Privacy Preserving of Public Health Record using Access Control Mechanism through Blockchain and AI

Kajal Umesh Kamthe
Computer Engineering,
S.K.N. College of Engineering, Pune, India,

Dr.Gitanjali Shinde
Computer Engineering,
S.K.N. College of Engineering, Pune, India.

***Abstract* – Good health is one of the most valuable resources for a human being. Large-scale technological advancements have significantly improved human health by a large margin. But there are still several diseases whose symptoms are highly painful and debilitating for a patient. The doctor is in charge of facilitating an effective diagnosis of the patient's illness and providing relief to the patient through various medications. For this purpose, the doctor has at his disposal repository of information in the form of PHR or Public Health Records which can be utilized for querying the symptoms experienced by the patient. Sometimes the Medical Institute does not have the PHR with the exact parameters and symptoms described by the patient. Due to the deficiency in the PHR database, the doctor can employ the use of a data aggregator or vendor for fulfilling the requirement from other medical institutions. The problem in this approach is that there is a lack of trust between the medical Institutions and the data aggregators, as PHR can contain sensitive and personally identifiable information that cannot be shared easily. Therefore, this Publication details the implementation of an effective access control mechanism through the use of the distributed blockchain framework along with the introduction of K-means clustering and Linear Regression catalysed by the Hidden Markove model and Fuzzy Classificaiton model to alleviate the trust issues. Extensive experimentation on the proposed methodology has confirmed its superiority in comparison to the conventional approaches.**

Keywords: Privacy Preserving, Blockchains, K-Means Clustering, Hidden Markov model, Linear Regression, Fuzzy Classification.

I. INTRODUCTION

There have been significant advances in the Healthcare system which have been instrumental in reducing the child mortality rates as well as increasing the life expectancy of individuals by a large margin. This has been responsible is steadily increasing the population of this planet as well as increasing the quality of health significantly. There have been technological advances that have allowed such a large-scale increase in the quality of Healthcare.

The significant increase in the population has been detrimental to the Healthcare system which is very ironic. There is also a significant difference between the health care that is offered in the developing countries compared to the Healthcare offered in the developed countries. This is due to the spending power of the respective countries into the Healthcare system. The significant advancements help the developed countries in achieving how much better Lifestyle for the citizens of their country. The various medical improvements have also been precipitated towards the developing countries in the form of improved hardware and medicines. This also leads to a reduction in the amount of pain and suffering for the people in developing countries.

But the diagnosing procedure still different in both types of countries. In the developing countries diagnosing is done very differently and it can have a significant effect on the patient's wellbeing and their mental health. When efficiently come to the doctor with the set of symptoms doctor diagnosis the patient according to do the symptoms that have been elucidated by the patient. Most of the time there is a discrepancy in the diagnosis as the patient might not be able to communicate the symptoms. Also due to various symptoms that are similar between two different diseases the doctor has to choose depending on his or her experience which one of the two diseases should he provide the medication. If the doctor cannot decide he has to try out the medication for one element and check if there is any improvement in the symptoms. If there is no improvement noticed in the symptoms then the doctor will diagnose The Other disease and provide medication for it. This technique is called the trial and error method and is widely used in various developing countries.

This is very different from how the diagnosis is dealt with in the developed countries. In the developed countries when a patient comes to the doctor for the purpose of diagnosing his or her symptoms the doctor note down the various parameters and the symptoms that are being encountered. The doctor then utilizes the PHR of Public Health record system for referencing the symptoms that the patient is experiencing. This is highly useful in diagnosing symptoms that are overlapping with other diseases as the doctor does not have to guess between the two. This significantly eliminates the use of the trial and error method as the matching parameters can achieve accurate and fast referencing which can be highly useful. The deployment of

personal health records is very useful in reducing the pain and suffering the patient goes through while being subjected to the trial and error method. The trial and error method puts the patient in undue stress of a new medication along with the pain caused by the underlying medical condition that is not being treated. This also allows the doctor to keep an effective track record of the patient which makes it significantly useful for the future diagnosis of the patients.

PHR paradigm is not without its downsides as it contains highly sensitive and personally identifiable information inside. And various different hospitals in medical Institutions maintain their own records which are accessible through the use of data vendors of stakeholders that can contact and provide relevant information according to the requirement of the doctor. This is highly problematic as there is a low level of trust that is associated with the data vendors which does not allow for effective data sharing. Therefore, to improve the security of the proposed methodology and provide an effective access control mechanism for the public health records the blockchain distributed platform is highly appropriate for this application.

