Intelligent Automation & Soft Computing
DOI:10.32604/iasc.2022.021920
Article

Service Level Agreement Based Secured Data Analytics Framework for Healthcare Systems

S. Benila[1,*] and N. Usha Bhanu[2]

[1]Department of Computer Science and Engineering, SRM Valliammai Engineering College, Tamil Nadu, 603203, India
[2]Department of Electronics and Communication Engineering, SRM Valliammai Engineering College, Tamil Nadu, 603203, India
*Corresponding Author: S. Benila. Email: sbenila@gmail.com
Received: 19 July 2021; Accepted: 09 September 2021

Abstract: Many physical objects are connected to the internet in this modern day to make things easier to work based on the convenience of the user, which reduces human involvement with the help of Internet of Things (IoT) technology.This aids in the capture of large amounts of data, the interchange of information via the internet, and the remote operation of machines. IoT health data is typically in the form of big data and is frequently coupled with the cloud for secure storage. Cloud technology provides a wide range of technological services via the internet, and it is a highly interoperable and on-demand network for a wide range of computing resources. The Service Level Agreement (SLA) is made between the cloud and the patient, and it outlines the services supplied as well as the level of security provided to the user. For fulfilling service, the deployed external cloud has challenges with load balancing and work scheduling. Furthermore, the gathered health data must be effectively processed by medical practitioners. To solve this issue, a Secure Cluster Naive Bayes (CNB) framework is proposed, both with and without Dimensionality Reduction. To preserve its anonymity, the obtained data is hashed and stored in the cloud using the blockchain technology. SLA sessions are organized to prioritize patient data for decryption and prediction. At the doctor's end, the decrypted data is first filtered and dimensionally reduced before being clustered using a dual K-means clustering technique and classified using the Naive Bayes algorithm. The web-based graphical user interface server is responsible for connecting the IoT device, the cloud, and the doctor. The security performance of the DRCNB and CNB frameworks is evaluated using block chain characteristics, the frequency of SLA violations, and processing and execution time. The DRCNB framework is 91.1% accurate, while the CNB model is 80.73% accurate, making it more accurate than previous models. The new models exceed the prior ones in terms of both security and prediction performance.

Keywords: IoT; healthcare; DRCNB; blockchain; security; prediction

1 Introduction

The Internet of things (IoT) is a technology that has been established with small, flexible smart devices that have the ability to gather and transmit data from any location, regardless of time or environment. Most IoT devices are transmitted either from machine to person or between people and machines [1], which is the most common scenario. The Internet of Things (IoT) has become well-established in the field of health care due to its simple and effective monitoring aspects. The majority of patient records today are kept in both electronic and physical formats. As a result, patients who attempt to switch to another health-care system may find their health records are inaccessible, which may lead to problems such as the use of harmful drugs [2]. It is possible for patients and the healthcare industry to transfer information in electronic form with the help of Internet of Things devices [3]. Additionally, the Internet of Things devices are capable of providing a variety of health-related benefits, such as monitoring chronic diseases and the fitness programmes associated with them, through the use of ubiquitous sensors [4].

In contrast, with the increasing complexity of networks and the need for security and privacy over massive amounts of data [5,6], the inherent disadvantages of IoT, in terms of available energy and computation capability, are becoming increasingly apparent. In addition, the unification of patient data at regular intervals of time represents a significant challenge. Furthermore, an IoT-based healthcare system must improve the personalization of processes in healthcare and support the management of clinical workload in order to be effective. Supporting self-care diagnostic processes [7] that are fully automated is the only way to achieve both goals.

Prior work has integrated IoT with cloud computing in the healthcare system in order to address the issues of resource constraint and to improve patient outcomes. The issues of security and privacy in relation to health data continue to be a concern. The proposed Dimensionality Reduction Cluster Nave Bayes (DRCNB) framework incorporates both the Internet of Things (IoT) cloud setup for handling patient data and block chain technology to provide a comprehensive solution. The block chain technology is used to add patient data to the chain if there is a need to include more at any time in the future, as well as to complete the transaction and store the data. It is necessary to use the hash function in the block chain in order to obtain the hash codes, and it will be used to connect the two blocks [8,9]. The proposed framework incorporates a dimensional reduction technique to reduce the vastness of large amounts of data collected from patients, and it is clustered based on similarities between the groups of data collected. The Nave Bayes algorithm is used to classify the clustered data, and the results are then provided to the doctor as a report.

The contribution to the proposed DRCNB framework for secure prediction of patient data is as follows:

a) The block chain technique is employed to secure and accumulate patient data effectively in the cloud environment.
b) The importance of dimensional reduction in handling health care data is discussed.
c) The clustering was carried out on the extracted data for effective prediction with less time consumption.
d) The Structure Extended Multinomial Naive Bayes (SEMNB) is used to classify the data along with the soothing technique.
e) The proposed model enhanced both security and predictive performance over the health care data.

