IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received November 21, 2020, accepted February 24, 2021, date of publication March 10, 2021, date of current version March 29, 2021.

*Digital Object Identifier 10.1109/ACCESS.2021.3065440*

# Blockchain-Based Secure Healthcare Application for Diabetic-Cardio Disease Prediction in Fog Computing

**P. G. SHYNU[1], (Member, IEEE), VARUN G. MENON[2], (Senior Member, IEEE), R. LAKSHMANA KUMAR[3], (Member, IEEE), SEIFEDINE KADRY[4], (Senior Member, IEEE), AND YUNYOUNG NAM[5], (Member, IEEE)**

[1]School of Information Technology and Engineering, Vellore Institute of Technology, Vellore 632014, India
[2]Department of Computer Science and Engineering, SCMS School of Engineering and Technology, Ernakulam 683576, India
[3]Head- Centre of Excellence for Artificial Intelligence and Machine Learning, Hindusthan College of Engineering and Technology, Coimbatore 641050, India
[4]Faculty of Applied Computing and Technology, Noroff University College, 4608 Kristiansand, Norway
[5]Department of Computer Science and Engineering, Soonchunhyang University, Asan 31538, South Korea

Corresponding author: Yunyoung Nam (ynam@sch.ac.kr)

This work was supported in part by the Korea Institute for Advancement of Technology (KIAT) Grant by the Korean Government through Ministry of Trade Industry and Energy (MOTIE) (The competency development program for industry specialist) under Grant P0012724, and in part by the Soonchunhyang University Research Fund.

**ABSTRACT** Fog computing is a modern computing model which offers geographically dispersed end-users with the latency-aware and highly scalable services. It is comparatively safer than cloud computing, due to information being rapidly stored and evaluated closer to data sources on local fog nodes. The advent of Blockchain (BC) technology has become a remarkable, most revolutionary, and growing development in recent years. BT's open platform stresses data protection and anonymity. It also guarantees data is protected and valid through the consensus process. BC is mainly used in money-related exchanges; now it will be used in many domains, including healthcare; This paper proposes efficient Blockchain-based secure healthcare services for disease prediction in fog computing. Diabetes and cardio diseases are considered for prediction. Initially, the patient health information is collected from Fog Nodes and stored on a Blockchain. The novel rule-based clustering algorithm is initially applied to cluster the patient health records. Finally, diabetic and cardio diseases are predicted using feature selection based adaptive neuro-fuzzy inference system (FS-ANFIS). To evaluate the performance of the proposed work, an extensive experiment and analysis were conducted on data from the real world healthcare. Purity and NMI metrics are used to analyze the performance of the rule based clustering and the accuracy is used for prediction performance. The experimental results show that the proposed work efficiently predicts the disease. The proposed work reaches more than 81% of prediction accuracy compared to the other neural network algorithms.

**INDEX TERMS** Fog computing, blockchain, clustering, classification, fuzzy, disease prediction.

## I. INTRODUCTION

Enduring technical advancements provide significant opportunities for biomedical innovation and cost savings, but also pose an obstacle for the integration of emerging technology into medical treatment [1]. A considerable volume of work is primarily focusing on smart healthcare to address conventional healthcare limitations and satisfy rising expectations for premium healthcare. Smart healthcare could be designed and developed as a range of devices, tools, software, facilities, and organizations with conventional healthcare, biosensors, connected apps, and smart emergency service systems [2]. The cornerstone of intelligent healthcare is IoT end nodes that include a wide range of medical equipment and applications that link to healthcare through the Internet. Fog computing is an extension of cloud computing which can process and archive vast quantities of data that IoT devices produce near their origins.

### A. MOTIVATION

Fog computing is considered to be one of the key technologies that contribute greatly to promoting IoT healthcare and surveillance applications as these systems are latency-sensitive and real-time tracking, data processing, and

The associate editor coordinating the review of this manuscript and approving it for publication was Amir Masoud Rahmani.

decision making are critical criteria in healthcare applications such as servicing the elderly by home nursing, heart care, diabetes and some other diseases. Health data is an important topic because it includes essential, confidential knowledge. With fog computing, the aim that patients take care of their own health data locally is realized. Those safety data are housed in fog nodes such as smart phones or smart vehicles [13]. Fog computing provides tremendous advantages for fog-based application which is prone to delay. Hong *et al.*, [12] introduced Mobile Fog, which is a globally dispersed and latency-sensitive programming paradigm for Internet applications. A variety of studies has looked into the use of fog in health care. This motivates to develop the fog based health care prediction.

