A Correlation Blockchain Matrix Factorization to Enhance the Disease Prediction Accuracy and Security in IoT Medical Data

P. Renuka and B. Booba

Abstract An IoT software product's reliability is the probability of the product working "correctly" under or over a given time. New opportunities are the result of expansion in the fast-paced Internet of Things (IoT) space. IoT technologies on the collected datasets improve disease progression technology, disease prediction, patient self-management and clinical intervention. To propose, the IoT with cipher block chaining in the traditional cryptographic operation mode will be used for cryptographic processing. Developing models for the supervised learning classification and security of imbalanced datasets is challenging, especially in the medical field. However, most real-time IoT datasets present most traditional machine learning algorithms challenging unbalanced datasets. Proposed a new framework for the Correlation Blockchain Matrix Factorization Classifier (CBMFC) related to comprehensive medical records. CBMFC uses a multiple class label machine learning that represents an independent population model based on disease meta functions such as profile age, group, or cognitive function keys. The Pairwise Coupling Multi-Class Classifier (PCMC) is used to prove the model's correctness. This produces more comprehensive data in various machine learning environments, such as predictive classification, similar to real data performance. For the results of security analysis confirmation, the proposed IoT application model's effectiveness can withstand various attacks, such as selected cryptographic attacks. In this proposed CBMFC system, classification accuracy, precision, recall, execution time and security matrix are used to evaluate performance.

Keywords Internet of Things · Correlation Blockchain Matrix · Pairwise Coupling Multi-Class Classifier · Cryptography

P. Renuka (✉) · B. Booba
VISTAS, Vels University, Chennai, India

K. Khanna et al. (eds.), *Cyber Security and Digital Forensics*,
Lecture Notes on Data Engineering and Communications Technologies 73,
https://doi.org/10.1007/978-981-16-3961-6_30

1 Introduction

The use of the Internet of Things (IoT) in healthcare is sharply increased among a variety of specific Internet-use cases. The Medical Internet of Things takes more effort to improve care itself, with remote monitoring as the broader primary application of remote medicine. The synergies between medicine and technology have taken great strides around the world. For example, the Internet of Things (IoT) data analytics is gaining popularity, providing the next generation of electronic healthcare and mobile healthcare services. They are transforming traditional institutions based on the management of large data with a blockchain solution. The chain's evolution is in favor and permitted by technical means to maintain the support of strategic applications needed for its potential growth. In the cloud-enabled network blockchain, there are trading and mining nodes both in the cloud and on-premises. According to embodiments, the node may be an enterprise-level server. The final level of overall collaboration on evolution is based on cloud and blockchain applications as the basis for distributed chains' operation. The devices configured to use public blockchain services and private blockchain nodes clouds to communicate securely via APIs. The IoT devices combining blockchain technology as a security framework IoT system using secure distributed key management techniques make it possible to discover each other and transaction encryption machine-to-machine.

In this, Fig. 1 represents the cloud blockchain network communication. Machine learning techniques have been widely used in the medical field. Most of the medical data resources can use to transfer valuable knowledge to assistive scientific decision making. It is stored in each hospital or other medical institution separately, which poses a major issue to the medical data applied to the constructed prediction model, its quality and efficiency. For medical professionals, researchers, there are several machine learning techniques to be used throughout disease research. Medical data keeps a history of patient records and will be unused in the future. This can be analyzed and considered for future research. This huge database is analyzed to identify machine learning techniques used by healthcare staff to predict patient risk.

The most commonly used machine learning technology, which may be registered in the category. Occupies a set of pre-classified patterns to create a demographic model classification. Learning and classification are involved in a process called data classification. The training data is being learned and analyzed by a classification algorithm. The test data is used for classification to approximate the classification rules. The rules can apply to the new data tuple with acceptable accuracy. The pre-classification example has been used in the classifier training algorithm to verify the group that requires the appropriate identification parameters. This section discusses the process of classification of disease analysis using IoT clinical data and machine learning methods.