2 Related Works

A Service Level Agreement (SLA) is a contractual method that specifies and governs service delivery among service providers, consumers, and other parties. Applications in IoT are mostly outsourced to the cloud to carry out some tasks, and the role of SLA is significant in inter-system

communication [10]. The majority of the time, Service Level Objectives (SLOs) are used to define the various QoS parameters. The several metrics that include availability, latency, scalability, and throughput. Quality of service metrics should be clearly defined in a system. The block chain can play an active role, such that activities are recorded on the shared ledger. Activities are represented as transactions, which must be completed in accordance with the SLA clauses that have been established. Participants, acting as block chain validators, need to ensure the ledger integrity by agreeing on the compliance of these activates [11]. The security of the block chain was proposed to secure the data in the cloud by adapting block chain security. The cloud, with its efficiency and availability, is easily adapted to the IT industry. To enhance data security, the block chain and its associated technology are discussed with their advantages and drawbacks [12]. Misura et. al. proposed a cloud-based mediator platform where automatic negotiation is performed between device owners and application owners. The mathematical model they adopted is a linear combination of previous offers, which is similar to the behavior-dependent tactic. However, this approach does not consider IoT service properties and context information [13]. Alkayal et. al. proposed a particle swarm optimization-based negotiation model to minimize the negotiation time and increase the throughput [14]. However, the common disadvantage shared by these approaches is the long negotiation time, which may increase significantly with the increase in tasks or negotiable parameters. System stability and resource usage should be at the same level as optimizing the IoT resource allocation [15].

One of the challenges in resource allocation is QoS limitations. The IoT system must guarantee QoS and maximize the provider's profit. Jiasi Weng et al. [16], proposed a deep chain framework with a deep learning algorithm and a block chain based incentive for auditable and privacy-preserving. Depending on the block chain mechanism, it forces the participants to behave correctly. It assured data privacy for participants and the whole training process to provide auditability. Jie Xu, Kaiping Xue et al. [17], proposed a privacy-preserving health chain scheme using the block chain for large scale health data. Two block chains are established to guarantee that both patient data and doctor's diagnosis cannot be tampered with. The proposed framework provides the user with leverage to revoke the doctors. The proposed health chain is found to be effective and feasible for health care systems. The IoT resource allocation challenges are discussed from an architectural point of view and grouped the works based on resource allocation tiers. However, the work on algorithmic optimization procedures for allocating resources in the IoT does not get much attention [18]. As a result, IoT services are supplied by monitoring data in the access network's physical domain and analysing data in the cloud's computing domain.

The server examines a large amount of data before making a conclusion. A data reduction system is required for the server's data analysis in order to receive trustworthy data. Fewer input dimensions frequently imply fewer parameters or a simpler machine learning model structure. A model with too many parameters is more likely to over fit the training data and so perform poorly on new data. Simple models that generalize effectively, as well as input data with few input variables, are preferred. This is especially true for linear models, where the number of inputs and the model's input variables are frequently linked. Prior to modelling, dimensionality reduction is applied as a data preparation technique [19]. It could be done after cleaning and scaling the data and before training a prediction model. Dimensionality reduction can be accomplished using a variety of methods. Wrapper methods and filter methods are the two basic types of feature selection approaches. The server leads to the correct decision by filtering the vast data and applying reliable data. Furthermore, by minimizing the computing load in data processing, the server can use less energy for computing. Because IoT devices generate a lot of data traffic, including malfunctioning data in the physical sensing domain, the data-filtering system should be able to pass useful normal data among them.

From the literature survey, it was observed that the block chain is found to be a solution for enabling effective data security and, hence, it can be used for accumulating individual patient data under a single

block. The SLA based scheduling with patient data is not yet explored effectively for connected healthcare applications.

3 Proposed System Framework

The proposed DRCNB and CNB framework comprises of two modules, namely, the security module and the prediction module. The entire proposed framework is shown in Fig. 1. The two modules are integrated with the web GUI server. The security module starts with collection of patient data and terminates with storing the data in the block chain. The prediction model begins with the decrypting of the patient data from the block chain and terminates with the analysis of predicted data by the doctor. The web GUI server gets patient information from the cloud server, as well as doctor recommendations, and transmits it to patients via IoT devices.