The blockchain platform was first introduced by a group of researchers for the purpose of organizing a distributed Framework for the purpose of storing timestamp documents. The blockchain was introduced for storing documents that cannot be altered in the form of a digital notary. The blockchain was not as popular until its utilization for creating a cryptocurrency. This allowed the blockchain platform to gain immense popularity and usage with researchers for the means of providing highly secure applications. This is due to the fact that the blockchain platform utilizes blocks of data and chaining them together with the help of hash keys to prevent any tempering on the data. This is the reason why it has been implemented in securing the various transactions that take place on the Bitcoin platform. This allows for effective implementation of the blockchain distributed Framework for the purpose of securing Public Health records and implementing an effective access control mechanism.

This research paper dedicates section 2 for analysis of past work as literature survey, section 3 deeply elaborates the proposed technique and whereas section 4 evaluates the performance of the system and finally section 5 concludes the paper with traces of future enhancement.

II. LITERATURE SURVEY

W. Dai [1] explain that data is one of the most important aspects of the economy right now as data has become increasingly valuable nowadays. Due to the importance of data in various algorithms and applications, there has been significant trading happening off data all over the world. But there are problems in data trading such as security issues that need to be addressed according to the author. Therefore, the researches in this Publication propose SDTE a platform for the data trading ecosystem that is secured through the use of blockchain. The proposed methodology has been experimented on to test its performance which has revealed that the platform is highly secure.

G. Magyar [2] states that the use of health information and electronic health records has been increasing in various countries. Earlier only developed countries had the resources to achieve electronic health record management, but due to various technological advancements, this has been achieved by developing countries also. Therefore, the increasing amount of data has various privacy and security issues related to the electronic health records. Thus, to improve the security the authors in this paper proposed the implementation of a blockchain framework that can effectively provide robust security to the platform.

Dr. M. Kumar [3] elaborate on the various storage techniques that have been utilized for storing different information specifically on the cloud platform. There has been increased usage of the cloud infrastructure for storage by organizations and various users. Due to the sensitive nature of different data that is stored on the cloud platform, there is a necessity to provide security to the data. The authors in this paper propose an effective utilization of the blockchain platform to secure the log storage in the cloud infrastructure. The proposed methodology has been tested extensively and reveals that it has improved security.

Z. Guan [4] discuss the issues that have been introduced due to the increase in the size of the internet platform and the amount of data that is being generated every day. This increasing amount of data has to be stored somewhere. This data cannot be stored just anywhere because there are security concerns related to the data as it might contain some sensitive and personally identifiable information success electronic or Public Health records. This type of data needs effective security on the platform and just cannot be stored anywhere. Therefore, the authors proposed and effective solution for securing the data through the use of blockchain. The proposed methodology can provide effective data trading without any loss of security.

M. Chowdhury [5] introduces the paradigm of electronic health records that have been utilized to provide a referencing mechanism for the doctors as well as keep a track of various treatments offered by the doctor to a particular patient. These health records are highly sensitive as they have personally identifiable information that is stored in them. The recent formation is highly valuable and can be used by attackers to get an unfair advantage over the user. Therefore, the authors have proposed an effective mechanism that provides notarization service for the storage of personal data through the use of the blockchain Framework. The proposed methodology has been tested for performance and reveals that it is significantly outperforming the traditional techniques.

M. Shen [6] explain that there has been an increase in the use of various devices an IoT devices that have been used to collect and store data. Most of these IoT devices are medical that perform data collection services heart rate monitoring BP etc. Therefore, it is important that the data on this device is to be secured effectively so that no data leakage can happen. Therefore, for this purpose, the authors have proposed an effective utilization of support vector machines along with the addition of the blockchain platform to achieve a privacy-preserving protocol that can be used to prevent data misuse from IoT devices.

U. Uchibeke [7] elaborate on the various advances made by technology that has been significant in achieving a lot of goals in which there are different fields across the world. This has led to a significant increase in the amount of data that is been generated every day. A large amount of data is collectively referred to as

big data and cannot be processed traditionally with traditional techniques. This data is also highly valuable as it contains Useful information that needs to be secured and provide an access control mechanism. Therefore, the authors in this paper provide an effective methodology for implementing an access control mechanism with the help of the blockchain Framework. The experimental analysis contributes that the proposed methodology is highly secure.