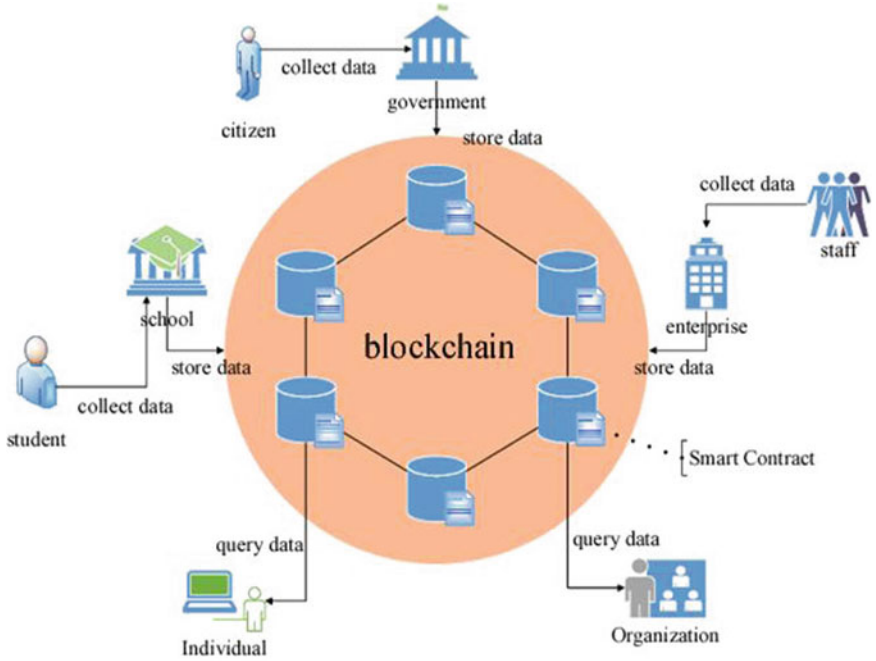

Fig. 1 Cloud blockchain network model

2 Related Work

The literature survey for various IoT blockchain security methods and disease classification strategies used to predict the result is discussed below.

Co-clustering is used simultaneously to group clinical features and patients to characterize block-wise data loss patterns [1] (A) group-based feature selection and data imputation for specific patient subgroups. (B) Miss the predictive model to consider data availability. The machine learning (ML) method of decision making and the medical field's data has proven highly predictive and supportive. ML technology's latest developments in the Internet of Things (IoT) [2] are being used. The task of breath estimation task [3] compares the performance of different machine learning techniques. Problems can be divided into two categories: high and low breathing work based on information extracted from pressure, volume and flow through signals recorded by mechanical ventilation.

Select the feature to solve the problem, and the proposed method is novel, rapid provision of mutual information to select the feature [4]. The feature selection algorithm improves the classification accuracy and is used to select the function to reduce the classification system's execution time. The neural network modification uses structured data and unstructured data and recommends a conventional neural network (CNN) basic peak disease risk prediction algorithm [5]. To the best of our

knowledge, existing work focuses on these two types of medical big data analytics. Machine learning technology has been used for [6] vegetation parameter estimation and disease detection; the effects of the disease symptoms on their performance have been small. Prevention to predict the actual occurrence of the pre epilepsy can help through therapeutic intervention [7]. Studies had found that abnormal activity in the brain begins minutes before the onset of a seizure, called a predictable condition.

Proposed DualFog-IoT is compared to the IoT-based architecture of the existing centralized data center. Having a genetic characteristic of blockchain, the proposed model system decreases rates and furthers existing IoT ecosystems unload minimum upper grayscale and [8, 9] to reduce the cloud data center. Blockchain technology's complexity is maintained for most developers or teams to build, usually expensive and difficult to monitor the support blockchain network in their application. This algorithm thereby forces the redesign of blockchain that uses cryptosystems to resist quantum attacks, creating quantum-fixed or quantum-resistant cryptosystems, called quantum proofs, called post-quantum [10]. Threatens public key cryptographic hash functions. If all copies of the opponent store segments [11], the opponent can leave the system, causing a permanent loss of segments due to the blockchain system's malfunction.

The unique geometry of [12] high-dimensional data using a manifold learning and support vector machine (SVM) proposed PVC detection and data visualization method. [13, 14] has proposed some of the blockchain-based storage systems in recent years. In most cases, the blockchain acts as a "witness of the agreement" with publishers. This method is, so far, because it is not managed to reduce the size of the blockchain itself, and it will not be able to avoid the storage model of Bitcoin. Maintaining this also a complete record of blockchain implementation in the IoT environment helps to insufficient storage capacity on edge devices. At the same time, the system does not require any significant transactions per second [15]. Even segment blockchain is used to improve blockchain sharing by separating transaction storage from transaction validation.

Our main contribution is to link the probability of failure as a group to each epoch by using the probability limit as the sum of the upper limit hypergeometric [16] and the two types of distribution. It is a reliable data management scheme based on blockchain in edge computing (BlockTDM called) [17] that has been proposed to solve the above problems. The flexible configuration of blockchain architecture, mutual authentication protocol, flexible consensus, smart contract management module and transaction data and blockchain node management and deployment are among them. Algorithms included artificial neural networks (ANN) [18], support vector machine (SVM), naïve Bayesian classifier (NBC), boosted decision trees (BDT) and multivariable logistic regression (MLR).

However, one of the important problems of fog computing and blockchain [19] integration is scalability. The group chain has proposed a new type of scalable public blockchain for the double-stranded structure of IoT services computing fog of computing. Despite its potential, there are some urgent problems to be solved to make the IoT services are widely used. Various loosely coupled distributed

intelligence [20] to adjust the connection operation requires an IoT device for managing the system. Developers' point of view analyzed blockchain from [21] blockchain, a larger software system, highlighting data storage and combination key concepts and considerations.