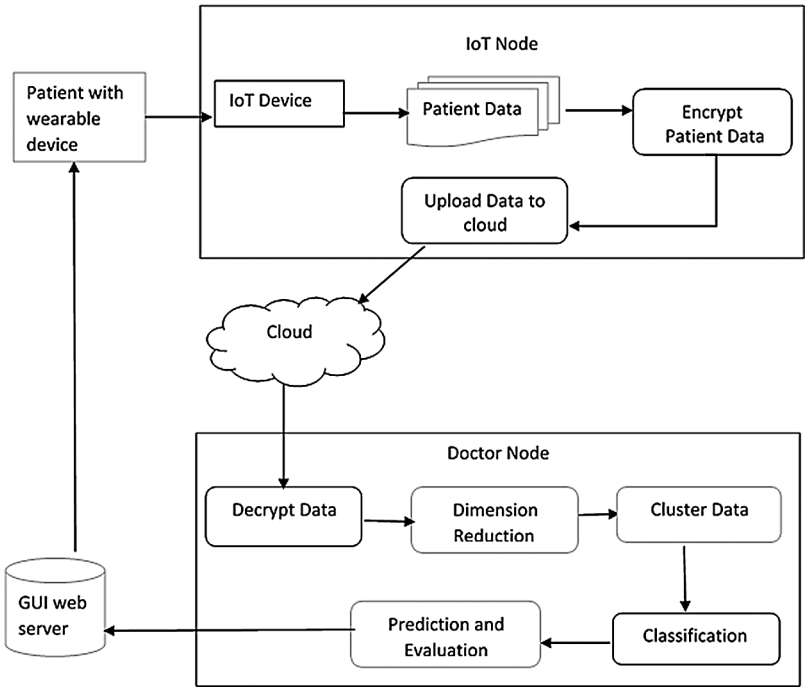

Figure 1: Proposed DRCNB and CNB framework

Steps involved in the DRCNB framework is given below.

Step 1: Wearable gadgets are supplied to the patient.

Step 2: The sensors capture the key parameters and transfer them to the cloud for processing and analysis.

Step 3: Using the SHA-256 encryption algorithm, encrypt the patient data, and then store the encrypted blocks in the blockchain hyperledger.

Step 4: The blocks are uploaded to a cloud server, processed and analysed

Step 5: The patients with varied degrees of disease severity are observed by the system. Depending on the severity of the disease, the doctor has access to the patient's data at the SLA level.

Step 6: The patient data is decrypted using the SHA-256 decryption algorithm.

Step 7: The decrypted data is reduced in dimension using the Range Set Filtering Algorithm.

Step 8: The data that has been reduced in dimensionality is supplied into the clustering module, which performs dual clustering.

Step 9: The data is clumped and provided to the classification module.

Step 10: Once the categorization and prediction results are complete, the doctor examines the results.

Step 11: The results are returned to the patient via the GUI webserver.

The functions performed by the DRCNB frameworks are as follows.

3.1 Security Module

The security module provides security to the patient data in the proposed framework. The security module consists up of following components to provide the essential security for the data.

IoT node: The IoT node is the primary node that collects the patient data. It records all the data through continuous monitoring. The data collected are stored in the form of packets and accumulated before transferring it into the cloud.

Cloud: The cloud serves to be an external storage to secure the patient data. The data that are collected in the IoT devices are transmitted to the cloud. The data are accumulated through the block chain technology using the hashing algorithm. The data of each patient is stored in distinct block chain and made available for the doctors after decryption. The cloud server shares the patient information to the web GUI server.

3.1.1 Block Chain

Block-chain is a method to compress the transactions in blocks form in block chain that are linked through the hash. In block-chain has the block, it consists of block header and a transaction representation as in Fig. 2.

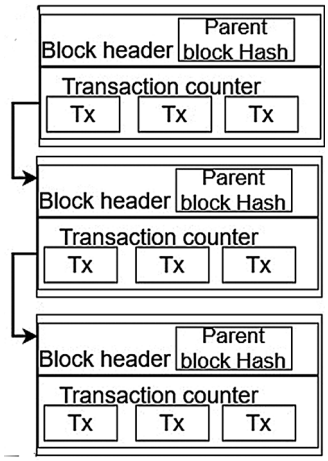

Figure 2: Block chain

Block version: It defines the set of rules for validating the block

Merkle Tree Root Hash: All transactions hash values are stored in the block

Time stamp: representation of universal present time in seconds

nbits: valid target block hash threshold

Parent block: hash value with 256-bit size that connects the previous block

In the present framework, the streaming patient data are block chained through the Hyperledger Fabric blockchain. The uniqueness of the hyper-ledger is that it contains three distinct steps in the transaction flow that includes the block creation, endorsement and validating step. The hash code function for the proposed model is SHA 256. The data is encrypted into a single line 256 bit encrypted text is secured in Hyperledger block storage. Every block enclosed the encrypted individual patient data which can be consulted by clinicians or staff for patient care. The Fig. 3 shows the input data and its fields that are transformed into hashes based on SHA 256. The hashed data is stored in the block chain, thereby protecting the privacy of the patient data. All the hashed data was de-hashed by the medical practitioner to perform the necessary analysis over it.

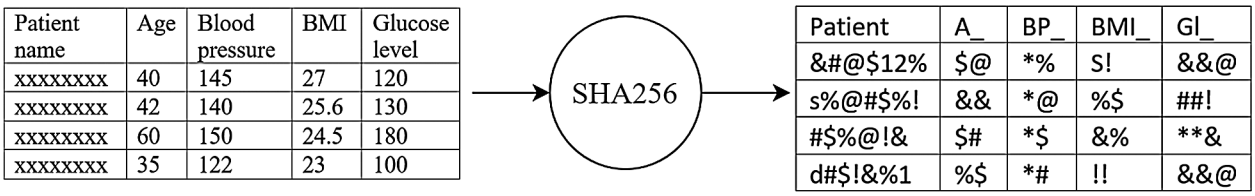

Figure 3: Hashing of fields in the patient data

3.1.2 SLA Negotiations

The DRCNB is centred on the SLA, which selects patients based on three distinct factors. The first criterion is the patient's health condition (HC), the second is disease severity (DS), which varies depending on the nature of the disease, and the final criterion is emergency (EM). All of these criteria are derived from patient streaming data across several health indicators. The criteria are graded on a scale of 0 to 10. The patient's priority is determined by the product of the three criteria stated in equation.