S. Ramamurthy [8] states that there has been an increase in the number of individual users and organizations that have been utilizing the cloud platform for storage. It is highly useful as cloud storage allows for increased convenience for the users as well as the organizations to access the data easily from anywhere. But the cloud storage platform stores the data on the remote server where the security of the data cannot be guaranteed. Some of the users utilize the storage for storing electronic health records that contain sensitive and personal information that cannot be leaked to the attackers. Before the authors in this paper proposed utilization of the blockchain platform for providing effective security to the data stored on the cloud platform. The cloud BC has been experimentally evaluated and produced satisfactory results.

W. Liang [9] expresses that there has been a significant increase in the number of IoT devices that have been utilized for supervision and data collection. Which of these devices are used in industries or in the Healthcare sector to monitor various events and parameters. The data that is stored and transmitted over the network from these devices is highly valuable and cannot be leaked to the attackers. Therefore, the authors in this paper propose an effective data transmission technique that works on a distributed framework through the implementation of the highly secure blockchain platform. The proposed methodology has been extensively tested for errors and resulted in a highly accurate and Secure methodology.

P. Urien [10] discusses the various improvement that has been made electronic over the past few years. These Electronic advancements have allowed the sensors and other devices to become highly cheap for implementation in everyday applications. Therefore, we see an increase in the usage of IoT devices to collect and transmit information mainly for the Healthcare industry. But there have not been significant improvements for transmitting this data which is highly private over unsecured networks. Therefore, the authors in this methodology proposed the utilization of the blockchain framework on these networks for securing the elements.

C. Cai [11] explains the increase in the number of cryptocurrencies ever since the Inception of the Bitcoin. An increasing number of cryptocurrencies that are utilizing the blockchain Framework that is the backbone of the Bitcoin network. This is since blockchain is a virtually tamper-proof and highly secure platform. Due to this fact, the authors in this paper propose an effective utilization of the blockchain platform in enabling security and trust in various applications other than cryptocurrency.

S. Sharma [12] discusses the significant increase in the number of users that have been utilizing the cloud platform to enable convenient and easy to access storage. This has been a valuable addition as it allows for the elimination of maintenance of local storage options. The cloud platform is an easy alternative to local storage which unlocks a lot of hidden potentials. But there have been instances where the data stored in the cloud platform is highly vulnerable to attacks and data leakages. Therefore, the authors in this paper propose the implementation of the blockchain platform to provide efficient and Temporary security to the data stored on the cloud platform.

X. Zheng [13] states that there has been a monument in an increase in the number of users utilizing the cloud platform. This is mostly due to the increased convenience and reliability of the cloud platform in maintaining the Stored data and making it accessible anywhere in the world with an Internet connection.

Due to this effect, many of the users store their personal information such as identification and personal health records on the cloud storage platform. This allows the sensitivity in personal information to be easily accessible in the case of a data leak or an attack on the cloud platform. Therefore, the authors propose to increase the security of the cloud service platform to be able to store personal health data without any worries. This is achieved through the implementation of the blockchain Framework that secures the data and does not allow any tempering to be done on the Stored data. The proposed methodology has been extensively tested and the results indicated that the proposed methodology has been significant in achieving robust security.

C. Harold expresses that there has been an increase in the number of internet of things devices due to the increase in the affordability of the electronic devices. The increasing affordability has seen among countries in the usage of these devices for various purposes mainly in the Healthcare sector for the collection and transmission of valuable health-related data.

IoT devices have allowed the remote monitoring of patients which has been a welcome change in the medical paradigm. But most of these data is highly personal and cannot be subjected to various leagues and attacks by people with malicious intent [14]. Thus, the authors in this paper proposed an effective utilization of the blockchain platform for deep reinforcement learning and providing security to the IoT data.