Two methods are currently introduced based on the heterogeneous network, protein interactions, genotype—is built using the phenotype correlations and phenotypic similarity. In HeteSim_MultiPath (HSMP) [22], HeteSim scores different routes contributing to the longer path and the damping constant. Therefore, a non-invasive diagnostic system based on machine learning (ML) has been developed to solve these problems. The decision-making system of experts based on the application of machine learning classifiers and artificial fuzzy logic is an effective diagnostic result; the mortality rate has decreased [23]. Therefore, no clear requirements should be treated as a technique to identify the most appropriate parameters during predictive analysis.

Instead, the feature extraction time-series EEG is converted to a time-frequency distribution (TFD) [24]. Gradient boosters use a set of programs aside to train the entire TFD directly. This paper proposes a machine learning method for predicting particulate matter concentrations from wind and precipitation levels, based on [25] two-year weather and air pollution data.

This paper [26] has been used to record valid composite minorities oversampling to generate new composites. Baldwinian learning and PSO (BLPSO) [27] are based on a novel hybrid algorithm to increase particle diversity and prevent premature convergence of PSO. Ability to compare calculated heart rate variability (HRV) baselines before the start of daytime antibiotics (LOS group) or during a randomly selected period (control group) during the calibration period [28].

In the future, Moderate Resolution Imaging Spectrum radix statistical model for predicting fire activity for 1–5 days to use satellite fire counts and meteorological data from temporary reanalysis [29] to develop. The component contains an apnea event [30], which automatically proposes the first use of EIT boundary voltage data from infants to obtain the main function of research apnea detection using machine learning. Factorial switching in linear dynamic systems is a common framework for addressing this issue [31].

This analysis of the previous method has a low classification accuracy and less security performance to introduce a new method in the next section.

3 Implementation of the Proposed Method

This IoT application model's main motivation is to provide a computationally secure key generation for protected data through encryption with blockchain technology. A master key is a secret key that is accepted between communicating parties before a communication protocol begins. The essential characteristic of machine learning is the design of heterogeneous data applications with different dimensionalities. Various IoT application consists of different information collected

by recording the data for producing diverse data representations. The collection of data is carried out by distributed and decentralized control with autonomous data sources.

Correlation Blockchain Matrix Factorization Classifier (CBMFC) algorithm is used to classify the given data into several columns and provide security in medical data information. Blockchain HMAC encryption is the enforcement of access control mechanisms, digital signatures, routing controls, notarization, etc., to provide data security services against attacks and prevention. The digest is calculated for ciphertext messages using the HMAC encryption algorithm. HMAC stands for the Hash-based Message Authentication Code. This authentication uses a key to implement the hash function product along with the content of the message. From Fig. 2, first, preprocess the data using Co-relational Matrix Data Preprocessing to remove the noise. The model is mainly executed in the preprocessing task. It presents a preprocessing task to remove noise and inconsistent data obtained from various sources. Interval-based measures are the discretization factors from large data applications, and the measured values reach the closest values. Data preprocessing is usually done to see if there is redundant information in the dataset. All attributes are initially considered separate subsets, with the final combination of attributes and features is marked as the highest subset of features. The hash value is stored on each block; if any new record enters the storage, it will update into the previous block. This proposed framework reduces the clustering task; a grouping of the same attributes occurs by using our framework for predicting the disease using a given disease dataset.