Priority of SLA = (0.2HC × 0.35DS × 0.45EM)

When the SLA priority value is larger than 7.5, it is regarded high priority, when it is between 3 and 7.5, it is considered medium priority, and when it is less than 3 it is considered low priority.

3.2 Prediction Module

The prediction module is employed to predict the patient data through proper dimensional reduction and classification. The data stored in the cloud is obtained through de-hashing of the block chain of patient.

3.2.1 Dimensional Reduction for DRCNB

Inappropriate data is abundant in raw data from many IoT sensors. To acquire the desired results, this raw data must be cleaned and processed using a variety of data cleaning and data analysis procedures. Furthermore, a large amount of filthy and meaningless data causes overutilization of the system and high computing costs, particularly when processing the raw data, which results in undesirable changes and modifications. To prevent analysing raw data, the high-dimensional input data is reduced to a finite set of information before data analysis using a dimension reduction process. Principal Component Analysis (PCA) is a popular dimension reduction technique. PCA has high computational complexity and time complexity. When dealing with huge data sets with a significant number of properties and instances, filter-based methods are quite successful.

In this paper, the Range Set filtering algorithm is used as a filteration method to reduce the dimension. The heart disease data set from the UCI repository has 14 attributes related to heart disease. There are a 'P' number of patients having multiple heart disease parameters v1, v2… vn. Each row represents a particular patient and the columns represent the corresponding disease parameters. The patient matrix *Pmat* can be represented as

$$Pmat = \begin{bmatrix} v11 & v12 & \ldots.v1n \\ v21 & v21 & \ldots v2n \\ & \vdots & \\ & \vdots & \\ vn1 & vn2 & \ldots\ldots vnn \end{bmatrix} \quad (1)$$

It is a time-consuming operation to consider a huge number of parameters for a cardiac patient, some of which are irrelevant to the disease. As a result, the range set filtering technique discards some parameters while keeping the most important ones. The properties are statistically measured, and the relationships between them are considered. The following measurements were used to determine the relationship: correlation co-efficient (μ) and threshold value (Ω). Multiple attributes are joined and the occurrences of each attribute are calculated. If the number of repetitions exceeds the threshold, the arrtibutes are recognised as primary components. If the value is less than the threshold, the attributes are discarded.

Algorithm 1: Range set filtering Algorithm

Input: All variable data set S= {v1, v2…..vn}

Output: Dimension reduced data set S_d = {j1, j2……jn}

Step 1: Access the data set

Step 2: Establish a threshold for each attribute.

Step 3: In Pmat, read each patient p.

Step 4: Go over each attribute v in V.

Step 5: compute the variance for each attribute

Step 6: Examine the level of variation

Step 7: If the variance is less than the threshold, the attribute is removed from the data set.

Step 8: Accept the variable if the variance exceeds the threshold.

Step 9: Repeat Steps 7–9 for all attributes.

Step 10: Find the correlation between two attributes.

Step 11: Determine the correlation threshold for each attribute pair.

Step 12: Repeat Step 10 for all accepted attributes in Step 9.

Step 13: Compute the occurrences of each acceptable pair in Step 10.

Step 14: Examine the occurrences in relation to the threshold.

Step 15: Accept if the value is greater than the threshold; otherwise, reject.

Step 16: Return the set of attributes and occurrences.

Step 17: The dimension reduced data set is obtained.

Once the dimension has been lowered, the data set is passed to the clustering algorithm for grouping based on the critical parameters.

3.2.2 Clustering of Data

The dimensionally reduced feature data are clustered using the dual K-means clustering technique in the proposed framework. The obtained data are initially clustered using the Simple K-means algorithm based on Euclidean distance and are subsequently scattered.

$$\text{Euclidean distance } d' = \sum_{i=1}^{n} \sqrt{(x_i - y_i)^2} \tag{2}$$

After the normalization, Hierarchical K-means is applied to form the clusters through the dissimilarity matrix. If points i and j are agglomerated into cluster i ∪ j, then the new dissimilarity between the cluster and all other points are calculated as:

$$d(i \cup j,\ k) = \alpha_i d(i,k) + \alpha_j d(j,k) + \beta d(i,j) + \gamma |d(i,k) - d(j,k)| \tag{3}$$

Here, αi , αj , β, and γ define the agglomerative criterion. This provides the effective clustering of the data based on the feature for analysis.