M. Singh states that the IoT devices have been police rating our environment due to the increasing affordability and ease of access that is been offered by this platform. The IoT devices have been significant in providing remote Healthcare to patients that are immobile or have difficulties in traveling to the hospital due to their old age or medical conditions. IoT devices allowed for the monitoring of various vital parameters of these patients from the comfort of their homes. But it has also exposed a lot of problems in how the data has been collected and transmitted to the doctor has not been highly secure [15]. Therefore, to improve the security of the IoT data the authors in this paper have proposed and innovative application of the blockchain Framework in achieving tamper-proof and Secure data collection. The experimental results indicate that the utilization of the blockchain platform has resulted in a significant increase in the security of the IoT data.

III PROPOSED METHODOLOGY

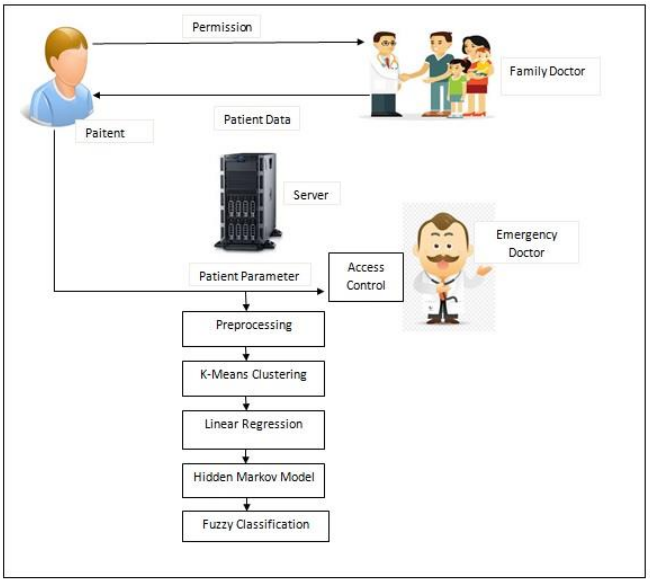


Figure 1: Proposed model System Overview

The presented technique to implement an emergency access control mechanism for secure sharing of Personal Health Records based on the distributed Blockchain framework has been outlined in the system overview diagram in figure 1 given above. There is a combination of steps that are required to implement the proposed methodology of providing effective access control mechanisms which are discussed in detail in the steps given below.

Step 1: *Data Seeking and Data preprocessing* – This is the first step in the proposed methodology which is one of the basic procedures that are followed to implement these features. When the doctor encounters a patient with a severe condition and a set of symptoms, the doctor seeks more information about the person's medical history and other related illnesses and diagnosis done by other doctors. The doctor utilizes the system to input the various known parameters about the patient and the symptoms and fires a query to the Data venders.

The query on reaching the Data Vender is then forwarded to the many Data Providers to seek the diagnosis Details of the current symptoms of the Patient. On receival of the Query on Data Provider's end it is subjected to preprocessing process which contain mainly four steps as detailed below.

Special Symbol Removal- This step performs the elimination of the special symbols from a query that is passed in the form of a string. Various special symbols are shredded in this process such as !,?,., etc.

Tokenization- Tokenization is the process of segregating a string into different parts in the space to be stored in the form of a well-indexed string. This is done to facilitate the conversion of the string eventually into an array which allows for much efficient and effective handling of the string.

Stopword Removal – Stopword is words that are used in the English language to facilitate the conjunction of letters and joining of sentences. These words are not as important as they do not contain explicit meaning other than just combining the sentences or words.

This is the motivation behind this process and keeping this in mind, the stop words encountered in the English language are eliminated in this process. This process removes the redundant parts of the sentence making it easier to process without changing the meaning of the query.

For example, if a phrase was there like: we are going to school. After Stopword removal it transforms into going school. It can be observed in this example that the removal of the Stopword has not changed the meaning of the sentence in any form.

Stemming- Stemming is the process by which a word is transformed into its root form. This is specifically done to reduce the redundancy in the data and effectively achieve significant improvements in the processing of the reduced data. This is achieved in the presented technique by utilization of the string replacement methodology that replaces the undesired postfix in the query by the relevant string.

For example, eating will be transformed into eat after substring "ing" is replaced with an empty character. It can be observed that the difference between eating and eat does not deviate from the core meaning which remains intact.

Step2: K means Clustering- The preprocessed Query string is utilized to cluster the Provider data. This Provider dataset is collected from the URL: https://data.world/arvin6/medical-records-10-yrs.

This dataset is all about the collection of clinical practice details of the different patients with different age, sex, symptoms, diseases and the details about the preliminary treatment.