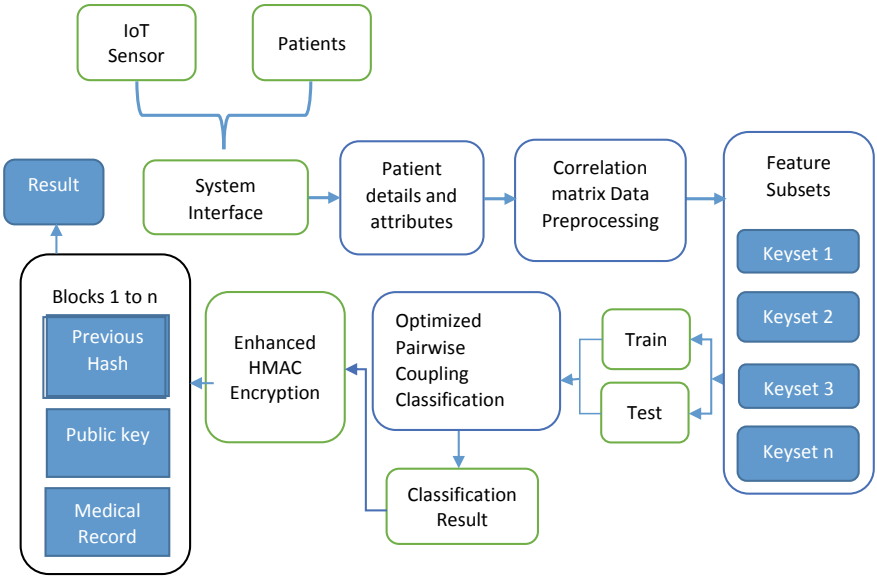

Fig. 2 Proposed Method CBMFC block diagram

3.1 Correlation Matrix Data Preprocessing

The initial stage in Correlation Matrix Data Preprocessing is finding the unrelated data and normalized datasets. While processing the data, it is essential to calculate the univariate stats like the mean value, standard error, and rate of recurrences to inspect the volume of missing data. In complete information, maximum likelihood the population criteria have approximated that would probably generate the estimate values from the sample set which is analyzed. The classification accuracy depends upon the accuracy of data, so data should be non-ambiguous, correct and complete. Data collection methods are loosely controlled, resulting in inappropriate values like out-of-range missing values. Correlation matrices are useful for showing the correlation coefficient (or degree of relevance) between variables. The correlation matrix is symmetric, just as the correlation between $V1$ and $V2$ is the same as the correlation between $V2$ and $V1$.

Let the data matrix V composed to n-dimensional observation dataset, and individual variable values (V) size of $[R \times n]$. To assume the row of R has centered, observe all the row values i to n. The correlational $\sum$ of each row V data.

$$\sum = cor(V) = \mathrm{E}\left[V^T\right] \tag{1}$$

To estimate the correlation data matrix.

$$\mathrm{E}\left[V^T\right] \approx \frac{V^T}{n} \tag{2}$$

Form the Eq. (2), matrices as the product of two simpler matrices E and L, using a procedure known as Eigenvalue Decomposition.

$$\sum = EL^{-1} \tag{3}$$

The data matrix E is resized $[R \times C]$ matrix, where each column to apply the eigenvector.

From Eqs. (2, 3), derivation of the correlation matrix is as follows,

$$\rho(R, C) = \mathrm{corr}(R, C) = \frac{\mathrm{cov}((R, C)V)}{\sigma_R \sigma_C} = \frac{E[\mathrm{cov}(((R - \mu_R)(C - \mu_C))V)]^{-1}}{\sigma_R \sigma_C} \tag{4}$$

All such data discrepancies can lead to wrong research results; thus, data is processed before applying an algorithmic technique for better and improved results. Data needs to be preprocessed to ease the entire data process to improve machine techniques' efficiency. The main steps used for preprocessing the data include data cleaning, data integration, data conversion and data reduction. In this, fill attribute

values with the correlation value for unknown instances and convert all disease databases in a single file format.

3.2 Medical Feature Selection

A feature is that a subset selection which is a preprocessing step used in machine learning. It aims to increase learning accuracy to reduce and eliminate invaluable and irrelevant data dimensions. It indicates specific problems and their functions and the type of prediction that will be useful.

The input dataset is fed into the feature selection method block, where the feature selection is made according to the given dataset. It will take this way to reduce the number of attributes selected for a given number of dependent attributes.

The medical algorithm feature selected by randomly sampling instances from the training data and the selection process is shown in Fig. 3. It has been found that the nearest value class (adjacent value) of the same class is opposite each time to select the best models. The characteristic weight is based on its value to distinguish the example for detecting the model or has been updated from the latest hit and latest features.

$$\text{Weight } (W_m) = \sum_{i,j=0}^{w} f_m - \frac{(U,\ V,\ S\ )^{\,2}(i)}{N} \tag{5}$$

$$\text{Gain } (G_m) = \sum_{i=o}^{n} S_i log_2 \tag{6}$$

where f_m, the weight for attributes U, V, which are randomly sampled instances, S is the latest hit and N is the number of randomly sampled instances. The function diff calculates the difference between two instances of a given attribute.

f_m is an attribute of U, V, if they are an approximate model example, S is the recent success and N is the approximate number of sample events. The functional differences are calculated as the difference between the two events of a given

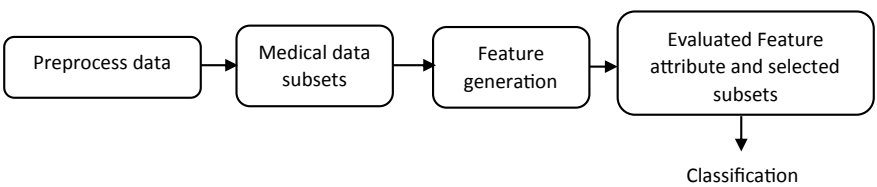

Fig. 3 Best feature selection process

attribute. The difference between the continuous attribute is the actual difference normalized to the interval [0, 1].

Algorithm steps

Input: training data D and feature subsets f_{n-1}

Input: training data D and feature subsets f_{n-1}

Step 1: Initialize the Attribute $S = S_1$

Step 2: Calculate the feature weight and gain using above eqs. (5) and (6)

Step 3: $\alpha_f = \max(S_i, G, W)$ to evaluate the best feature (α_f)

Step 4: For each current iteration $\leq N$

If $\alpha_f > th$, then // where th is represented a threshold value.

$$f'(D) = S \cup \left(\frac{G(n)+W(n)}{\alpha_f}\right) \quad (7)$$

Feature ⚽ f';

End

If available in the presence and corresponding class distribution of the feature, feature selection measures the amount of information about the class prediction bit.

