Integrating Blockchain and Machine Learning: A Framework for Enhancing Healthcare

Neha Mishra[1], Chetan Agrawal[2], Rashi Yadav[3]
Dept. of CSE, Radharaman Institute of Technology & Science, Bhopal, India[1, 2, 3]
mneha1078@gmail.com[1], chetan.agrawal12@gmail.com[2], rashi6yadav@gmail.com[3]

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.*

Keywords:- Blockchain, SSN.

Introduction

Blockchain and ML technologies offer unique benefits in healthcare, permitting organizations to implement them. Such benefits encompass preserving and exchanging medical evidence, predicting illness risk, picture categorization in magnetic resonance imaging (MRI) scans, and so on. Modern activities were actually done by medical experts and staff members, but, since the emergence of these innovations, activities are now executed by multiple computers [1]. ML is indeed a form of AI in which systems are permitted to train through the use of specialized strategies [2]. Learners evolve through particular processes and instruction, and machine intelligence is similarly trained to acquire segregated data that will eventually allow the system to run picture detection, categorization, and recognition [3]. Hence the more the data, the more accurate the outcomes; therefore ML techniques are the significant methods to be employed in the proposed technique. On either end, blockchain can be characterized like a record that exists within a system and maintains specific sources of evidence. To put it another way, a blockchain is indeed an electronic way of preserving in- formation which cannot be tampered with, hacked, or defrauded [4]. As a result, blockchain and machine learning are both distinct techniques, with blockchain being employed to safely store some sorts of data and machine learning being utilized for other objectives (defined beneath). In the medical industry, machine learning is utilized to recognize and classify medical pictures, allowing for more diagnosis. The healthcare system, for example, uses MRI scan images to assess that there may not be a danger of serious illness. The ML technologies can recognize pictures from chest X-rays or MRIs. Following that, the equipment analyzes the photos using the classification model, and lastly, an outcome with the reasonable degree of certainty is generated. The more dataset it has, the more appropriate the outcome will be. As a result, medical professionals no longer justify and classify MRI images on their own; instead, ML machines perform the identification. The accuracy of classification ranges from 85 to 99 percentage which has piqued the curiosity among healthcare providers. Aside from the benefits, there have been a couple of obstacles that will be investigated in this study. Blockchain, at the other side, is often used to securely store medical evidence that cannot be manipulated or changed by anyone. The conventional pencil and board methodology collecting and distribution are being replaced by a digital information gathering and utilization approach in the medical industry. As a result, a confidential scheme is essential to gather and process medical data. Inventors have observed blockchain solutions that are similar to Bitcoins [5, 6]. This approach is shown to have a high level of immutability when it comes to securing medical data. Furthermore, after obtaining agreement from clients and their associated practitioners, medical information can be collected with other parties. The blockchain is classified into three categories: personal, open, and mixed. Due to its ownership and surveillance by several organizations, the hybrid system can be considered the most secured digital categorization among them [7]. As a result, distinct chain structures have different benefits and drawbacks.

Rest of the paper is organized as follow: in section II literature review are explained, in section III proposed method is represented, section IV explains training & preprocessing, in section V represents system architecture, section VI shows results of proposed method and finally we conclude our work in section VII.

Literature Review

This section presents an overview of the relevant work conducted by various authors in the field of EHR Management in Blockchain-Cloud integration using various approaches.

Zhang et al. [8] suggested a blockchain-based privacy-preserving e-health system to solve the security issues pursuing the current cloud-assisted EHRs. This study discussed the dangers of EHRs being tampered with or leaked by unscrupulous medical professionals or cloud storage service providers. The authors offer pairing-based cryptography to create immutable records incorporated into blockchain transactions, therefore protecting the privacy of electronic health information. The electronic health records of the patients are protected from unauthorized changes and may be verified with this method. The study also covers the development of safe payment protocols utilizing blockchain-based smart contracts for trustworthy payments between patients and hospitals for diagnostic and storage services. Validation via security analysis and performance assessment demonstrates the efficacy and low computational cost of the proposed approach.

