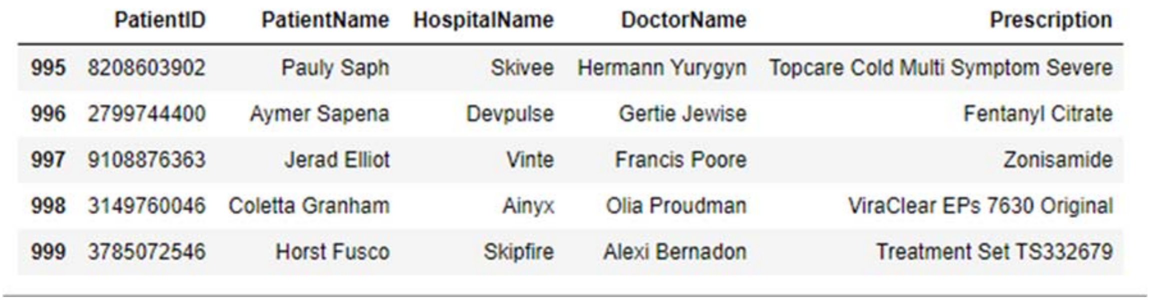
4.2.2 Clusters Generation

The Bayesian Information Criterion (BIC), was used to determine the K number of clusters to be used in segmenting the users. Figure 12 shows the output of the BIC scoring for the K means cell behaviour. The K number selected was 4 because it was the least point from the BIC plot.



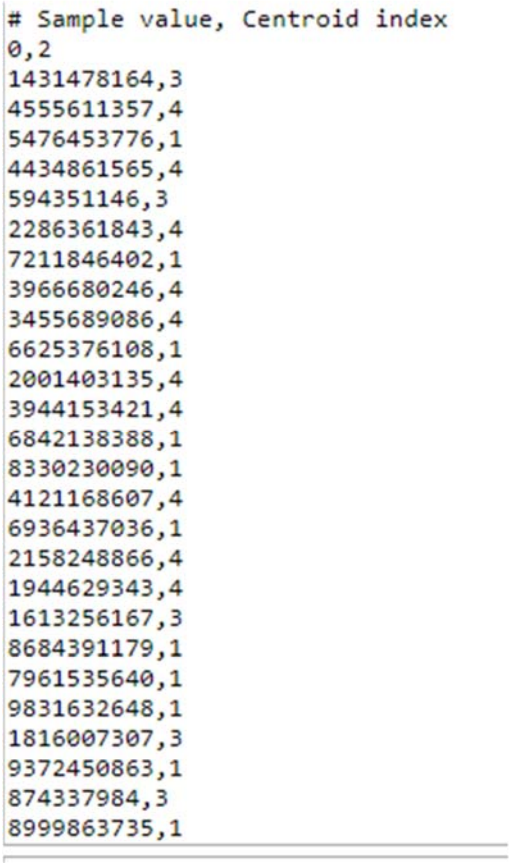
	PatientID	PatientName	HospitalName	DoctorName	Prescription
0	1431478164	Kaspar Meggison	Browsecat	Ulla Tume	DIVALPROEX SODIUM
1	4555611357	Harvey Wane	Babblestorm	Tonnie Sacchetti	Tygacil
2	5476453776	Ortensia Kingsmill	Skibox	Gianni Ochterlony	Stool Softener Plus Stimulant Laxative
3	4434861565	Charlie Able	Skippad	Julieta Rudyard	Mineral Oil
4	594351146	Nolly Liebermann	Digitube	Chrissie Malia	Lemon Glycerin

Figure 9: Rows showing the patient details i.e. the patientID, PatientName, HospitalName, DoctorName and Prescription



	PatientID	PatientName	HospitalName	DoctorName	Prescription
995	8208603902	Pauly Saph	Skivee	Hermann Yurygyn	Topcare Cold Multi Symptom Severe
996	2799744400	Aymer Sapena	Devpulse	Gertie Jewise	Fentanyl Citrate
997	9108876363	Jerad Elliot	Vinte	Francis Poore	Zonisamide
998	3149760046	Coletta Granham	Ainyx	Olia Proudman	ViraClear EPs 7630 Original
999	3785072546	Horst Fusco	Skipfire	Alexi Bernadon	Treatment Set TS332679

Figure 10: Tail result after missing values had been handled



```
# Sample value, Centroid index
0,2
1431478164,3
4555611357,4
5476453776,1
4434861565,4
594351146,3
2286361843,4
7211846402,1
3966680246,4
3455689086,4
6625376108,1
2001403135,4
3944153421,4
6842138388,1
8330230090,1
4121168607,4
6936437036,1
2158248866,4
1944629343,4
1613256167,3
8684391179,1
7961535640,1
9831632648,1
1816007307,3
9372450863,1
874337984,3
8999863735,1
```

Figure 11: The obtained statistical data of the first 26 medical data after segmentation.