Algorithm 2: Dual K means clustering

Input: Dimension reduced data, d is the minimum distance

Output: Clusters

Step 1: Load the dimension reduced data

Step 2: Establish the initial centroid data

Step 3: Estimate the Euclidian distance between the centroid and other data, d' with Eq. (1)

Step 4: if d < d'

Step 5: Repeat step 2 and 3

Step 6: else

Step 7: Perform scattering of data and establish the subplot

Step 8: Estimate the dissimilarity matrix

Step 9: Merge the closest paired clusters

Step 10: Update the dissimilarity matrix

Step 11: Repeat step 9 and 10, until maximal clusters are obtained

The K-means method is crucial in this strategy for identifying the proper number of data groups. This approach computes centroids and Euclidian distance for numerous diagnostic attributes. The mean value is applied for the given data set, and the patient condition is now established. The patient is more likely to have heart disease if his or her mean value is closest to the sample mean value.

3.2.3 Classification

Naive Bayes classifiers are extremely flexible, demanding a number of parameters continuous in the number of features in a learning issue. Despite their simplicity, Naive Bayes classification algorithms routinely outperform more advanced classification approaches in terms of accuracy. The Naive Bayes model is used to determine the characteristics of persons who are suffering from heart disease based on the conditional probability between two attributes. It indicates the chance of predictable state for each of the input attributes that are given.

The Naive Bayesian classifier (NBC) is based on the assumption that, given the class variable, all characteristics are independent:

$$P(Ci \mid A1,\ A2, \ldots,\ An)\ =\ P(Ci)P(A1 \mid Ci)P(A2 \mid Ci)\ldots P(An \mid Ci)\ P(An) \tag{4}$$

The classification is defined as

$$\arg_{cv}\ [max[P(C_v \mid A)\ =\ P(C_v)P(A \mid C_v)/P(A)]]) \tag{5}$$

P(A) does not vary with class and may be viewed as a constant for the maximization process.

The basic framework of a naive Bayesian classifier is updated using Semi Naive Bayes classifier (SEMNB), while keeping the structure of the network. The SEMNB removes or joins properties that are dependent on the provided class. The classification is carried with the novel Naïve Bayes that includes the smoothing function over the existing SEMNB [20] as,

$$C_M = \arg max_{Cv}\, P(C_v) \prod_{i=1}^{n} P_\lambda \left(w_i | \prod_{w_i}, C_v \right)^{f_i} \tag{6}$$

$$C_M = \arg max_{Cv}\, P(C_v) \prod_{i=1}^{n} (1-\lambda) \left(w_i | \prod_{w_i}, C_v \right)^{f_i} \tag{7}$$

where

C_M = predicted class label for data (M)

C_v = class variable

$P(C_v)$ = Probability of class variable

λ = smoothing factor

w_i = word in the document

f_i = Feature vector

This framework generates two types of classification results. One is the presence of cardiac disease, while the other is the absence of heart rate. The accuracy, precision, recall, F measures as well as the AUC for ROC, are all measured.

4 Simulation and Setup

The proposed DRCNB framework is implemented to assess its performances. The doctor node and user node set are developed with 64-bit Intel core processor with frequency of 2.45 GHz. Java programming language is employed for transaction prototyping. XAMPP open source web server is used provide GUI interface to connect patient and doctor nodes. For training and testing the model, the dataset from the UCI repository has been used which contain the data for heart disease is employed for estimating the performance of DRCNB and CNB framework. Although there are 76 attributes in this database, all published studies only use a subset of 14 of them. The data set contains 303 instances. The parameters are given in Tab. 1.

Table 1: Dataset-parameter

S.no	Feature	Data Type	Description
1	Age	Numeric	Age in years
2	Gender	Numeric	Gender of person
3	Cp	Numeric	class of pain
4	Trestbps	Numeric	Resting blood pressure in mmHg
5	Chol	Numeric	Serum cholesterol in mg/dl
6	Fbs	Numeric	Fasting blood sugar in mg/dl
7	Restecg	Numeric	pattern of ECG in rest
8	Thalach	Numeric	heart rate
9	Exang	Numeric	angina due to exercise
10	Oldpeak	Numeric	ST depression induced by exercise relative to rest
11	Slope	Numeric	Peak exercise ST segment slope
12	Ca	Numeric	No.of fluoroscopy colored vessel
13	Thal	Numeric	Status of defect
14	Target	Numeric	Predicted class; 1 for heart disease; 0 for no heart disease

5 Result and Discussion

5.1 Capacity of Block

The block header capacity is the total bytes in each patient block header. In the proposed model, the updated block header and block body capacity is given in Tab. 2.

Table 2: Block header and body parameters

Block Header		Block body	
Parameters	Length (Bytes)	Parameters	Length (Bytes)
Pre-hash	16	ID	32
Merkle root	32	Hash	64
Index	32	Signature	64
Nonce	8	Ts	4
Gtime	8	Asymmetric encryption	256

5.2 SLA Violations

The foremost focus of the proposed framework is to ensure that there is a minimization of SLA violation in the IoT-cloud based health care system. The performance of the proposed framework is compared with the existing Parallel Semi Naïve Bayes model (PSNB) [21] based on the priorities of jobs which can have low, medium and high categories. The Tab. 3 and Fig. 4 provides the performance in SLA violation.

Tab. 3 shows the proposed CNB and DRCNB provides lower SLA violations compared with the existing PSNB model with respect to low, medium and high SLA priority categories. In high priority patients, the SLA violations of DRCNB is zero. It indicates no violations found for high SLA priority.