Ismail et al. [9] explored healthcare blockchain-cloud integration (BcC), or the use of blockchain technology with cloud computing. The study utilized the scalability and effectiveness of cloud computing in combination with the decentralized nature of blockchain to address security and privacy issues. In the study, the authors surveyed all aspects of BcC integration in healthcare, including the various architectures, apps, and development tools currently in use. Challenges, solutions, and plans for the future of the field were also discussed. The study's findings can aid the healthcare sector in improving patient care through the use of new data management systems.

Velmurugadass et al. [10] developed a new method of criminal investigation that makes use of blockchain technology. Mobile nodes, an open-flow switch, blockchain-based controllers, a cloud server, an Authentication Server (AS), and investigators are all parts of the framework's Cloud-based Software Defined

Network (SDN). For information safety, the system makes use of cryptographic algorithms and cryptographic hash functions based on the Elliptic Curve Integrated Encryption Scheme (ECIES). Based on a Logical Graph of Evidence (LGoE), the investigators carry out several tasks, such as identification, evidence collecting, analysis, and report preparation. Response time, accuracy, throughput, and security characteristics were all significantly enhanced in experimental findings. Criminal investigations and evidence management might benefit from the integration of blockchain, SDN, and encryption methods, as demonstrated by this study.

Benil and Jasper [11] presented a novel approach to deal with EHR security concerns called Elliptical Curve Certificateless Aggregate Cryptography Signature (EC-ACS). The system safeguards private medical records and prevents unauthorized access by utilizing approved blockchain technology. Medical records are encrypted using Elliptic Curve Cryptography (ECC), and digital signatures are generated using the Certificateless Aggregate Signature (CAS) approach, both of which help to make cloud storage and sharing possible. The suggested technique safeguards the cloud-based healthcare system by enforcing confidentiality and preventing illegal access. Integrating blockchain technology further ensures the integrity, traceability, and secure cloud storage of medical records.

Shi et al. [16] performed a comprehensive literature assessment of blockchain options for EHR systems with an emphasis on security and privacy. The study set out to investigate blockchain's potential utility in EHR systems and to spot gaps and openings in the field. The writers emphasized the rising interest in blockchain's revolutionary potential in the healthcare industry. They did, however, note that several obstacles remain in the way of the complete integration of blockchain technology with traditional EHR systems. Some of these difficulties were reviewed, and potential topics for further study were highlighted, including the Internet of Things (IoT), big data, ML, and edge computing. The aging society may greatly benefit from the creation of next-generation EHR systems, which the authors of this study intend to facilitate.

Bhattacharya et al. [12] proposed Blockchain-Based Deep Learning as a Service (BinDaaS) as a framework to solve the issues of confidentiality, security, as well as data integrity in EHRs. The system combines blockchain technology with deep learning algorithms to enable the safe transfer of EHR data between different medical institutions. In the first stage, lattice-based cryptography is presented as an authentication and signature technique that can withstand collusion attempts across healthcare authorities. To forecast future illnesses based on patient indications and attributes, the second step entails employing Deep Learning as a Service (DaaS) on archived EHR information. Accuracy, end-to-end latency, mining time, and computation and transmission expenses are only a few of the metrics used to gauge the success of the suggested system. The results show that BinDaaS performs better than competing solutions across all of these measures, making it the best option for managing and predicting EHRs.

Guo et al. [13] suggested a multi-authority attribute-based signing technique to guarantee the integrity of blockchain-stored EHRs. Patients can now recommend attribute-based messaging without disclosing any more personal information using this system. Since the blockchain is a decentralized ledger, using several authorities eliminates the requirement for a single point of failure. The protocol protects itself against collusion attacks by having its authorities share private pseudorandom function (PRF) seeds. When compared to existing approaches, the suggested attribute-based signature system proved to be both secure and efficient. This method, when combined with blockchain technology, gives patients more control over their medical records and allows for the safe, distributed administration of electronic health records.

Al Omar et al. [14] focused on the growing interest of cybercriminals regarding medical data and recommended a blockchain-based, patient-centric approach to managing healthcare records. MediBchain was a solution that used a decentralized network of peers to protect user privacy and data integrity. In order to maintain the privacy of their patients' information and attain pseudonymity, the researchers used Elliptic Curve Cryptography (ECC) for encryption and cryptographic functions. The review highlighted the benefits of the

proposed platform while also presenting a privacy-preserving approach for healthcare data. The goal of this study was to create a decentralized system that would improve the patient experience on the web while protecting their privacy.