The preprocessed query is subject to estimate the count of the query words in each of the treatment details of the different patients in the dataset which is denoted by the attribute called soapnotes. Then a double dimension list is created which contain two columns like encounter id and count to call it as count list.

This count list is then subjected to estimate the cluster using the K_means Clustering Technique, which contains mainly 6 Steps as described below.

Distance Evaluation- This is the initial step of the K means clustering , where the counter list is subjected to estimate the distance of each of its row with all other rows for the value of the count, thereby obtained all other rows mean is calculated to append at the end of the respective Row. The distance is estimated using the Euclidean Distance Equation as mentioned in 1.

$$\boldsymbol{ED} = \sqrt{(\boldsymbol{x1} - \boldsymbol{x2})^2 + (\boldsymbol{y1} - \boldsymbol{y2})^2} \text{-------} \quad (1)$$

$$M_{ED} = \sum_{k=0}^{n} ED\,/n \text{\_\_\_\_\_} (2)$$

Where
x1, x2,y1,y2 are the entities of the Rows.
E_D- Euclidean distance of a specific row
M_{ED} - Mean Euclidean Distance
n= Number of Rows

Sorting- The obtained list, which is appended with the Row Euclidean distance is then sorted based on this Euclidean distance using the bubble Sort.

Data Point Selection – The Sorted list, then used to select K Random rows that eventually denote the number of the clusters to be produced.

Centroid Evaluation- The obtained Data point rows are used to collect the respective Row Euclidean distance to call them as centroid.

Boundary Evaluation- In this phase of clustering each of the centroid Euclidean distance is then tend to find the boundaries of the clusters. This boundary can be estimated by subtracting the centroid Euclidean distance with the mean Euclidean distance of the whole list to form the lower limit of the cluster boundary.

In the same way the upper boundary of the cluster can be formed by adding the mean Euclidean distance to the centroid distance. So these Lower and upper boundaries are stored in a double dimension list to call it as boundary list.

Cluster Formation- This is the last point of the K- mean clustering, where each of the rows are segregated into the specific clusters based on their Row Euclidean Distance. This is done by comparing Row Euclidean Distance with the each cluster boundary range like min and max to form K clusters. This process of Cluster formation can be shown in the below algorithm 1.

ALGORITHM 1: CLUSTER FORMATION

//Input : Boundary List B_L, ED_L [Euclidean Dstance List]
//Output: Cluster C_L
1: Start
2: **FOR** k=0 to Size of B_L
3: $SC_L = \emptyset$ [Single Cluster List]
4: $TP_{LST} = B_{L[k]}$ [TP_{LST} = Temporary Set]
5: MIN= $TP_{LST[0]}$, MAX= $TP_{LST[1]}$
6: **For** j=0 to size of ED_L
7: $T_{LST} = ED_{L[j]}$ T_{LST} = Temporary Set]
8: ED_{SCRE}= $TMP_{LST[2]}$ [Euclidean Score]
9: **IF**(ED_{SCRE} >= MIN AND ED_{SCRE} <=MAX)
10: $SC_L = SC_L + T_{LST}$
11: **End IF**
12: **End For**
13: $C_L = C_L + SC_L$
14: $SC_L = \emptyset$
15: **END For**
16: return C_L
17: **Stop**

Step 3: Linear Regression – Here in this step all the preprocessed query words tend to form an array according to the position of the words to call it as x [] . If there are 5 preprocessed words from the user quey, then x [] forms as [1,2,3,4,5]. On the other hand the Y[ ] forms an array for the each stored treatment history string with respect to the fired Query. If the fired query words match with some of the words of the stored instance, then Y[] becomes as [1,0,0,4,5]. Then these two arrays are subjected to estimate the Slope intercept by using the equation 3.

The linear regression is calculated using equation 3 given below.

Y=Mx+B ____ (3)

Where:
x = how far up (Array of Attribute)
M = Slope or Gradient (how steep the line is)
B = the Y Intercept (where the line crosses the Y axis)
Y=Intercept value

Then, based on the threshold intercept values for each of the instances in the formed clusters, each of the clusters is assigned a rank. Based on this rank clusters are sorted in descending order to feed to the next model of Hidden Markov model.