Table 3: Performance on SLA violations

Jobs	10	20	30	40	50	60
Low-PSNB	0	1	1	2	3	3
Low-CNB	0	0	1	1	2	2
Low-DRCNB	0	0	0	0	1	1
Medium-PSNB	0	0	1	1	1	2
Medium-CNB	0	0	0	1	1	1
Medium-DRCNB	0	0	0	0	0	1
High-PSNB	0	0	0	1	1	2
High CNB	0	0	0	0	1	1
High-DRCNB	0	0	0	0	0	0

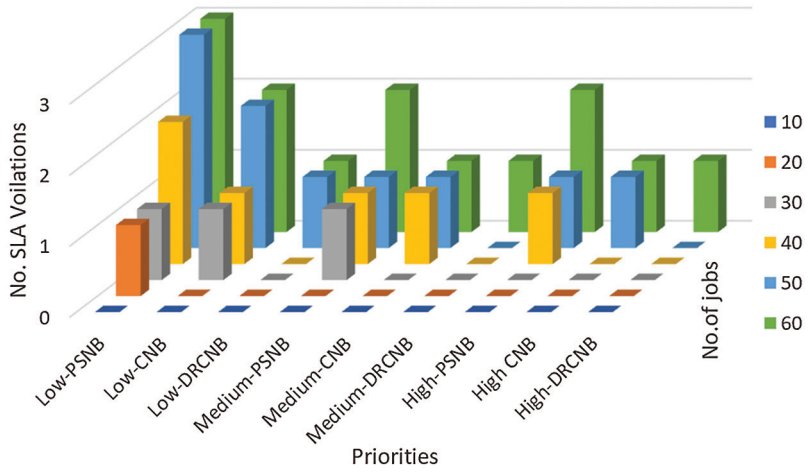

Figure 4: SLA violations for various priority jobs

5.3 Processing Time of Transactions

The proposed model involves the hashing and de-hashing process in the cloud along with the storage and retrieval of data in the security module. The time for dimensional reduction, clustering and classification along with predictive analysis is enclosed in the prediction model. The processing time differ with the number of nodes involved in the framework. The usage of hyper-ledger improved storage efficiency with auditing and privacy of logs and data. Additionally, the process of dimensionality reduction improved the processing of data and henceforth the processing time for patient data is less comparatively. The comparison on processing time is carried out with the existing PSNB model [20] under fixed 1 GB data. The processing time for both the proposed DRCNB and CNB along with existing PSNB is given in Tab. 4 and Fig. 5.

According to Tab. 4 and Fig. 5, the processing time in CNB is long. Because dimensionality reduction is not used in CNB, it must process a vast volume of data. As a result, the execution time is prolonged. However, because dimensionality reduction is used in DRCNB, the processing time is reduced when compared to other approaches. As compared to other dimensionality reduction techniques, filtering-based algorithms require less processing time.

Table 4: Processing time

No. of nodes	PSNB	CNB	DRCNB
1	1360	1440	1000
2	1100	1168	910
3	900	952	750
4	580	619	480
5	400	439	320
6	360	388	295
7	320	344	260
8	280	302	230
9	250	282	210
10	240	268	190

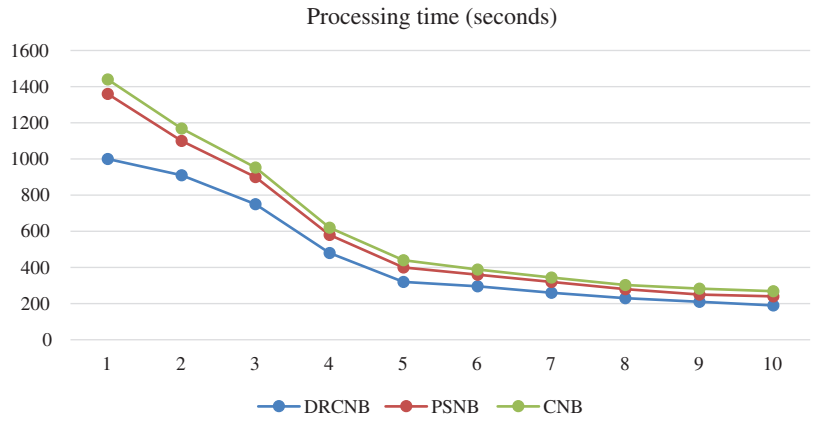

Figure 5: Comparision of processing time

5.4 Execution Time

The execution time for the proposed DRCNB and CNB model for a fixed node (10) is evaluated and compared with the existing PSNB model. The execution time involves the time taken for block storage and prediction through Navies Bayes. The comparison over the proposed and existing model is given in Tab. 5 and Fig. 6.

Similar to the processing time the usage of dimensional reduction approach and the hyper ledger reduced the time complexity in executing the large dataset in an effective manner.