3.3 Optimized Pairwise Coupling Classification (OPCC)

It also serves as a classification mark supervised learning method and a statistical method of classification. It considers a basic natural model and assigns it to us by determining the probability of imprisonment with uncertainty in the moral model. This feature is characterized by information gain, and then, the best ranking features are selected as the best attributes to use in the classification.

Algorithm steps:

Step 1: Read data Ds= $\{\{P_i, Q_i\}/i=1, 2, 3\ldots n\}\}$ be the set of training data.

Step 2: Initialize the random weight support values, $W(0)$.

Step 4: For each training data $(P_i, Q_i) \in$, Ds.

To analysis the predicted output $Y_i\hat{}\ (k)$

For each weight, we do.

Update the weight $(k+1) = W_j\ (k) + (y_i - y_i\hat{}(k))\ x_{ij}$.

End for.

End for.

Step 5: Until disease predicts the output.

The attributes are determined for each attribute's information to gain to classify a set of data segments. Then the information gain must be selected maximized attributes. After the classification result was stored in the database using HMAC encryption with blockchain.

3.4 HMAC Encryption and Blockchain Security Model

The Hash-based Message Authentication Code (HMAC) algorithm is implemented using binary operations and hash functions. HMAC is calculated with any cryptographic hash function; the resulting MAC algorithm is called HMAC-MD5. The security strength provided by the HMAC algorithm depends on the HMAC key, the most basic hash algorithm and the security features of the MAC Tag length. MAC is calculated using the data on the HMAC function, and the following operation is performed:

$$\text{HMAC}\ (k_1, k_2,\ \text{data}) =\ \text{hash}\ ((k_1,\ \text{inner})\ ||\ (\ k_2,\ \text{outer}))\ \text{t} \tag{8}$$

where t = time, k_1 *and* k_2 = a pair of keys (encryption and authentication).

Algorithm steps

Input: document, MK- master key, HMAC encryption key (k1, k2).

Step 1: To initialize the MK, $k1$, $k2$.

Step 2: compute the HMAC tag for authentication to create a block.

Step 3: To encrypt the document using the XOR function

Step 4: read the data hash function to generate the MK, $k1$, $k2$

XOR⊕generate 64{MK, $k1$, $k2$}

Step 5: If the tag T equal, then

to connect the blockchain network and authenticate the user.

Else

return

Step 6: Store the document SD

Step 7: mapping the authentication key to the array and remove the special characters, spaces from SD.

Step 8: for n to string

$K_{SD} = H_{k1}, (A_i)$// where, $H_{k1(x)}$is hash key 1 and A_i authentication key

Convert to K_{sD} to integer list and remaining truncate bits

End

Step 9: Decryption process and verification key

for n to the number of character

If (SD is an integer), then.

$P(C_i - K_{SD(i)})$ mod 10 // where p is integer array and C_i,

character array.

Else

$P(C_i - K_{SD(i)}) \bmod 26$

Return

End

End

Step 10: Mapping and validate the p array data to x.

It is considered an easy way to compute the MAC value $h(x)$ for a given k and an arbitrary input x (Fig. 4).

$H_{k1(x)}$ is mapped to a message x of arbitrary length value having n number of fixed bits.

It is considered impossible to calculate the MAC value $hk(m)$ of a new message m if the key k is unknown, though we get MAC values of other messages.

Protecting HMAC is used to demonstrate the correct data relationship between the designer's embedded strength and the HMAC hash function. In terms of an embedded hash function, the HMAC function's strength for basic hash function encryption within security depends in a certain way. HMAC function is usually a safe and predetermined number of successes based on the probability of fraud. Created based on the time spent using the same key to create a MAC message.

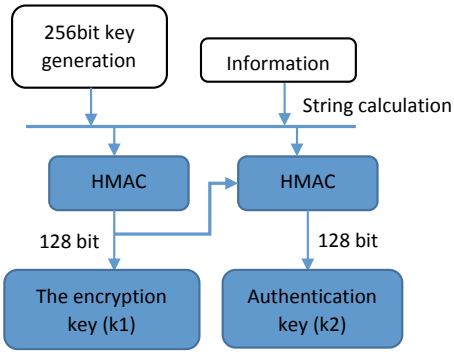

Fig. 4 Process of Key Generation

3.5 Blockchain Network Construction Algorithm:

Step 1: To initialize the index of the variable (IN), Timestamp (Ts), Node Information List (NI), Previous Hash (PH), Data Identity (DI).

Step 2: Create a new block.

IN = Chain.Count // in this chain count is the total number of the node to enter the network

TS = DateTime.UTCNow

NI = {user request, request ID, ssLocation}.to list

PH = Get hash (Chain.Last)

DI = current data-id

Chain.add(IN, TS, NI, PH, DI)

Step 3: To create a user transaction data-id (DI)

DI =0

While (Is not valid (last DI, Current DI, PH))

DI++

Return DI

Step 4: To register the users in blocks.

