Preventing Forgeries by Securing Healthcare Data Using Blockchain Technology

V. Vetriselvi, Sridharan Pragatheeswaran, Varatharajan Thirunavukkarasu and Amaithi Rajan Arun

Abstract Decentralization has gained a lot of attention due to its application in diverse fields. It is pioneered largely by bitcoin, a blockchain technology, and a financial application of decentralization, which has impacted a lot on how financial transactions happen in a secure manner. The advantage of using this technology is that there is no central authority to rely on. Thus, a decentralized storage of medical records would allow forgeries on the records to be reduced. We propose a solution to avoid forgery in healthcare sector using blockchain. The blockchain network in the proposed system will time-stamp and store healthcare management data and its associated files in the network storage. The network is decentralized; thus, the data is inherently secure. Yet this approach may create a storage exploitation and may lead to breakdown of the system. However, a machine learning-based classification model is used to decide upon which records that get into the blockchain to reduce the required storage. Hence, a system to securely store healthcare data using blockchain technology can be implemented or created.

Keywords Decentralization · Blockchain · Healthcare · Medical records · Classification · Machine learning

V. Vetriselvi (✉) · S. Pragatheeswaran · V. Thirunavukkarasu · A. R. Arun
Department of Computer Science and Engineering, College of Engineering, Guindy, Chennai, India
e-mail: vetri@annauniv.edu

S. Pragatheeswaran
e-mail: sripragathees29@gmail.com

V. Thirunavukkarasu
e-mail: vthirunavukkarasu96@gmail.com

A. R. Arun
e-mail: arunamaithirajan@gmail.com

M. Tuba et al. (eds.), *Information and Communication Technology for Sustainable Development*, Advances in Intelligent Systems and Computing 933,
https://doi.org/10.1007/978-981-13-7166-0_15

1 Introduction

Blockchain technologies are really fascinating the researchers largely due to their success of the bitcoin cryptocurrency. The main reason for this fascination is that the applications which ran only through a centralized authority can now operate as a fully decentralized authority and achieve the same functionality with the same amount of certainty, with blockchain in that place [1]. Satoshi Nakamoto explained functionalities of blockchain and proof of work [2]. Figure 1 shows the structure of blockchain [3]. With this huge advancement, researchers are trying to implement blockchain in different scenarios rather than this cryptocurrency scenario. Some of the areas where researchers are trying to implement blockchain are finance [4, 5], healthcare to utilities, IoT field [6], music industry, real estate, and the government sectors.

Mizrahi [7] proposed how assets that can be uniquely identified by one or more identifiers can be registered in blockchain. This can be used to verify ownership of an asset and trace the transaction history. Any physical or digital property such as real estate, automobiles, laptops, and other valuables can potentially be registered in blockchain. The ownership and transaction history can be validated by anyone.

Mettler [8] projected how blockchain can prevent the counterfeiting of given drugs in the pharmaceutical industry. Estonia [9] is implementing a blockchain-based healthcare management record to store it in a hacker-proof manner. In that project, logs of access of healthcare data and audit data are stored in blockchain. All the users can see when the data is accessed, but the access log cannot be modified.

Nowadays, healthcare data is more important since it can be modified by higher authorities. Hence, the security of this healthcare data becomes questionable. If patients get affected by medications or treatments given that cannot be proved by them, it can be easily counterfeited. This happens because we completely rely on the centralized record management system maintained. So, we need to build a system which records the medications or treatments given in such a way that it is impossible to be counterfeited even by the higher authorities. Hence, the original treatments

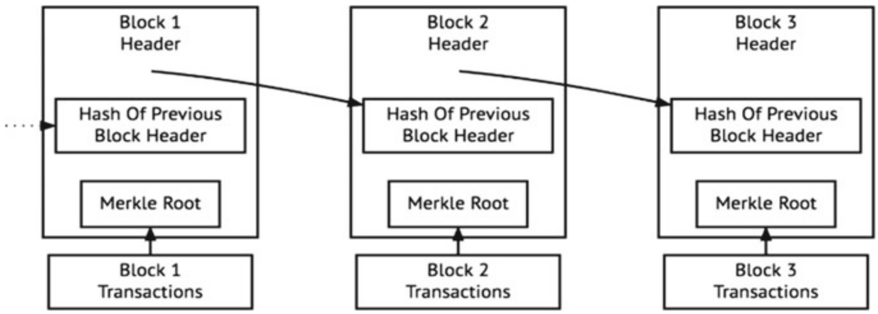

Fig. 1 Structure of blockchain

given cannot be hidden from the eye of society. To achieve this, we have to store all surgery details from patient's surgery approval date to surgery date in blockchain. Especially heart, brain, and cancer surgeries are major in India, which has to be monitored very seriously.