5.5 Performance of Prediction Model

The performance of the prediction model with Navies Bayes in the proposed framework is estimated with the parameters like accuracy, recall and precision along with its F measures based on confusion matrix. The accuracy of the prediction is about 91.10% for DRCNB and it is about 80.73% while the precision and the recall of the prediction is about 96.4% and 88.88% for DRCNB. The corresponding precision and recall value for CNB is about 86.21% and 88.03% for CNB. The F measure for the prediction in the proposed frameworks is about 94.08% for DRCNB and 87.11% for CNB. The AUC for

ROC is about 90.5% for DRCNB and it is about 81.5% for CNB. The result obtained through prediction is presented in the Fig. 7. It was found that the proposed prediction model performance is better than the existing SVM and decision tree [22] as shown in Tab. 6 and Fig. 7.

Table 5: Execution time

Size of data (GB)	PSNB	CNB	DRCNB
0.2	200	230	150
0.3	210	245	160
0.4	200	262	165
0.5	200	262	170
0.6	205	295	170
0.7	220	295	170
0.8	230	295	180
0.9	250	320	195
1	270	320	210

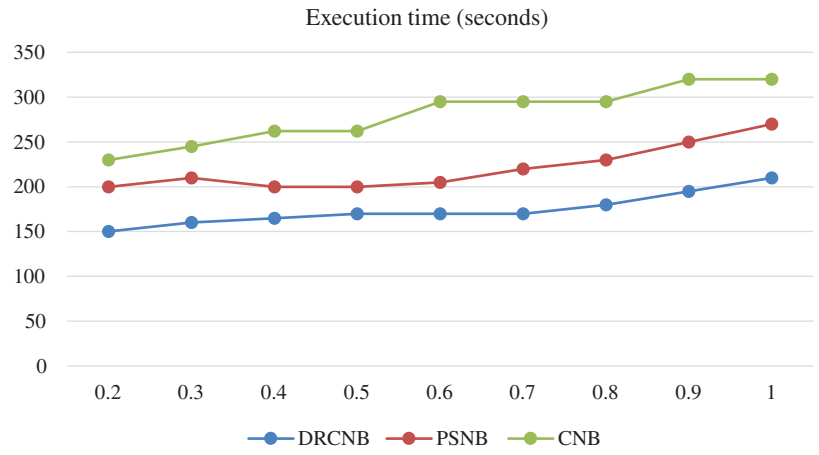

Figure 6: Comparison on execution time

Without the dimensionality reduction the time incurred in the proposed model is higher the PSNB. The reduced dimensionality provided the most significant features from the dataset and hence the prediction performance over the proposed DRCNB is better than CNB model.

Table 6: Performance of the proposed CNB, DRCNB with existing models

Metrics	Decision tree	SVM	CNB	DRCNB
Accuracy	0.738	0.651	0.8073	0.911
Precision	0.735	0.425	0.8621	0.964
Recall	0.738	0.651	0.8803	0.888
F-measures	0.736	0.513	0.8711	0.9408
AUC in ROC	0.751	0.5	0.815	0.905

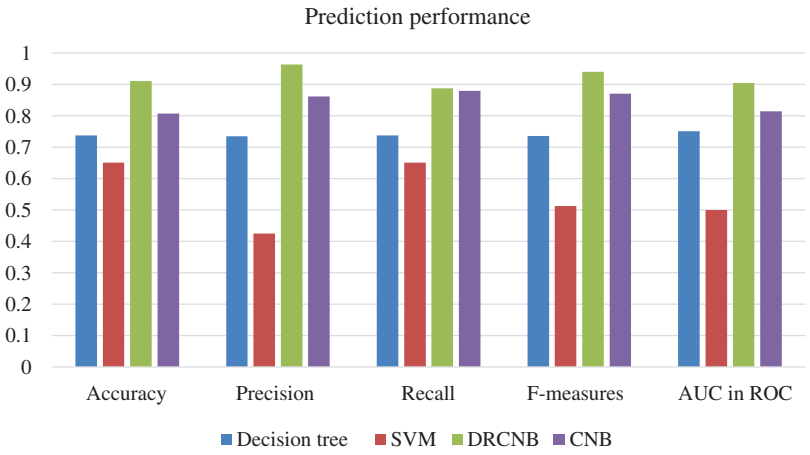

Figure 7: Comparison on prediction performance

6 Conclusion

To secure IoT-cloud-based health care data, the DRCNB and CNB frameworks have been proposed. Two modules are included in the proposed frameworks, one for security and the other for prediction. IoT data is gathered and transferred to the cloud for processing and storage. The data is stored in a block chain with a hash function and de-hashed for prediction. With the Range Set Filtering algorithm, de-hashed data is reduced in dimension. Using the dual K-means clustering algorithm, dimensionally reduced data is clustered. A Semi-Naive Bayes Classifier with a smoothing factor is used to classify clustered data. The cloud, IoT devices, and doctors all benefit from the web GUI's ability to transfer and accumulate information. With fixed nodes and fixed data sizes, the performance of the proposed DRCNB framework is measured for both execution time and processing time. The execution time increases as the file size increases, and the processing time decreases as the number of nodes increases. Also, dimensionality reduction and hyper ledger usage were found to provide an advantage over time complexity in processing and executing the data. The DRCNB framework has an accuracy rate of 91.1%, while the CNB model has an accuracy rate of 80.73%, making it more accurate than existing models. The future scope of the proposed framework is to improve security and prediction performance with real-time implementation.