Hasanova et al. [15] presented a machine learning-based algorithm called Sine Cosine Weighted K-Nearest Neighbour (SCA-WKNN) is proposed for the early prediction of heart disease. The algorithm utilizes data stored in the tamper-resistant blockchain, ensuring data authenticity and secure storage for patient information. The performance of SCA-WKNN is compared to other algorithms using metrics such as accuracy, precision, recall, F-score, and root mean square error. The results show that SCA-WKNN achieves higher accuracy compared to W K-NN and KNN, with an improvement of 4.59% and 15.61%, respectively. Furthermore, the study compares blockchain-based storage with peer-to-peer storage in terms of latency and throughput, finding that decentralized blockchain storage has a maximum throughput 25.03% higher than peer-to-peer storage. This research highlights the potential of IoT integration and blockchain technology in facilitating early detection and secure management of heart disease.

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.

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 aday, 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 a non-letter character, 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 intelligencetechnologies 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 ForestSTEP 1: START
STEP 2: SPLIT dataset into 67 percent training set, 33percent testing set
STEP 3: FOR train dataset
CALL RFClassifier TRAIN RFClassifierSTEP 4: FOR test dataset
CALL RFClassifier PREDICT the label COMPUTE AccuracyScore SAVE AccuracyScore DISPLAY ConfusionMatrix
STEP 5: STOP
```