We are implementing blockchain for healthcare data. To reduce the storage in blockchain, we are storing only the patient's details who needs surgery by classifying their heartbeat signals.

Dokur and Olmez [10] developed a method for abnormal heartbeat sound decision-making process comprising three main stages. At the first stage, S1–S2 sounds are segmented [11], i.e., their timings are determined. S1–S2 sounds are used to extract the features of the heartbeat signals. At the second stage, feature vector elements are formed by using the wavelet plane. At the last stage, classification process is realized by a SVM learning algorithm [12].

Our system mainly focuses on securing the healthcare data using the blockchain technology. Since the healthcare data is very large in size, we have restricted ourselves to heart diseases. To fasten the process of diagnosis, CNN machine learning algorithm has been implemented to classify the heartbeat as normal or abnormal using the heartbeat signal recorded. The heartbeat signals have been first preprocessed. Then, noise has been removed by applying a low-pass filter. The features are extracted from the signal by FFT method. CNN model has been built using the features extracted, which is then later used for classifying heartbeat signals.

2 Proposed System

The machine learning phase classifies the heartbeat signals into normal or abnormal to find critical cases which have to be stored securely. We used blockchain to prevent those critical data from tampering. Additionally, we are analyzing the data stored in blockchain and identifying disease spread in the country with the help of GeoSmart contract.

2.1 Heartbeat Classification Phase

In this phase, we have implemented CNN to classify heartbeat signals into normal or abnormal. It includes training as well as testing phases. It consists of data preprocessing, feature extraction, and classification process.

Normalizer and Noise Remover. This module takes a normal WAV file as input. WAV file contains time domain signal values. There are some complex operations on these signals. We have applied random normalizing function which converts the amplitude values within -1 to 1 to make those operations easier. The heartbeat signals

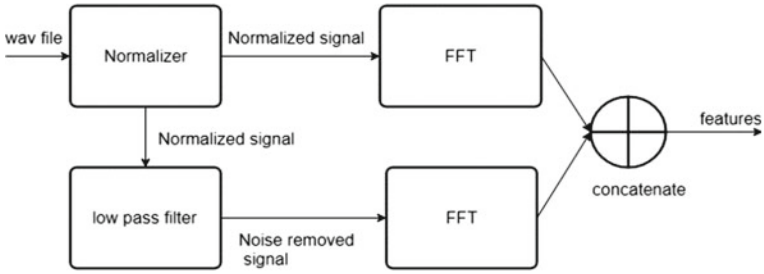

Fig. 2 Process of feature extraction

recorded might have extra noises which affect the accuracy of the classification. So, a Butterworth low-pass filter has been created to filter out frequencies greater than 2.5 Hz. However, normal human heartbeat frequencies maxes up to 2.5 Hz.

Feature Extraction. Features of each WAV file are needed to train the CNN classification algorithm. The coefficients of FFT have been taken as features. Formula of FFT is given in Eq. (1) [13].

$$X(k) = \sum_{n=0}^{N-1} x(n) \cdot \mathrm{e}^{-j(2\pi|N)nk} \tag{1}$$

where $X(k)$ is frequency domain samples, $x(n)$ is time domain samples, N is the size of FFT, and k varies from 0 to $N-1$. To improve the accuracy of the classification, FFT has been done two times before and after noise removal and concatenate them. Flow of this procedure is given in Fig. 2.

CNN Classification Process. Training part of classification model has been done using convolutional neural network to get higher accuracy than other classification algorithms. For this CNN training, we have feature vectors as input. There are two hidden layers. We have two nodes in output layer to classify heartbeat signal as normal or abnormal. Initially, weights and biases are randomly assigned in this CNN; then weights and biases will be updated in upcoming iterations. Finally, this model predicts the heartbeat signal as normal or abnormal. A full classification process has been shown in Fig. 3.

