Blockchain and Machine Learning in Healthcare

Mr.Aryan K.[1], Mr.Vimpy P.[2], Mr.Gopal K.B.[3], Mr. Bal K.B.[4], Mrs.P Sushmita S[5], Dr.Aruna M G[6], Dr. Malatesh S H[7]

[1,2,3,4]*Student, Department of CSE, M S Engineering College, Bangalore, India*
[5]*Assistant Professor, Department of CSE, M S Engineering College, Bangalore, India*
[6]*Associate Professor, Department of CSE, M S Engineering College, Bangalore, India*
[7]*Professor and HOD, Department of CSE, M S Engineering College, Bangalore, India*

Abstract - **Now days so much inaccurate information and fraud in the healthcare industry, so we need to explore a secure and trusting environment to better the system. Nonfinancial blockchain provides secure and immutable data sharing to data management in the diverse medical workflow. The data breaches reached a record high and for the past few years, the healthcare field has had the second highest number of breaches compared to other sectors. The frequency of medical data breaches has been highly concerning. In particular, armed with someone's medical information, thieves can easily commit medical identity theft to get drug prescriptions, or make false insurance claims under the victim's name. Medical data mostly comes with personal and private information which includes Social Security Numbers (SSNs), as well as financial information. This is done using trained algorithms. After storing we use Blockchain for data sharing and its reliability. For securing the medical data use Decentralized server for secure storage of medical data.**

Index Terms - **Blockchain, SSN.**

I.INTRODUCTION

In Blockchain like a distributed chain, singular transactions are encrypted into blocks by the applicable encryption, added to the ledger and never deleted. The information in Blockchain is verified at basic level by a linked list of encoded exchanges that utilizes a hash.

The hash function generates a hash by encrypting the information fed in Blockchain. It provides the shape of foundation of a decentralized medicinal service stage shared by the patients and suppliers, acting as an interface to the patient's record.

In healthcare, Blockchain has a wide range of applications and functions. The ledger technology helps healthcare researchers uncover genetic code by facilitating the secure transfer of patient medical records, managing the drug supply chain, and facilitating the safe transfer of patient records.

With blockchain technology, health care system will able to store medical records confidentially, updating patient data across multiple facilities and locations in real time and with security Blockchain technology could also increasingly support monetary transitions between patient and care center, especially at the level of micro payments.

II.EXISTING SYSTEM

Meditab, a software company, claims itself as one of the foremost electronic medical records software creators for medical institutions. This company includes electronic faxes and this is method is still used to share patient data to other users. But this method of sharing data has proved to be insecure and unreliable making it very less preferable.

Spider Silk, a cybersecurity company in Dubai, had a fax server, which was running an Elastic search database which included not less than 6 million health care records. The server did not even had passwords for security, which indirectly granted access to everyone and thus anyone could read the transmitted faxes in real-time .The fax that were sent consist of many recognizable information regarding patients such as their medical history, treatment undertaken in the past, their Social Security Numbers and other records.

Disadvantages of Existing System:
There is no security of patients medical data.
There is no efficient access control mechanism while sharing our medical data in cloud.

To provide the efficient access control data owner relies on the third party.

The below table I gives a literature summary about the papers being reviewed for this project work

TABLE I: LITERATURE SUMMARY

Sl. No.	Paper Title	Proposed system	Advantages
1	Blockchain-Based Medical Records Secure Storage and Medical Service Framework	We design a storage scheme to manage personal medical data based on blockchain and cloud storage. Furthermore, a service framework for sharing medical records is described.	It provides secure storage of medical service
2	Blockchain: Solving the privacy and research availability trade-off for EHR data.	We introduce how the blockchain technology can help solving the problem of secure data storing and ensuring data availability at the same time.	It provides secure storage of medical service
3	The biggest medical data breaches in history	The effects of medical insurance fraud on victims include large financial losses, poor credit rating.	The effects of medical insurance fraud on victims.
4	Supervised Machine Learning: A Review of Classification Techniques	Hackers gained access to parts of its network that contained the protected health information of 2.4 million individuals.	We have studied about ML algorithms.