```
B . KNN ClassifierPseudo code Procedure Train()
  // Input: train set, test set
// Output: Trained model
Step 1: Read Train set and test setStep 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, 33percent 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
```

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.

System Architecture

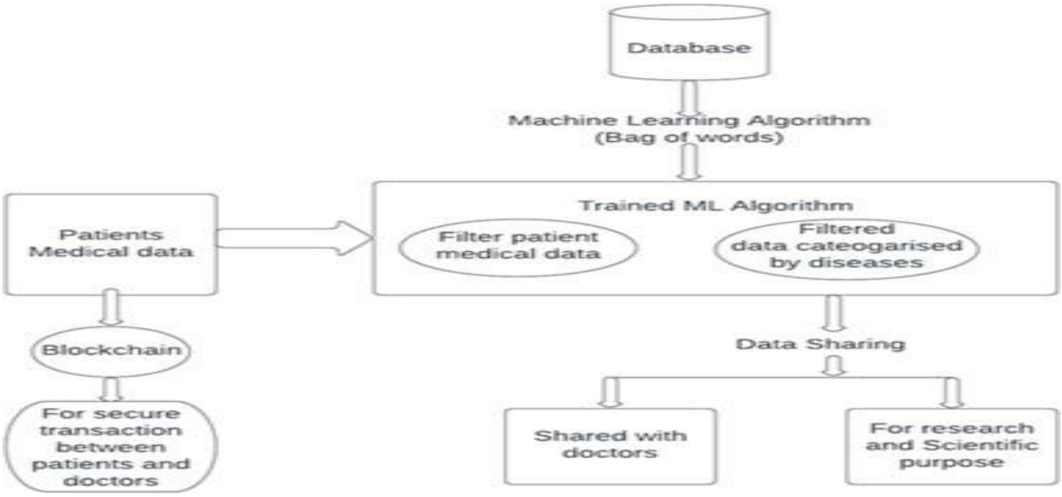

Database
Machine Learning Algorithm (Bag of words)
Trained ML Algorithm
Patients Medical data
Filter patient medical data
Filtered data cateogarised by diseases
Blockchain
Data Sharing
For secure transaction between patients and doctors
Shared with doctors
For research and Scientific purpose

Fig.1: System Architecture.

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.

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.

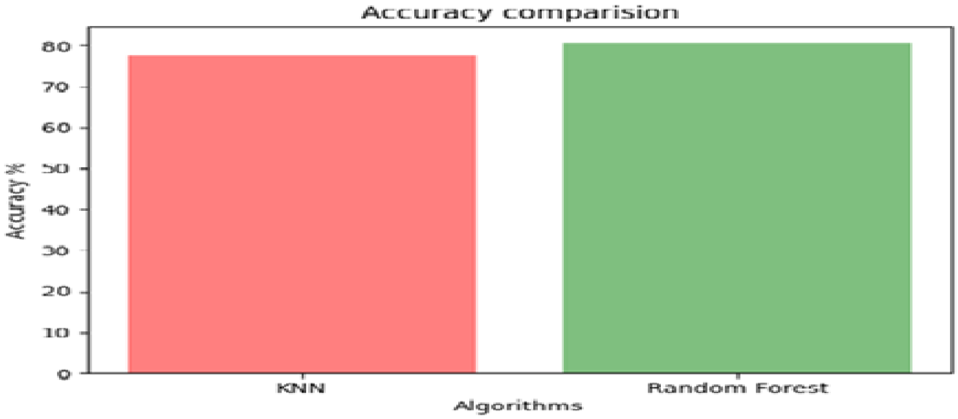


Fig.2: Accuracy comparison between algorithm From the values calculated in the confusion matrix, an accuracy graph (Fig2) is generated for each algorithm for comparison for best algorithm with highest accuracy.

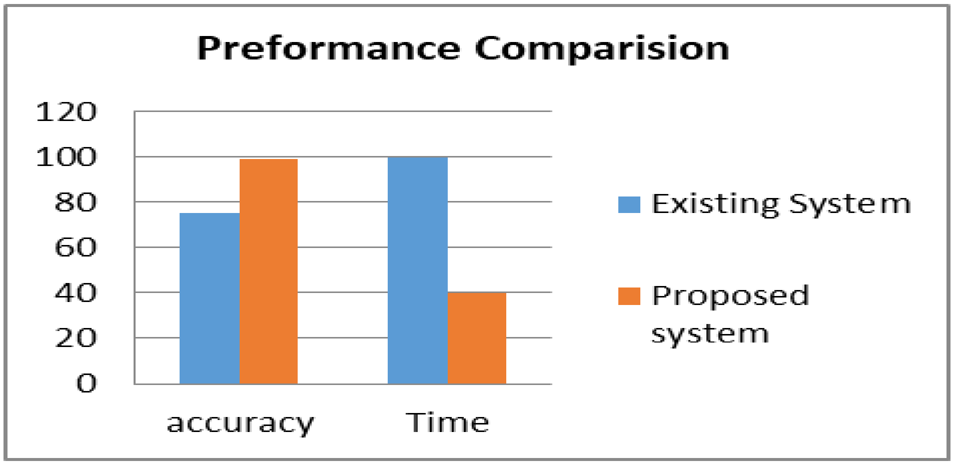


Fig.3: Performance Comparison.

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 trainingand prediction by a model.

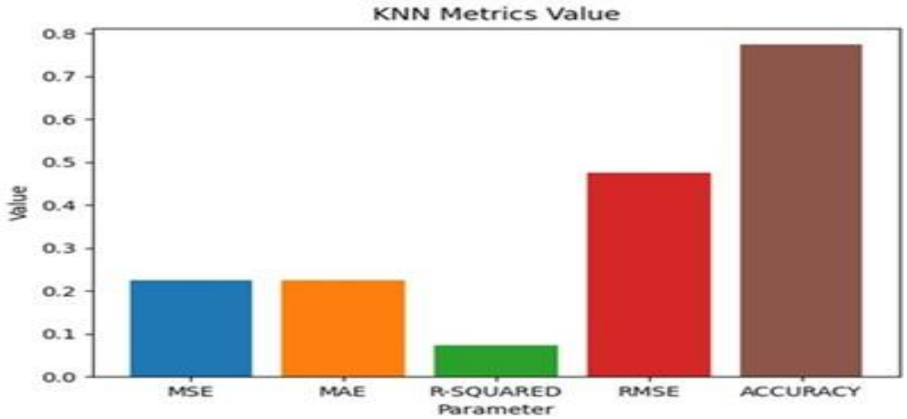
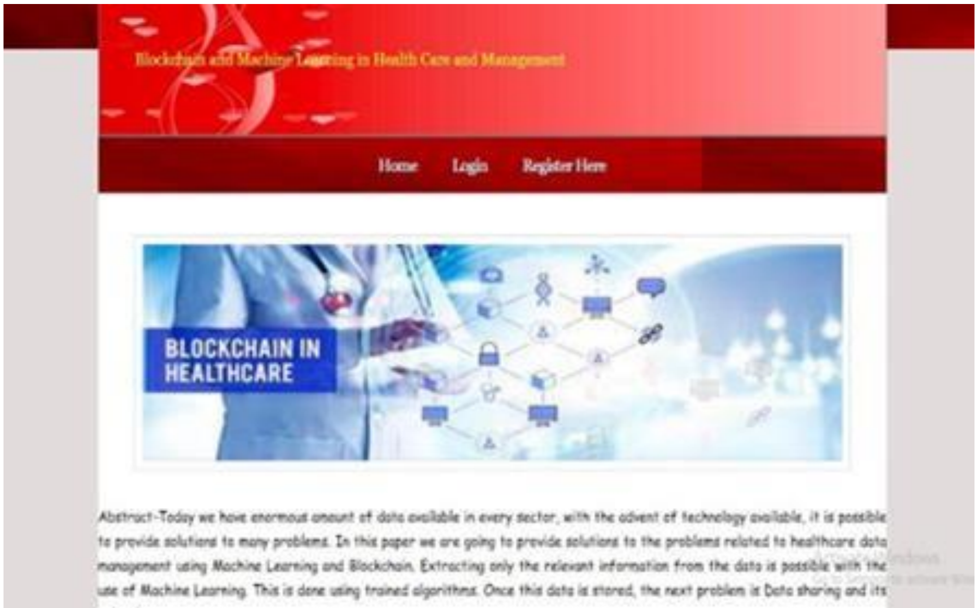


Fig.4: KNN Metrics Value.

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

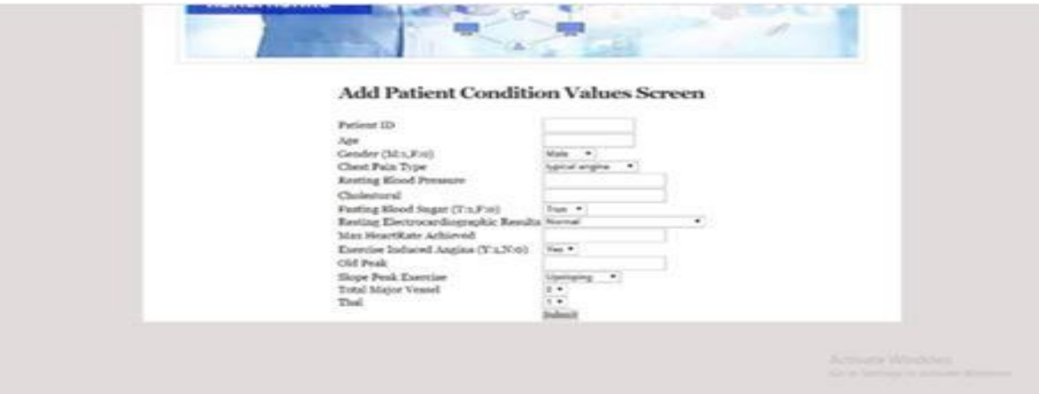

Blockchain and Machine Learning in Health Care and Management
Home
Login
Register Here
BLOCKCHAIN IN HEALTHCARE

Abstract-Today we have enormous amount of data available in every sector, with the advent of technology available, it is possible to provide solutions to many problems. In this paper we are going to provide solutions to the problems related to healthcare data management using Machine Learning and Blockchain. Extracting only the relevant information from the data is possible with the use of Machine Learning. This is done using trained algorithms. Once this data is stored, the next problem is Data sharing and its