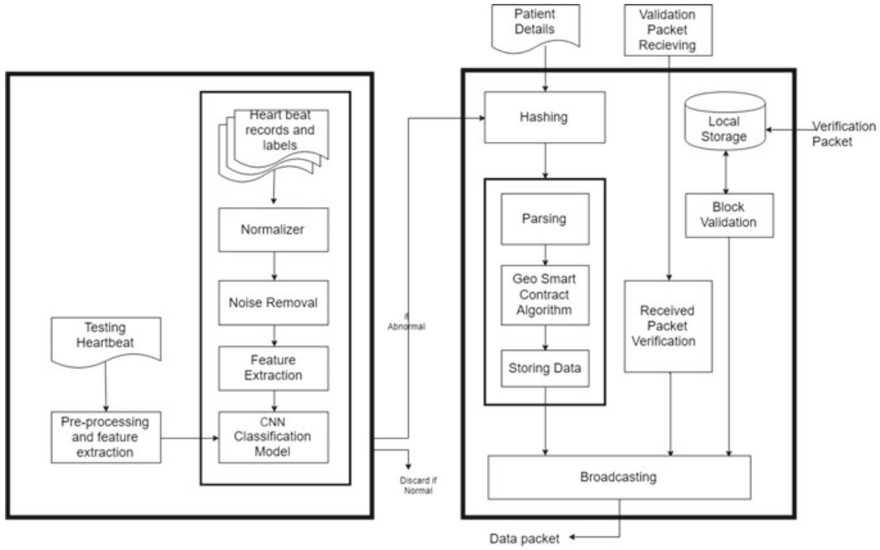

Fig. 3 Block diagram of a proposed system

2.2 *Transaction Generation Phase*

When the heartbeat signal is classified as abnormal, the patient has to undergo many screening processes. Once the surgery has been confirmed, the patient's details should be stored in blockchain. This phase comprises hashing the patient details, analysis of disease spread, and verification processes. Every transaction begins in this phase. Refer Fig. 3.

GeoSmart contract. Geo location data is extracted from patient details. This module will be triggered whenever a transaction is added to the blockchain. Here the flow of smart contract is given in Figure 4. Initially, smart contract parses the transaction, gets the location of the patient, and then alters the cumulative spread identity. GeoSmart contract gives us the survey of heart disease spread in all cities using geolocation data from the patient details.

3 Experimental Results

This section explains about the results obtained in CNN classification process and GeoSmart contract procedure which are described in previous sections.

3.1 Dataset Used

Heart signal recordings in the physionet heart signal database were sourced from several contributors, collected at either a clinical or nonclinical environment, from both healthy subjects and patients. The training set consists of five folders (A through E) containing a total of 3126 heart signal recordings, lasting from 5 to 120 s. The recordings were collected from different locations on the body. In both training and testing sets, heart signal recordings were divided into normal as well as abnormal heart signal recordings. The normal recordings were from healthy subjects, and the abnormal ones were from patients with a confirmed cardiac diagnosis. All recordings have been provided as .WAV format [14].

3.2 Accuracy in Classification of Heartbeat Signals

The cardiac sounds are selected from the physionet heart sound database. We used sixth-order Butterworth filter with cutoff frequency 2.5 Hz. All signals above 2.5 Hz are filtered out. Then we applied FFT algorithm on it and got features from that to train the CNN model. Accuracy of the prediction over the different size of dataset has been shown in Fig. 5.

When the size of dataset increases, the accuracy of the classification model also increases.

3.3 Comparison with Previous Approaches

We compare accuracy of our system with previous works with the help of following Eq. (2).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TF} + \text{TN} + \text{FN}} \tag{2}$$

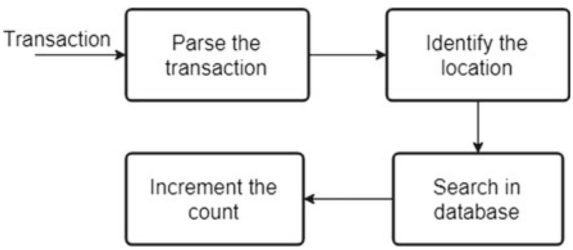

Fig. 4 Flow of GeoSmart contract

Table 1 Comparison with previous approaches

Approach	Accuracy
NN-2 layers [15]	0.79
DNN [16]	0.80
AdaBoost and CNN [17]	0.86
Our system	0.88

TP, FP, TN, and FN represent true positive, false positive, true negative, and false negative, respectively. Comparison with previous works with our work is shown in Table 1. Accuracy of our system is 88%.