III.PROPOSED SYSTEM

In our proposed model we have used both Blockchain technology and Machine Learning Algorithms to provide a better solution in terms of security.

By using Machine Learning, we provide additional features which can be the base of ideas for further implementations on this subject.

Machine Learning is based on the concept of centralization of data, while Blockchain technology uses decentralization of data to provide high security.

IV. TRAINED DATA AND PRE-PROCESSING

Get all the news and updates from Deccan Chronicles, Times of India, The Hindu and others. Global media outlets have newscast programmers that cover international news 24 hours a day, seven days a week.

A. Pre- Processing

Prior to training and data evaluation using machine learning, data processing is a normal first step. Algorithms for machine learning are always as useful as information you fed them. It is important to format correct data and to include relevant items so that they are consistent enough to produce best outcomes possible. Stop word removal, tokenization, lower case and punctuation removal are all examples of data refinement. This allows us to reduce the size of the real data by removing irrelevant information. We created a simple processing function for each document to remove punctuation and non-letter characters, followed by the letter case in the document was lowered. Make different steps to clean text (remove all non- alphanumeric characters, delete stop words, delete missing rows, etc.).

A. Feature Extraction

Feature selection is the method of reduction that reduces an original batch of actual data to even more controllable computing categories. Ngram are a type of grammatical unit. Every news channel's word bag is mined for unigrams and bigrams. Tfidf Vectorizer is used to score the relative importance words in a document. Count Vectorizer is used for creating vectors that have a dimensionality equal to the size of our vocabulary, and if text data features vocab word, we will put a one in that dimension. Result of this will be very large vectors, if we use them on real text data, however, we will get very accurate counts of the word content of our text data.

Trained data:

The idea to use data from training in machine learning programs is a simple idea, however the way such innovations work is also really simple. The training process is an initial piece of facts used to help a program to realize how computational intelligence technologies can be applied and specialized results produced.

Prediction:

Usually, a data set is separated into a training and test set. The majority of the data is used for training, while only a small portion of the data is used for testing. Using web application module to display the interface for taking input from the user, by using the trained data machine it can predict output and display it to the user. Test data is also applied for feature extraction and preprocessing.

Algorithms:

For the prediction, multiple supervised learning algorithms are trained using the training set, after which using the testing set performance evaluation occurs. The algorithms are:

A. Random Forest
STEP 1: START
STEP 2: SPLIT dataset into 67 percent training set, 33 percent testing set
STEP 3: FOR train dataset
CALL RFClassifier TRAIN RFClassifier
STEP 4: FOR test dataset
CALL RFClassifier PREDICT the label COMPUTE AccuracyScore SAVE AccuracyScore DISPLAY ConfusionMatrix
STEP 5: STOP

B . KNN Classifier
Pseudo code
Procedure Train()
 // Input: train set, test set
// Output: Trained model
Step 1: Read Train set and test set
Step 2: Build KNN classifier
Step 3: Train the model using fit()
Step 4: Performance Graph Returned Trained Model

D. Support Vector Machine(SVM)
STEP 1: START
STEP 2: SPLIT dataset into 67 percent training set, 33 percent testing set
STEP 3: FOR train dataset CALL SVMClassifier TRAIN SVMClassifier
STEP 4: FOR test dataset CALL SVMClassifier PREDICT the label COMPUTE AccuracyScore DISPLAY ConfusionMatrix
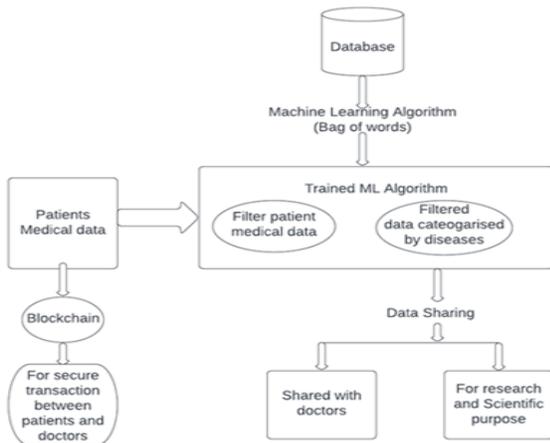STEP 5: STOP