Acknowledgement: We thank all the people who have given suggestions and contributions to bring this work for publication.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] P. Sethi and S. Sarangi, "Internet of things: Architectures, Protocols and Applications," *Journal of Electrical and Computer Engineering*, vol. 2017, pp. 1–25, 2017.

[2] F. Ozair, N. Jamshed, A. Sharma and P. Aggarwal, "Ethical issues in Electronic Health Records: A general overview," *Perspectives in Clinical Research*, vol. 6, no. 2, pp. 73–76, 2016.

[3] K. J. Serrano, M. Yu, W. T. Riley, V. Patel, P. Hughes *et al.,* "Willingness to exchange health information via mobile devices: Findings from a population-based survey," *Annals of Family Medicine*, vol. 14, no. 1, pp. 34–40, 2016.

[4] K. Ullah, M. A. Shah and S. Zhang, "Effective ways to use Internet of Things in the field of medical and smart health care," in *Proc. ICISE*, Islamabad, Pakistan, 2016.

[5] M. Seliem, K. Elgazzar and K. Khalil, "Towards privacy preserving IoT environments: A survey," *Wireless Communications and Mobile Computing*, pp. 15, 2018.

[6] N. N. Srinidhi, S. M. Dilip Kumar and K. R. Venugopal, "Network optimizations in the Internet of Things: A review," *Int. Journal of Engineering Science and Technology*, vol. 22, no. 1, pp. 1–21, 2019.

[7] D. V. Dimitrov, "Medical Internet of Things and big data in healthcare," *Healthcare Informatics Research*, vol. 22, no. 3, pp. 156–163, 2016.

[8] P. K. Kaushal, A. Bagga and R. Sobti, "Evolution of bitcoin and security risk in bitcoin wallets," in *Proc. Comptelix*, pp. 172–177, Jaipur, India, 2017.

[9] D. Wood, "Ethereum: A secure decentralised generalised transaction ledger," *Ethereum Project Yellow Paper*, vol. 151, pp. 1–32, 2014.

[10] A. V. Papadopoulos, S. A. Asdollah, M. Ashjaei, S. Mubeen, H. Breivold *et al.,* "SLAs for industrial IoT: Mind the gap," in *proc. FiCloudW*, IEEE, 2017.

[11] A. Ali, S. Ellis, P. Pankesh and M. Karan, "Blockchain-based SLA management in the context of IoT," *IT Professional*, vol. 21, no. 4, pp. 33–40, 2019.

[12] J. H. Park and J. H. Park, "Blockchain security in cloud computing: Use cases, challenges, and solutions," *Symmetry*, vol. 9, no. 8, pp. 164, 2017.

[13] K. Misura and M. Zagar, "Internet of Things cloud mediator platform," in *Proc. MIPRO*, Opatija, Croatia, IEEE, 2014.

[14] M. F. Abulkhair, E. S. Alkayal and N. R. Jennings, "Automated negotiation using parallel particle swarm optimization for cloud computing applications," in *Proc. ICCA*, IEEE, pp. 26–35, 2017.

[15] B. Manate, T. Fortis and V. Negru, "Optimizing cloud resources allocation for an Internet of Things Architecture," *Scalable Computing: Practice and Experience*, vol. 15, no. 4, pp. 345–355, 2014.

[16] J. Weng, J. Zhang, M. Li, Y. Zhang, W. Weng *et al.,* "Deepchain: Auditable and privacy-preserving Deep Learning with Blockchain-based Incentive," *IEEE Transactions on Dependable and Secure Computing*, vol. 14, no. 8, pp. 1, 2019.

[17] J. Xu, K. Xue, S. Li, H. Tian, J. Hong *et al.,* "Healthchain: A Blockchain-based privacy preserving scheme for large-scale health data," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 8770–8781, 2019.

[18] F. C. Delicato, P. F. Pires and T. Batista, "The resource management challenge in IoT," *Resource Management for Internet of Things*. 1st ed., vol. 1. Springer, pp. 7–18, 2017.

[19] L. V. D. Maaten, E. Postma and J. V. D. Herik, "Dimensionality reduction: A comparative review," *Tilburg Centre for Creative Computing*, pp. 36, 2009.

[20] L. Jiangab, S. Wanga, C. Lic and L. Zhanga, "Structure extended multinomial naive Bayes," *Information Sciences*, vol. 329, no. 2–3, pp. 346–356, 2016.

[21] P. K. Sahoo, S. Mohapatra and S. Wu, "SLA based healthcare big data analysis and computing in cloud network," *Journal of Parallel and Distributed Computing*, vol. 119, no. Suppl. C, pp. 121–135, 2018.

[22] D. Sisodia and D. S. Sisodia, "Prediction of diabetes using classification algorithms," *Procedia Computer Science*, vol. 132, pp. 1578–1585, 2018.