3.4 Blockchain GeoSmart contract and Transaction Acknowledgement

Figure 6 states that the system is updated on execution of each transaction by GeoSmart contract. The bundled transaction is parsed, and the frequency of occurrence of surgeries is viewed in the front end.

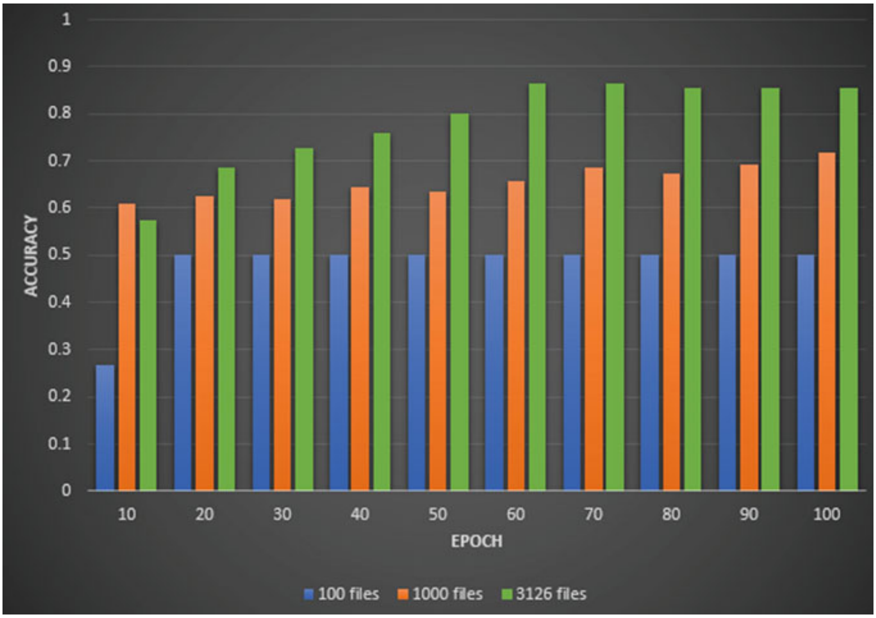

Fig. 5 Comparison between accuracy when varying size of dataset

State	Surgeries
State 1	23
State 2	45
State 3	90

Fig. 6 Result of GeoSmart contract

```
2018-04-24 08:05:44.594 UTC [shim] beforeInit -> DEBU 181 [78832eb8]Received INIT, initializing chaincode
2018-04-24 08:05:44.594 UTC [qscc] Init -> INFO 182 Init QSCC
2018-04-24 08:05:44.594 UTC [msp] GetLocalMSP -> DEBU 183 Returning existing local MSP
2018-04-24 08:05:44.594 UTC [shim] func1 -> DEBU 184 [78832eb8]Init get response status: 200
2018-04-24 08:05:44.594 UTC [shim] func1 -> DEBU 185 [78832eb8]Init succeeded. Sending COMPLETED
2018-04-24 08:05:44.594 UTC [shim] func1 -> DEBU 186 [78832eb8]Move state message COMPLETED
2018-04-24 08:05:44.594 UTC [shim] handleMessage -> DEBU 187 [78832eb8]Handling ChaincodeMessage of type: COMPLETED(state:ready)
2018-04-24 08:05:44.594 UTC [shim] func1 -> DEBU 188 [78832eb8]send state message COMPLETED
2018-04-24 08:05:44.594 UTC [chaincode] processStream -> DEBU 189 [78832eb8]Received message COMPLETED from shim
2018-04-24 08:05:44.594 UTC [chaincode] HandleMessage -> DEBU 18a [78832eb8]Fabric side Handling ChaincodeMessage of type: COMPLETED in state read
y

2018-04-24 08:05:44.594 UTC [chaincode] HandleMessage -> DEBU 18b [78832eb8-dfb9-4e79-bbe5-b9179b2120a8]HandleMessage- COMPLETED. Notify
2018-04-24 08:05:44.594 UTC [chaincode] notify -> DEBU 18c notifying Txid:78832eb8-dfb9-4e79-bbe5-b9179b2120a8
2018-04-24 08:05:44.594 UTC [chaincode] Execute -> DEBU 18d Exit
```

Fig. 7 Internal log of peer (hyperledger fabric)

Figure 7 is the internal log of one of the peer nodes. The peer is installed with chaincode. Logs state that the peer has received a transaction for chaincode invocation and the transaction identity (Txid) of executed transaction is notified. Thus, we can infer that the fabric is storing transaction in the blockchain storage.