V. SYSTEM ARCHITECTURE

Fig:1 System Architecture

In the above fig1 represent the System Architecture of Blockchain structure is used in System. The second part of the structure works on Ethereum and performs all application and services. Medical information is very sensitive and personal so a closed Blockchain such as Hyperledger Fabric helps in retaining necessary privacy required. Majorly blockchains are classified as public Blockchains and permissioned Blockchains.

In the proposed model, we use the "Bag of Words" algorithm which will extract only the required dataset and ignore the various other things like the Name, Age, Address and other personal details of the patient to maintain the privacy In the proposed model, we use the "Bag of Words" algorithm which will extract only the required dataset and ignore the various other things like the Name, Age, Address and other personal details of the patient to maintain the privacy trouble and provide data trends, collected data must go through a cleaning process. This process includes data transformation, metadata enrichment exploitation, exploration or removing unnecessary or invalid data that is not required to obtain data trends, then data validation. Then we are applying KNN algorithm to get the recommendation.

VI. RESULTS AND DISCUSSION

The system was created using Windows 10 as well as a 64-bit processor with 8 GB of RAM. The model implemented with the help of Python v3.6.
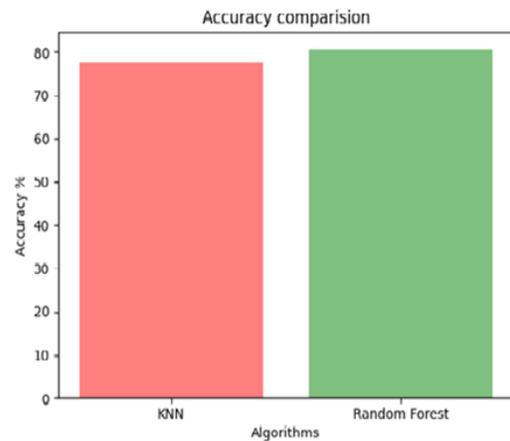

Fig:2 Accuracy comparison between algorithm
From the values calculated in the confusion matrix, an accuracy graph (Fig2) is generated for each algorithm

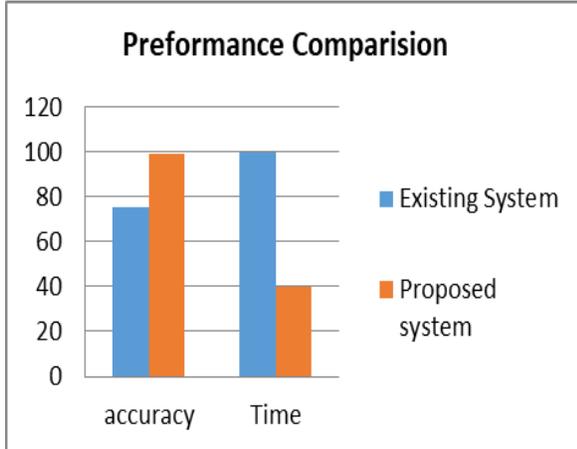for comparison for best algorithm with highest accuracy.

Fig:3 Performance Comparision

This phase involves the evaluation of performance of machine learning algorithms on training and building the model, and then predicting the label of the news article given by the user. The impact is measured in average accuracy and time taken to train and predict.