Step 4: Hidden Markov Model- The obtained sorted cluster regression list is subjected to fetch top two clusters as they contain best matching data for the fired query by the Provider. Then these two clusters are merged to get a single list to input to the Hidden Markov Model and it is termed as the HMM input list. The maximum and minimum values of the query count is evaluated from this list to call them as Target 1 and Target 2 variable of HMM.

10 Random weights in between the ranges 0 to 1 are assigned to the entities like b1,b2,W1,W2,W3,W4,W5,W6,W7,W8 by using the Random class of the Java. Where B1 and b2 are referred as the bias weights.

The HMM list is used to estimate the hidden layer for all its rows using the following equation.

X=(AT1* W1)+(AT2*w2)+b1 _____(4)

$H_{LV}= \frac{1}{(1+\exp(-X)}$ _____(5)

Where, AT1 and AT2 are the query matching count and Euclidean distance respectively. Then Equation 5 represents the sigmoid function of the HMM.

In the same way the output layer also been calculated to estimate the target error with the variables Target1 and Target 2 to get the Probability values of the each row of the provider data using other weight entities.

Step 5: Fuzzy Classification -Then these rows are optimized for the 99% accuracy for the access control mechanism value of select the best matching rows. Then these row's prediction value frequency will decide the access control level for each row, according to their different attributes. This row's prediction value is decided based on the Fuzzy Crisp values which are distributed in the segment of VERY LOW, LOW, MEDIUM, HIGH and VERY HIGH. Once the access control is decided for the respective attributes, then they will be hidden with some characters like '*' to return back to the Data vendor using the blockchain mechanism.

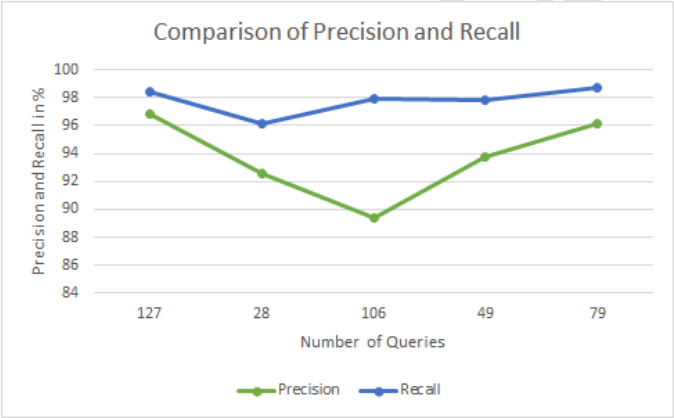
Step 6: Blockchain- Once the access control data are ready, then it is considered as the block body and this will be sent to the

MD5 hashing algorithm to provide the hash key of the block body. This hash key is known as the head key which is then optimized for constrained length to send the data with the secured way through the Email to the Data Seeker through the vendor. This process eventually maintains the data security for the applied access control.

IV. RESULTS AND DISCUSSIONS

The presented technique for the purpose of establishing an efficient access control mechanism has been developed on the NetBeans IDE in the Java programming language. For the demonstration of the proposed methodology, 3 development machines are required which are equipped with Intel i5 processor handling the processing requirements along with 500 GB of storage and 4GB of primary memory as RAM. The MySQL database server fulfilled the database responsibilities along with a D-Link WIFI router for networking.

Extensive Experimentation was conducted to evaluate the performance metrics of the proposed methodology. For the purpose of analyzing the accuracy of the proposed methodology, the Precision and Recall analysis technique was applied which accurately outlines the performance metrics of the proposed technique. The performance metrics were evaluated to determine



that the access control mechanism based on the distributed blockchain framework proposed in this paper is performing adequately.

Performance Evaluation based on Precision and Recall

Precision and Recall can provide valuable information related to the performance of the proposed methodology. These parameters are one of the most accurate and insightful techniques that are utilized to extract the absolute performance of the system. Precision in this evaluation extracts the relative accuracy of the presented technique by evaluating the accurate values of the level of precision in the system.

Precision in this approach is being evaluated as the ratio of the combined sum of all the queries that have been matched to the number of accurate predictions performed. Therefore, the evaluation of the values of precision allows for an in-depth analysis of the effectiveness of the proposed methodology.

The Recall parameters are utilized for extracting the absolute accuracy of the system which is quite different from the precision parameters. The Recall parameters are extracted through the ratio of the number of accurate predictions for the given query matched versus the total number of inaccurate predictions for the given query matched. This analysis technique gives valuable insight as it extracts the absolute accuracy of the system. Precision and recall are mathematically elaborated in the equations detailed below.