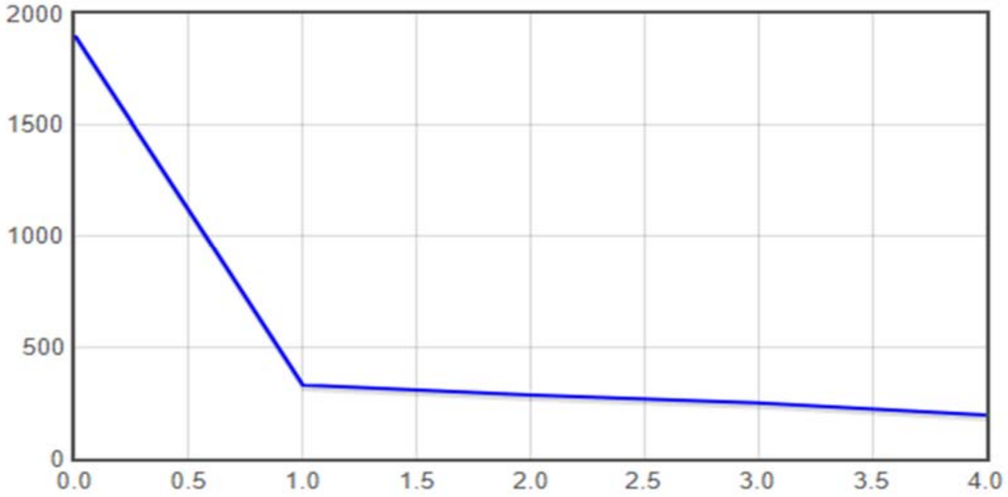
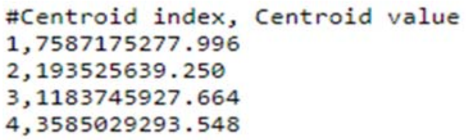


Figure 12: BIC score used to determine the K number of clusters

4.3 Result Discussion

Four (4) clusters were finally generated from the data using the K means algorithm. From Figure 13.

Centroid index and Value after clustering patient data into 4 clusters and Figure 14 shows cluster visualization showing the segmentation of the medical data, i.e. all segments are well separated from each other which implies that the Bayesian Information Criterion(BIC) method performed quite well in this work.



```
#Centroid index, Centroid value
1,7587175277.996
2,193525639.250
3,1183745927.664
4,3585029293.548
```

Figure 13: Centroid index and Value after clustering patient data into 4 clusters