The above Fig 3 shows the accuracy determined by accuracy score of the ML models, which is measured in percentage. The average time taken is determined by comparing the evaluation time taken for training and prediction by a model.
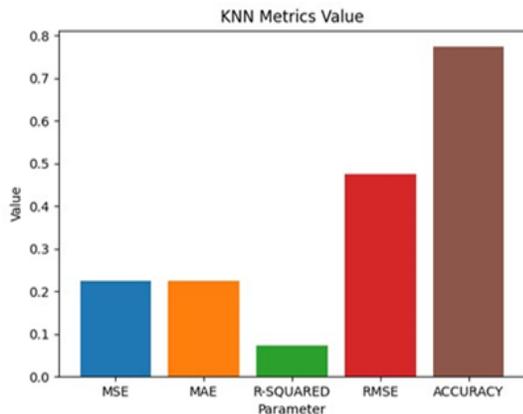

Fig:4 KNN Metrics Value

Fig:4 Fig KNN metrics value are generated for calculation of accuracy of Algorithm The above fig6 shows the KNN Metrics value generated after training an algorithm using the pre-processed data. MSE is Mean Square Error and MAE is Mean Absolute Error. RMSE is Root Mean Square Error and accuracy is calculated for KNN algorithm.

SCREENSHOT 1

Fig:5 Main screen of web application

The above Fig 5 shows the main screen of web application. From this page user will able to start performing task by register through it.

SCREENSHOT 2

Fig:6 Adding patient details

The above Fig 6 contains patient details to feed in the system. It is having details of patients such as patient id, age and their disease related information.

VII. CONCLUSION

Blockchain Technology has been evolving with time, financial sectors are already using Blockchain keeping in mind the unparalleled advantages it offers, with the significant increase in health data breach through hacking, and application of Blockchain for security becomes important and imperative. It will not be wrong to say that Blockchain based Health care model is the future in healthcare sector and has the potential to change the way health care records are managed and secured. With the emergence of 5-G networks and faster than ever data transfer facilities it will encourage advancement of Machine Learning, Blockchain and

other data-based techniques in various sectors including Healthcare. As this new technology ecosystem emerges, Blockchain promises significant improvements in managing patient health records. Continuous efforts are being made to increase the accuracy of wearable health tracking devices and if these data could provide more accurate and reliable results there will be brighter chances of integrating these devices with the health records to provide more information and also share some of these medical data securely with authorized doctor without actually visiting. The ideas based on implementing Blockchain and Machine Learning is not much explored.

REFERENCE

[1] Zack Whittaker (2019, March), "A huge trove of medical records and prescriptions found exposed.", a.[Online]. Available: https://techcrunch.com/2019/03/17/medical-health-data-leak/

[2] Centres for Medicare & Medicaid Services," Electronic Health Records", [Online]. Available: b.https://www.cms.gov›Medicare›E-Health› EHealth Records

[3] S. B. Kotsiantis,Supervised Machine Learning: A Review of Classification Techniques, Department of Computer Science and Technology University of Peloponnese, Greece.

[4] Izhar Alam, 'Patient Data Recommendation' [Data set] 2020.Available: https://www.kaggle.com/ izharalam150/data-recommendation.

[5] G.Magyar,Blockchain: Solving the privacy and research availability tradeoff for EHR data: A new disruptive technology in health data management, Budapest, Hungary,24-25 Nov. 2017

[6] Aimee O'Driscoll (2019, July), "The biggest medical data breaches in history", [Online].Available: https://www.comparitech.com/blog/vpn-privacy/biggest-medical-data-breaches/.

[7] K.T. Win. A review of security of electronic health records. Electronic Health Records: security, safety and archiving, 34, 2005.

[8] Chen, Y.; Ding, S.; Xu, Z.; Zheng, H.; Yang, S. "Blockchain-Based Medical Records Secure Storage and Medical Service Framework", Journal of Medical Systems, vol.43, no. 5, 2018.

[9] K.T. Win. A review of security of electronic health records. Electronic Health Records: security, safety and archiving, 34, 2005

[10] S. B. Kotsiantis,Supervised Machine Learning: A Review of Classification Techniques, Department of Computer Science and Technology University of Peloponnese, Greece.