Precision can be concisely explained as below

✓A = The number of accurate Access control applied for the given query using HMM

✓B= The number of inaccurate Access control applied for the given query using HMM

No of Dataset Instances	Accurate Predictions (A)	Inaccurate Predictions (B)	Accurate Predictions not done (C)	Precision	Recall
127	121	4	2	96.8	98.37398374
28	25	2	1	92.59259259	96.15384615
106	93	11	2	89.42307692	97.89473684
49	45	3	1	93.75	97.82608696
79	75	3	1	96.15384615	98.68421053

✓C = The number of accurate Access control not applied for the given query using HMM

So, precision can be defined as

Precision = (A / (A+ B)) *100
Recall = (A / (A+ C)) *100

The equations given above are used for performing intensive experimentation on the proposed methodology through the evaluation of the results of the Artificial Neural Network module. The evaluation of the experimental results is detailed in table 1, given below.

Table 1: Precision and Recall Measurement Table for the performance of HMM

Figure 2: Comparison of Precision and Recall for the performance of HMM

The graph given above indicates that the Artificial Neural Network in the proposed system attains expected parameters of precision and recall for the purpose of access control mechanism. The model achieved the precision of 93.74% and Recall of 97.78% and this precision and recall parameters indicate that the HMM module executes with utmost accuracy and efficiency in the proposed methodology.

These experimentation results specify that the HMM analysis module is being implemented correctly in the proposed methodology and is extremely accurate in its performance. The Hidden Markov Model is one of the principal components in the proposed methodology and the successful and accurate execution of this module significantly contributes to the performance of the proposed Access control mechanism which further improves the process.

The experimental outcomes attained for the Precision and Recall of the presented methodology are compared to [16] which implements an innovative technique reliant on the K Nearest Neighbors clustering. The outcomes demonstrate that our approach outperforms the one defined in [16].

In [16] the model is attributed to the K Nearest Neighbor algorithm as it amplifies the number of iterations which increase the complexity of the system. The proposed methodology, on the other hand, deploys the access control search using the combination of K- means and Hidden Markov Model for searching the data, which improves the accuracy which is evident in the results. This also reduces the overall complexity of the system considerably. The comparison is tabulated in the Table 2 and the subsequent graph plotted from these values in Figure 3.

Table 2: Comparative Results of Precision and Recall between KNN Search and K means HMM Access Control Search

Methodology	Average Precision	Average Recall
KNN Search	84.5	97
K Means HMM Access Control Search	93.74	97.78

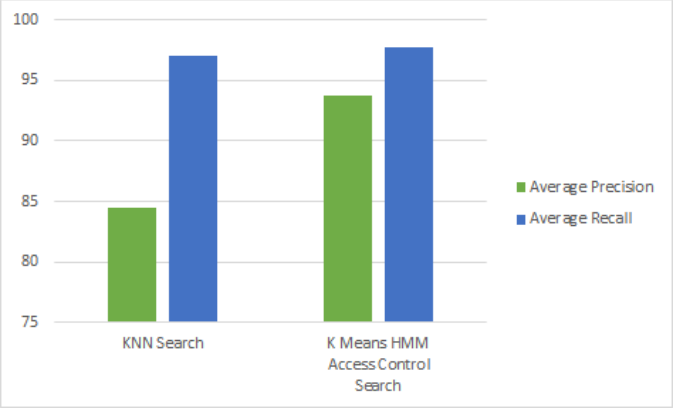


Figure 3: Comparative Results of Precision and Recall, KNN Search V/s K means HMM Access Control Search.

V. CONCLUSION AND FUTURE SCOPE

The public health record paradigm is a highly valuable resource that contains useful information that can be utilized to effectively treat a patient. Even though most of the medical institutions have their PHR storage mechanism, some of the times these hospitals encounter a patient whose symptoms cannot be referenced in their PHR database. Therefore, they employ the use of a data aggregator or vendor that provides the hospital with PHR from a different medical institution related to their query. As there is a lack of trust between the data vendors and the medical institutions due to the fact that the public health records contain sensitive personal information that cannot be shared without effective implementation of an access control mechanism. Therefore, to secure the public health records and provide an effective access control mechanism a blockchain-based methodology has been elaborated in this publication. The decentralized framework of the blockchain platform along with the utilization of k-means clustering and Regression analysis catalyzed by the Hidden Markov model and Fuzzy Classification attains an effective and Secure access control system efficiently.