For each user in users

URL_{node} =$"http// {user id}"

Register user (URL_{user})

Insert (user.count()) //new user is added.

End

Step 5: Check the user information and database block

If (block is empty == 0)

Go to step 2

Else

Verify the block attacker or not

Record = data size

IN = block.IN, TS = block.TS, NI = block.NI, PH = block.PH

For the user in u //. u=users

PH (k) = HMAC [k] // k is keys

To evaluate the user (Li) using.

$$\mathrm{U}(i) = IN_{n(i)}.\mathrm{getvalue} + TS_{n(i)}.\mathrm{getvalue} + NI_{n(i).}\mathrm{getvalue} + PH_{n(i)}.\mathrm{getvalue}$$

End for

End if

From the above algorithm step to form the network group, the user can upload their document to a centralized server with encrypted format help of HMAC encryption. The HMAC authenticates the user using the master key for help to encrypt and decrypt the documents. This blockchain network incorporates patient medical data to diagnose and predict disease, and the resultant data is stored securely.

4 Result and Discussion

Statistics provide a strong basic background for quantifying and assessing results. However, it needs to be modified and tweaked for statistics-based algorithms before being applied to the IoT blockchain method. This section presents the results of a work that proposed a technique for predicting disease using machine learning.

Table 1 Simulation parameters

Parameters	Values
Language	C#
Dataset name	UCI Machine learning repository
Diseases	Cardiology
Total number of data	1500
Train data	1000
Test data	500

Section analysis to take cardiology disease data and some parameters (Temperature, heart rate and blood pressure) are considered to predict the disease level. Table 1 represents the simulation parameter of the proposed method to use.

The number of correctly classified files with patient data according to the total number of files is defined as classification accuracy. This proposed method evaluates the following Eqs. (9, 10, 11, 12) for classification accuracy, precision, recall, F1 score, security and authentication time analysis. The Correlation Blockchain Matrix Factorization Classifier (CBMFC) method prediction accuracy is compared to the random forest, Bayesian classifier and SVM methods. Similarly, the proposed method's security analysis, Enhanced HMAC Blockchain (EHMACB) security, compares to existing method Blowfish, MD5 and MidChain methods (Fig. 5).

Different bytes of data are taken to estimate the method proposed in this execution time analysis. In this proposed method, EHMAC is compared to existing methods Blowfish, MD5 and MidChain. The proposed method is to authenticate users and store the data to the network within 270 ms less execution time than the HMAC blockchain method.

Figure 6 represents the comparison of the proposed and existing method graph. In this analysis of security result, the proposed method EHMACB provides a 93.5% security compare to existing methods MidChain has 91.3%, MD5 has 85.6% and blowfish has 83.4% security in the medical blockchain network. The machine

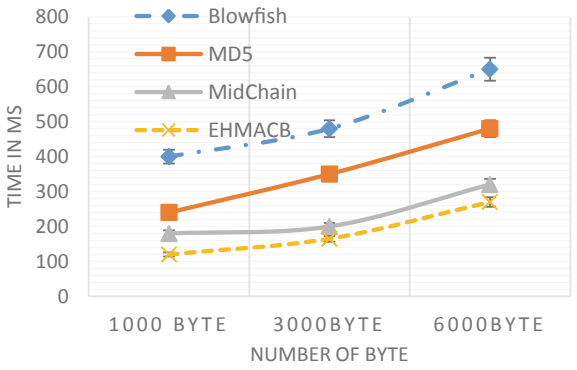

Fig. 5 Execution Time Analysis

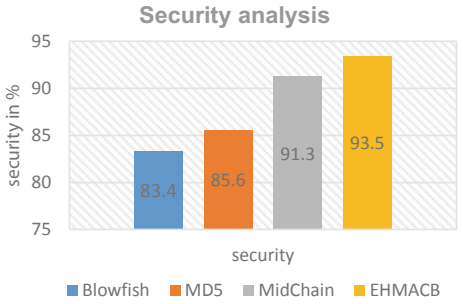

Fig. 6 Comparison of Security Analysis

learning performance of the clinical dataset classification after the analysis is evaluated using the equation below.

$$CA = \frac{\text{Number of classified files}}{\text{number of files}} * 100 \tag{9}$$

$$\text{Precision} = \frac{\text{true positive}}{\text{true positive} + \text{false positive}} * 100 \tag{10}$$

$$\text{Recall} = \frac{\text{true positive}}{\text{true positive} + \text{false negative}} * 100 \tag{11}$$

$$F1 = \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{12}$$

Table 2 shows a comparison of the existing and proposed method disease prediction performance. Table 2 shows the classification accuracy, precision, recall, proposed method CBMFC, existing methods SVM, random forest and Bayesian classifier (Fig. 7).

Using a pairwise coupling method to evaluate a two-pair feature matrix improves classification and prediction accuracy. The proposed machine learning method results provide higher performance than SVM, random forest and Bayesian classifier.