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



Add Patient Condition Values Screen

Patient ID	
Age	
Gender (M:1,F:0)	Male
Chest Pain Type	typical angina
Resting Blood Pressure	
Cholesterol	
Fasting Blood Sugar (T:1,F:0)	True
Resting Electrocardiographic Results	Normal
Max HeartRate Achieved	
Exercise Induced Angina (Y:1,N:0)	Yes
Old Peak	
Slope Peak Exercise	Upsloping
Total Major Vessel	0
Thal	1

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.

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 Blockchainand Machine Learning is not much explored.

References

[1]. M. Ho¨lbl, M. Kompara, A. Kamiˇsalic´, and L. Nemec Zlatolas, “A systematic review of the use of blockchain in healthcare,” Symmetry, vol. 10, no. 10, p. 470, 2018.

[2]. “Ibm.com,” 2022, https://www.ibm.com/in-en/cloud/learn/ machine-learning.

[3]. G. Carleo, I. Cirac, K. Cranmer et al., “Machine learning and the physical sciences,” Reviews of Modern Physics, vol. 91, no. 4, 045002 pages, 2019.

[4]. A. A. Monrat, O. Schele´n, and K. Andersson, “A survey of blockchain from the perspectives of applications, challenges, and opportunities,” IEEE Access, vol. 7, pp. 117134–117151, 2019.

[5]. P. S. Kohli and S. Arora, “Application of machine learning in disease prediction,” in 2018 4th International Conference on Computing Communication and Automation (ICCCA), pp. 1–4,

IEEE, 2018.

[6]. F. Alam Khan, M. Asif, A. Ahmad, M. Alharbi, and H. Aljuaid, "Blockchain technology, improvement suggestions, security challenges on smart grid and its application in healthcare for sustainable development," Sustainable Cities and Society, vol. 55, p. 102018, 2020.

[7]. L. Ismail, H. Hameed, M. AlShamsi, M. AlHammadi, and N. AlDhanhani, "Towards a blockchain deployment at uae university: performance evaluation and blockchain taxon- omy," in Proceedings of the 2019 International Conference on Blockchain Technology, pp. 30–38, 2019.