4 Conclusion and Future Works

In our system, CNN has been used for the classification of heartbeat signals and hyperledger has been used for building blockchain. If the heartbeat is classified abnormal, then the corresponding patient has to undergo a surgery if needed. The surgery details have been considered as a transaction and broadcast in the blockchain network. Hence, the surgery details cannot be modified. In future, any machine learning algorithm which gives more efficient results than CNN can be used. X-ray reports, blood reports, etc., are not being stored in blockchain; these important reports also can be stored in blockchain. Due to scalability issues, we have restricted ourselves to heart surgeries; this can be extended to support all kinds of treatments and surgeries.

References

1. Owenson, G., Dennis, R., Aziz, B.: A temporal blockchain: a formal analysis. In: International Conference on Collaboration Technologies and Systems, pp. 430–437 (2016)

2. Nakamoto, S.: Bitcoin: a peer-to-peer electronic cash system. https://bitcoin.org/bitcoin.pdf (2008)
3. Bitcoin: Bitcoin developer guide. https://bitcoin.org/en/developer-guide#block-chain (2017)
4. Pattanayak, P., Verma, S., Crosby, M., Nachiappan, Kalyanaraman, V.: Blockchain technology: beyond bitcoin. http://scet.berkeley.edu/wpcontent/uploads/AIR-2016-Blockchain.pdf (2016)
5. Tschorsch, F., Scheuermann, B.: Bitcoin and beyond: a technical survey on decentralized digital currencies. IEEE Commun. Surv. Tutor. (2015)
6. Christidis, K., Devetsikiotis, M.: Blockchains and smart contracts for the internet of things. IEEE Access J. Rapid Open Access Publ. **04**, 2292–2303 (2016)
7. Mizrahi, A.: A blockchain based property ownership recording system. https://chromaway.com/papers/A-blockchainbased-property-registry.pdf
8. Mettler, M.: Blockchain technology in healthcare the revolution starts here. In: IEEE 18th International Conference on e-Health Networking, Applications and Services (2017)
9. Estonian citizens will soon have the world's most hack proof health-care records. https://qz.com/628889/thiseastern-european-country-is-moving-its-health-records-to-theblockchain/
10. Dokur, Z., Olmez, T.: Heart sound classification using wavelet transform and incremental self-organizing map. Digital Signal Proc. **18**, 951–959 (2008). Elsevier
11. Shervegar, M.V., Bhat, G.V.: Automatic segmentation of phonocardiogram using the occurrence of the cardiac events. Inf. Med. Unlock. J. **9**, 6–10 (2017). Elsevier
12. Azman, A., Jantan, A., Safara, F., Doraisamy, S., Ranga, A., Ramaiah, A.: Multi-level basis selection of wavelet packet decomposition tree for heart sound classification. Comput. Biol. Med. **43**, 1407–1414 (2013). Elsevier
13. Wada, T.: 64 point fast fourier transform circuit. http://www.ie.u-ryukyu.ac.jp/wada/design07/spece.html (2006)
14. http://physionet.org/physiobank/database/challenge/2016/. Accessed date: 29 Dec 2017
15. Grzegorczyk, I., Soliński, M., Łepek, M., Perka, A., Rosiński, J., Rymko, J., Stępień, K., Gierałtowski, J.: PCG classification using a neural network approach. In: 2016 Computing in Cardiology Conference (CinC), pp. 621–624. IEEE (2016)
16. Nassralla, M., El Zein, Z., Hajj, H.: Classification of normal and abnormal heart sounds. In: 2017 Fourth International Conference on Advances in Biomedical Engineering (ICABME), pp. 1–4. IEEE (2017)
17. Potes, C., Parvaneh, S., Rahman, A., Conroy, B.: Ensemble of feature based and deep learning-based classifiers for detection of abnormal heart sounds. In: 2016 Computing in Cardiology Conference (CinC), pp. 621–624. IEEE (2016)