Table 2 Analysis of proposed method prediction performance

Methods	Classification accuracy in %	Precision in %	Recall in %
Bayesian classifier	79	80	78
Random forest	81	82	85
SVM	82	84	81
CBMFC	90	86	87

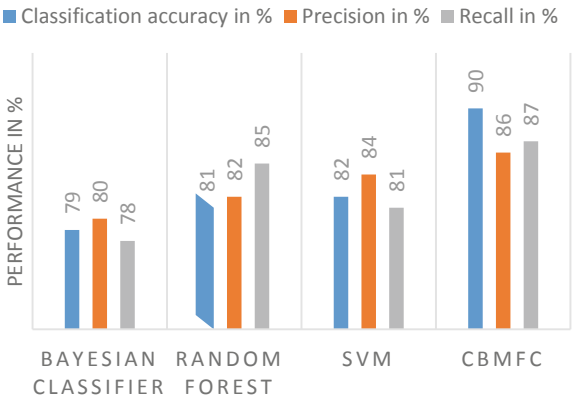

Fig. 7 Proposed method performance of prediction analysis

5 Conclusion

A given patient's cardiology disease needs to be diagnosed accurately and in time. The proposed Correlation Blockchain Matrix Factorization Classifier (CBMFC) analysis the IoT data disease prediction. First, to apply the co-relational matrix for preprocessing to remove the noise from the IoT dataset. Finally, a Correlation Blockchain Matrix Factorization Classifier (CBMFC) method uses the train data and predicts it. The Blockchain Enhanced Hash-based Message Authentication Code (EHMAC) Encryption is used to provide security, it encrypts the user request, and records are stored on blocks. This analysis of the proposed method simulation result has a 90% of classification accuracy, 86% of precision, 87% of recall values and 93.5% of security with 270 ms execution time more efficiently compared to the existing method. This proposed method to implement into hospitals for analysis disease and secure the data form unknown person.

References

1. Wang, H., Huang, Z., Zhang, D., Arief, J., Lyu, T., Tian, J.: Integrating co-clustering and interpretable machine learning for the prediction of intravenous immunoglobulin resistance in kawasaki disease. IEEE Access **8**, 97064–97071 (2020)
2. Mohan, S., Thirumalai, C., Srivastava, G.: Effective heart disease prediction using hybrid machine learning techniques. IEEE Access, 1–1 (2019)
3. Castro, L.F.B., Santacruz, L.F.E., Sánchez, M.B.S.: Work of breathing estimation during spontaneous breathing test using machine learning techniques. In: 2020 IEEE Colombian Conference on Applications of Computational Intelligence (IEEE ColCACI 2020), Cali, Colombia, pp. 1–6 (2020). https://doi.org/10.1109/ColCACI50549.2020.9247855