[8]. Zhang G, Yang Z, Liu W. Blockchain-based privacy preserving e-health system for healthcare data in cloud. Computer Networks. 2022, 203: 108586. doi: 10.1016/j.comnet.2021.108586.

[9]. Ismail L, Materwala H, Hennebelle A. A Scoping Review of Integrated Blockchain-Cloud (BcC) Architecture for Healthcare: Applications, Challenges and Solutions. Sensors. 2021, 21(11): 3753. doi: 10.3390/s21113753.

[10]. Velmurugadass P, Dhanasekaran S, Shasi Anand S, et al. Enhancing Blockchain security in cloud computing with IoT environment using ECIES and cryptography hash algorithm. Materials Today: Proceedings. 2021, 37: 2653- 2659. doi: 10.1016/j.matpr.2020.08.519.

[11]. Benil T, Jasper J. Cloud based security on outsourcing using blockchain in E-health systems. Computer Networks. 2020, 178: 107344. doi: 10.1016/j.comnet.2020.107344.

[12]. Bhattacharya P, Tanwar S, Bodkhe U, et al. BinDaaS: Blockchain-Based Deep-Learning as-a-Service in Healthcare 4.0 Applications. IEEE Transactions on Network Science and Engineering. 2021, 8(2): 1242-1255. doi: 10.1109/tnse.2019.2961932.

[13]. Guo R, Shi H, Zhao Q, et al. Secure Attribute-Based Signature Scheme With Multiple Authorities for Blockchain in Electronic Health Records Systems. IEEE Access. 2018, 6: 11676-11686. doi: 10.1109/access.2018.2801266.

[14]. Omar AA, Bhuiyan MZA, Basu A, et al. Privacy-friendly platform for healthcare data in cloud based on blockchain environment. Future Generation Computer Systems. 2019, 95: 511-521. doi: 10.1016/j.future.2018.12.044.

[15]. Hasanova H, Tufail M, Baek UJ, et al. A novel blockchain-enabled heart disease prediction mechanism using machine learning. Computers and Electrical Engineering. 2022, 101: 108086. doi: 10.1016/j.compeleceng.2022.108086.

[16]. Shi S, He D, Li L, et al. Applications of blockchain in ensuring the security and privacy of electronic health record systems: A survey. Computers & Security. 2020, 97: 101966. doi: 10.1016/j.cose.2020.101966.