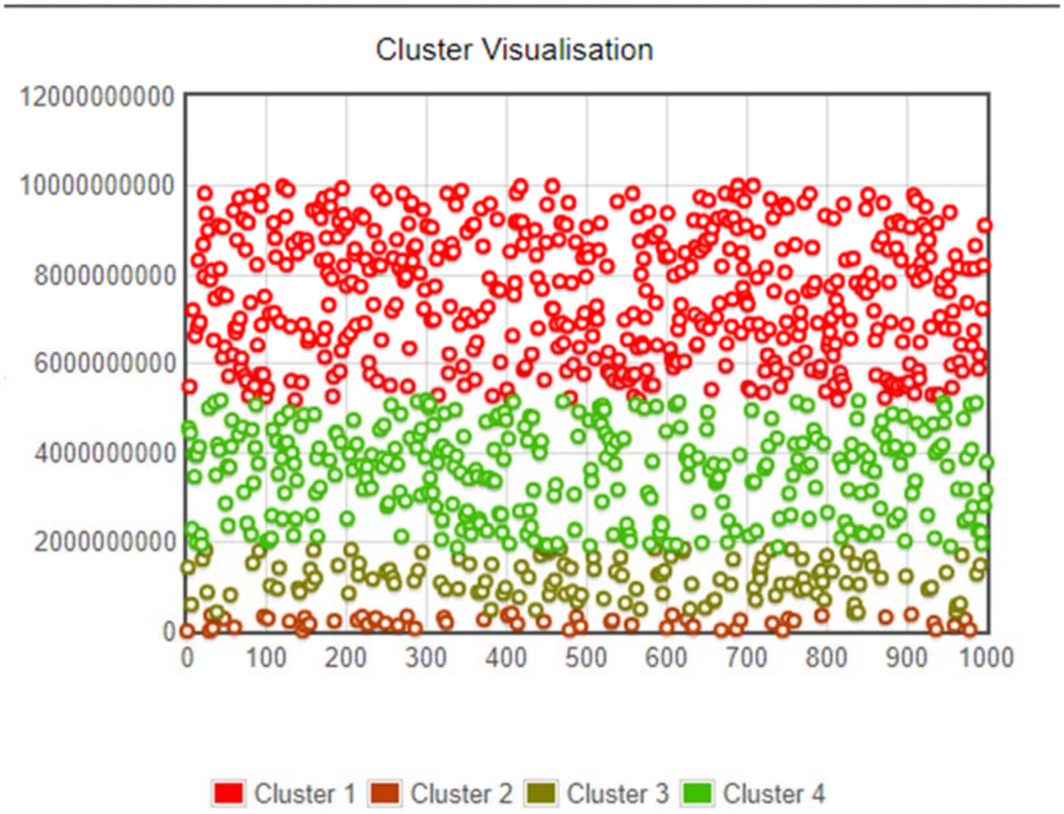


Figure 14: Cluster Visualization showing the segmentation of the medical data

5. Conclusion

A decentralised blockchain based application was created to ensure that every patient's medical records are appropriately and easy access granted to those in the patient's peer network. This will bring about transparency as users will also be able to access their medical records during emergency or transfer of doctors. Smart contracts were also implemented to enforce and detect any form of medical data theft which would automatically reject access to any individual caught in the act.

The work was also able to build a resilient system that prevent data loss as it is stored in a reliable system which aids in maintenance of data integrity, assures individual of their records privacy, forms trusted partnerships and is immutable**.**

References

[1] Azaria, Asaph, Ariel Ekblaw, Thiago Vieira, and Andrew Lippman. 2016. "MedRec: Using Blockchain for Medical Data Access and Permission Management." In *Proceedings - 2016 2nd International Conference on Open and Big Data, OBD 2016*,.

[2] Casino, Fran, Thomas K. Dasaklis, and Constantinos Patsakis. 2019. "A Systematic Literature Review of Blockchain-Based Applications: Current Status, Classification and Open Issues." *Telematics and Informatics* 36: 55–81. https://doi.org/10.1016/j.tele.2018.11.006.

[3] Chen, Yi *et. al.* 2018. "Blockchain-Based Medical Records Secure Storage and Medical

Service Framework." *Journal of Medical Systems* 43(1).

[4] Dimitrov, Dimiter V. 2016. "Medical Internet of Things and Big Data in Healthcare." *Healthcare Informatics Research* 22(3): 156–63.

[5] Eisenberg, Aviram. 2018. "Five Business Opportunities for Blockchain for Blockchain Business Applications." https://igniteoutsourcing.com/blockchain/blockchain-business-applications/

[6] Health, Info and Law. 2010. "TOPICS FEDERAL STATE ANALYSIS RESOURCES Who Owns Medical Records : 50 State Comparison." http://www.healthinfolaw.org/comparative-analysis/who-owns-medical-records-50-state-comparison

[7] Kamieńska, W. 1977. "Medical Record." *Pielegniarka i polozna* (8): 9–10. http://www.ncbi.nlm.nih.gov/pubmed/588019.

[8] Mandl, K. D. 2001. "Public Standards and Patients' Control: How to Keep Electronic Medical Records Accessible but Private Commentary: Open Approaches to Electronic Patient Records Commentary: A Patient's Viewpoint." *Bmj* 322(7281): 283–87. http://www.bmj.com/cgi/doi/10.1136/bmj.322.7281.283.

[9] Blockchain, En.wikipedia.org

https://en.wikipedia.org/wiki/Blockchain

[10] Ng, H.P. et al. 2006. "Medical Image Segmentation Using K-Means Clustering and Improved Watershed Algorithm." *2006 IEEE Southwest Symposium on Image Analysis and Interpretation*, Denver, CO, 2006, pp. 61-65, doi: 10.1109/SSIAI.2006.1633722.

[11] Ølnes, Svein, Jolien Ubacht, and Marijn Janssen. 2017. "Blockchain in Government: Benefits and Implications of Distributed Ledger Technology for Information Sharing." *Government Information Quarterly* 34(3): 355–64.

[12] Rosic, Ameer. 2017. "17 Blockchain Applications That Are Transforming Society." *Blockgeeks*. https://blockgeeks.com/guides/blockchain-applications/.

[13] Sun, Wencheng et al. 2018. "Security and Privacy in the Medical Internet of Things: A Review." *Security and Communication Networks* 2018: 1–9.

[14] Wu, Jionglin, Jason Roy, and Walter F. Stewart. 2010. "Prediction Modeling Using EHR Data." *Medical Care*. June 2010-Volume 48 - Issue 6 - p - S106-S113 doi: 10.1097/MLR.0b013e3181de9e17

[15] Fran Casino, Thomas K. Dasaklis, Constantinos Patsakis, "A systematic literature review of blockchain-based applications: Current status, classification and open issues", Telematics and Informatics, Volume 36, 2019, Pages 55-81, ISSN 0736-5853, https://doi.org/10.1016/j.tele.2018.11.006. (http://www.sciencedirect.com/science/article/pii/S0736585318306324)