4. Li, J.P., Haq, A.U., Din, S.U., Khan, J., Khan, A., Saboor, A.: Heart disease identification method using machine learning classification in E-healthcare. IEEE Access **8**, 107562–107582 (2020)
5. Chen, M., Hao, Y., Hwang, K., Wang, L., Wang, L.: Disease prediction by machine learning over big data from healthcare communities. IEEE Access **5**, 8869–8879 (2017)
6. Ashourloo, D., Aghighi, H., Matkan, A.A., Mobasheri, M.R., Rad, A.M.: An investigation into machine learning regression techniques for the leaf rust disease detection using hyperspectral measurement. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **9**(9), 4344–4351 (2016)
7. Muhammad Usman, S., Khalid, S., Aslam, M.H.: Epileptic seizures prediction using deep learning techniques. IEEE Access **8**, 39998–40007 (2020)
8. Memon, R., Li, J., Nazeer, I., Khan, A., Ahmed, J.: DualFog-IoT: additional fog layer for solving blockchain integration problem in the internet of things. IEEE Access **7**, 169073–169093 (2019). https://doi.org/10.1109/ACCESS.2019.2952472
9. Zhang, W., Zheng, Z., Chen, X., Dai, K., Li, P., Chen, R.: NutBaaS: a blockchain-as-a-service platform. Comput. Sci. IEEE Access **7**, 134422–134433 (2019)
10. Fernandez-Carames, T.M., Fraga-Lamas, P.: Towards post-quantum blockchain: a review on blockchain cryptography resistant to quantum computing attacks. IEEE Access, 1–1 (2020)
11. Xu, Y., Huang, Y.: Segment blockchain: a size reduced storage mechanism for blockchain. IEEE Access, 1–1 (2020)
12. Ribeiro, B.R., Henirques, J.H., Marques, A.M., Antunes, M.A.:Manifold learning for premature ventricular contraction detection. In: 2008 Computers in Cardiology, Bologna, Italy, pp. 917–920 (2008). https://doi.org/10.1109/CIC.2008.4749192
13. Xu, Y.: Section-blockchain: a storage reduced blockchain protocol, the foundation of autotrophic decentralized storage architecture. In: 2018 23rd International Conference on Engineering of Complex Computer Systems (ICECCS), pp. 115–125. IEEE (2018)
14. Xu, Q., Aung, K.M.M., Zhu, Y., Yong, K.L.: A blockchain-based storage system for data analytics in the internet of things. In: New Advances in the Internet of Things, pp. 119–138. Springer (2018)
15. Ren, Y.J., Leng, Y., Cheng, Y.P., Wang, J.: Secure data storage based on blockchain and coding in edge computing. Math. Biosci. Eng. **16**, 1874–1892 (2019)
16. Hafid, A., Hafid, A.S., Samih, M.: New mathematical model to analyze security of sharding-based blockchain protocols. IEEE Access **7**, 185447–185457 (2019)
17. Zhaofeng, M., Xiaochang, W., Jain, D.K., Khan, H., Hongmin, G., Zhen, W.: A blockchain-based trusted data management scheme in edge computing. IEEE Trans. Ind. Inf. 1–1 (2019)
18. Mueller, M., Wagner, C.C., Stanislaus, R., Almeida, J.S.: Machine learning to predict extubation outcome in premature infants. In: The 2013 International Joint Conference on Neural Networks (IJCNN), Dallas, TX, USA, pp. 1–6 (2013). https://doi.org/10.1109/IJCNN.2013.6707058
19. Lei, K., Du, M., Huang, J., Jin, T.: Group chain: towards a scalable public blockchain in fog computing of IoT services computing. IEEE Trans. Serv. Comput. **13**(2), 252–262 (2020)
20. Li, Z., Kang, J., Yu, R., Ye, D., Deng, Q., Zhang, Y.: Consortium blockchain for secure energy trading in the industrial Internet of Things. IEEE Trans. Ind. Informat. **14**(8), 3690–3700 (2018)
21. Paik, H.-Y., Xu, X., Bandara, H.M.N.D., Lee, S.U., Lo, S.K.: Analysis of data management in blockchain-based systems: from architecture to governance. IEEE Access **7**, 186091–186107 (2019)
22. Zeng, X., Liao, Y., Liu, Y., Zou, Q.: Prediction and validation of disease genes using HeteSim scores. IEEE/ACM Trans. Comput. Biol. Bioinf. **14**(3), 687–695 (2017)
23. Ansarullah, S.I., Kumar, P.: A systematic literature review on cardiovascular disorder identification using knowledge mining and machine learning method. Int. J. Recent Technol. Eng. **7**(6S), 1009–1015 (2019)

24. Murphy, B.M., Goulding, R.M., O'Toole, J.M.:Detection of transient bursts in the EEG of preterm infants using time–frequency distributions and machine learning. In: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, QC, Canada, pp. 1023–1026 (2020). https://doi.org/10.1109/EMBC44109.2020.9175154
25. Mbarak, A., Yetis, Y., Jamshidi, M.:Data—based pollution forecasting via machine learning: case of Northwest Texas. In: 2018 World Automation Congress (WAC), Stevenson, WA, USA, pp. 1–6 (2018). https://doi.org/10.23919/WAC.2018.8430438
26. Davagdorj, K., Lee, J.S., Park, K.H., Ryu, K.H.: A machine-learning approach for predicting success in smoking cessation intervention. In: 2019 IEEE 10th International Conference on Awareness Science and Technology (iCAST), Morioka, Japan, pp. 1–6 (2019). https://doi.org/10.1109/ICAwST.2019.8923252
27. Leon, C., Carrault, G., Pladys, P., Beuchee, A.: Early detection of late onset sepsis in premature infants using visibility graph analysis of heart rate variability. IEEE J. Biomed. Health Inform. https://doi.org/10.1109/JBHI.2020.3021662
28. Wang, W., Chen, L., Jie, J., Wang, H., Xu, X.: A novel hybrid algorithm based on baldwinian learning and PSO. In: 2010 International Conference on Computational Aspects of Social Networks, Taiyuan, China, pp. 299–302 (2010). https://doi.org/10.1109/CASoN.2010.73
29. Graff, C.A., Coffield, S.R., Chen, Y., Foufoula-Georgiou, E., Randerson, J.T., Smyth, P.: Forecasting daily wildfire activity using poisson regression. IEEE Trans. Geosci. Remote Sens. **58**(7), 4837–4851 (2020). https://doi.org/10.1109/TGRS.2020.2968029
30. Vahabi, N., Yerworth, R., Miedema, M., van Kaam, A., Bayford, R., Demosthenous, A.: Deep analysis of EIT dataset to classify apnea and non-apnea cases in neonatal patients. IEEE Access **9**, 25131–25139 (2021). https://doi.org/10.1109/ACCESS.2021.3056558
31. Quinn, J.A., Williams, C.K.I., McIntosh, N.: Factorial switching linear dynamical systems applied to physiological condition monitoring. IEEE Trans. Pattern Anal. Mach. Intell. **31**(9), 1537–1551 (2009). https://doi.org/10.1109/TPAMI.2008.191