For future research directions, the proposed technique can be implemented in a real-time scenario. The access control technique could also be implemented in various applications across different fields such as agriculture and law.

REFERENCES

[1] Weiqi Dai, Chunkai Dai, Kim-Kwang Raymond Choo, Changze Cui, Deiqing Zou, and Hai Jin," SDTE: A Secure Blockchain-based Data Trading Ecosystem" 1556-6013 (c) IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission 2019.

[2] Gábor Magyar," Blockchain: solving the privacy and research availability tradeoff for EHR data" IEEE 30th Jubilee Neumann Colloquium • November 24-25, 2017 • Budapest, Hungary 2017.

[3] Dr. Manish Kumar, Ashish Kumar Singh, Dr. T V Suresh Kumar," Secure Log Storage Using Blockchain and Cloud Infrastructure", IISC, Bengaluru, India July 10-12, 2018.

[4] Zhangshuang Guan, Xiaobei Shao and Zhiguo Wa," Secure, Fair and Efficient Data Trading without Third Party Using Blockchain" IEEE Confs on Internet of Things, Green Computing and Communications, Cyber, Physical and Social Computing 2018.

[5] Mohammad Jabed Morshed Chowdhury, Alan Colman, Muhammad Ashad Kabir, Jun Han, and Paul Sarda," Blockchain as a Notarization Service for Data Sharing with Personal Data Store" 17th IEEE International Conference on Trust, Security and Privacy in Computing and Communications 2018.

[6] Meng Shen, Xiangyun Tang, Liehuang Zhu, Xiaojiang Du, and Mohsen Guizani, "Privacy-Preserving Support Vector Machine Training over Blockchain-Based Encrypted IoT Data in Smart Cities" 2327-4662 (c) IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission in 2018.

[7] Uchi Ugobame Uchibeke, Sara Hosseinzadeh Kassani, Kevin A. Schneider, Ralph Deters," Blockchain access control Ecosystem for Big Data security" IEEE Confs on Internet of Things, Green Computing and Communications, Cyber, Physical and Social Computing, 2018.

[8] S. Ramamoorthy, B. Baranidharan," CloudBC-A Secure Cloud Data Access Management system" 978-1-5386-9371-1/19/$31.00c 2019 IEEE.

[9] W. Liang, M. Tang, J. Long, X. Peng, J. Xu, and K. C. Li, "A Secure Fabric Blockchain-based Data Transmission Technique for Industrial Internet-of-Things" IEEE. Personal use is permitted, but republication/redistribution requires IEEE 2018.

[10] Pascal Urie," Towards Secure Elements for Trusted Transactions in Blockchain and Blockchain IoT (BIoT) Platforms" 2016, Gainesville, FL, USA.

[11] Chengjun Cai, Huayi Duan, and Cong Wang," Tutorial: Building Secure and Trustworthy Blockchain Applications" 2018 IEEE Secure Development Conference.

[12] Shweta Gaur Sharma, Dr. Laxmi Ahuja," Building Secure Infrastructure for Cloud Computing using Blockchain" IEEE Xplore Compliant Part Number: CFP18K74-ART; ISBN:978-1-5386-2842-3.

[13] Xiaochen Zheng Raghava Rao Mukkamala, Ravi Vatrapu, Joaquin Ordieres-Mer," Blockchain-based Personal Health Data Sharing System Using Cloud Storage" IEEE 20th International Conference on e-Health Networking, Applications and Services (HealthOP) 2018.

[14] C. Harold et al, "Blockchain-enabled Data Collection and Sharing for Industrial IoT with Deep Reinforcement Learning", IEEE Transactions on Industrial Informatics, 2018.

[15] M. Singh et al, "Blockchain: A Game Changer for Securing IoT Data", IEEE 4th World Forum on the Internet of Things (WF-IoT), 2018.

[16] Cengiz Orencik, Erkay Savasy and Mahmoud Alewiwiz, " A United Framework for Secure Search Over Encrypted Cloud Data ", IACR Cryptology ePrint Archive 2017.