### B. PROBLEM STATEMENT

Health care contributors are widely implicit in producing large volumes of information in a variety of formats, together with records, economic papers, clinical test findings, imaging tests, and vital sign assessments, etc., [10]. The comprehensive database created in care environments is expanding rapidly, with healthcare information struggling from numerous problems, with data access, and how information can be obtained beyond the healthcare ability. Blockchain provides the ability to enhance the data's authentication and legitimacy. It also helps to disseminate data inside the network or services. Such apps affect the cost, quality of data, and importance of providing health care within the system. Blockchain is a transparent, decentralized network without the middleman [14]. Blockchain healthcare networks do not need several verification rates which have access to data for anyone who is part of the infrastructure of blockchain. Data is rendered available to consumers and is transparent. Such innovations will continue to overcome the numerous problems facing the healthcare domain today.

Disease prediction is one of the main real-world problems in healthcare domain. Many classification algorithms [31], [33] are used to predicts the diseases accurately. Artificial neural network (ANN) is one of the classification algorithms. ANN is a massively computational parallel model with self-adaptive and self-learning capabilities, because of its large parallel structure; it takes more time to predict the outcome. ANN is not appropriate for dealing with such issues, such as ambiguous and imprecise data for which problems of uncertainty may occur at any point of the process of classification.

Fuzzy logic is used to resolve this issue in order to translate the numeric input features into their corresponding linguistic terminology. Based on linguistic properties such as low, medium and high, each input function is transformed into its corresponding membership values in this fuzzification process. Similarly, from the input features, all linguistic characteristics are extracted. By deciding the membership value in different linguistic terms, fuzzy logic is also sufficient to deal with the ambiguity problem. Adaptive Neuro Fuzzy Inference System (ANFIS) is a hybrid model which adopts the characteristics of ANN and fuzzy logic.

### C. CONTRIBUTIONS

The objective of this paper is to develop the disease prediction model using feature selection and ANFIS. Feature selection is the one of the pre-processing technique which reduces the size of the dimensionality of the dataset. This paper use Cronbach's alpha [41] for optimal feature selection.

The significant findings in this paper as follows:

- A semi-centralized Blockchain-based digital healthcare network for the protection and sharing of patient data is introduced to ensure safe and effective data storage and data sharing.
- The rule-based clustering algorithm is used to group the diabetic and cardio disease patient records.
- After this clustering, diabetic and cardio disease is predicted using Feature selection based ANFIS.
- Finally, the model is created to evaluate the performance of the proposed work in terms of various metrics.

### D. PAPER ORGANIZATION

The remaining of the paper is organized as follows: The background of fog computing and blockchain explained in section II and Section III describes the reviews of the related work. Section IV explains the system and data model. The proposed methodology is defined in Section V. The experimental results are analyzed in Section VI and finally, Section VII, concludes the paper.

Fog computing is considered to be one of the key technologies that contribute greatly to promoting IoT healthcare and surveillance applications as these systems are latency-sensitive and real-time tracking, data processing, and decision making are critical criteria in healthcare applications such as servicing the elderly by home nursing, heart care, diabetes and some other diseases. Health data is an important topic because it includes essential, confidential knowledge. With fog computing, the aim that patients take care of their own health data locally is realized. Those safety data are housed in fog nodes such as smartphones or smart vehicles [13]. Fog computing provides tremendous advantages for fog-based application which is prone to delay. Hong *et al.*, [12] introduced Mobile Fog, which is a globally dispersed and latency-sensitive programming paradigm for Internet applications. A variety of studies has looked into the use of fog in health care.

Health care contributors are widely implicit in producing large volumes of information in a variety of formats, together with records, economic papers, clinical test findings, imaging tests, and vital sign assessments, etc., [10]. The comprehensive database created in care environments is expanding rapidly, with healthcare information struggling from numerous problems, with data access, and how information can be obtained beyond the healthcare ability. Blockchain provides the ability to enhance the data's authentication and legitimacy. It also helps to disseminate data inside the network or services. Such apps affect the cost, quality of data, and importance of providing health care within the system. Blockchain is a transparent, decentralized network without the middleman [14]. Blockchain healthcare networks do not need several verification rates which have access to data for anyone who is part of the infrastructure of blockchain. Data is rendered available to consumers and is transparent. Such innovations will continue to overcome the numerous problems facing the healthcare domain today.

Disease prediction is one of the main real-world problems in healthcare domain. Many classification algorithms [31], [33] are used to predicts the diseases accurately. Artificial neural network (ANN) is one of the classification algorithms. ANN is a massively computational parallel model with self-adaptive and self-learning capabilities, because of its large parallel structure; it takes more time to predict the outcome. ANN is not appropriate for dealing with such issues, such as ambiguous and imprecise data for which problems of uncertainty may occur at any point of the process of classification.

Fuzzy logic is used to resolve this issue in order to translate the numeric input features into their corresponding linguistic terminology. Based on linguistic properties such as low, medium and high, each input function is transformed into its corresponding membership values in this fuzzification process. Similarly, from the input features, all linguistic characteristics are extracted. By deciding the membership value in different linguistic terms, fuzzy logic is also sufficient to deal with the ambiguity problem. Adaptive Neuro Fuzzy Inference System(ANFIS) is a hybrid model which adopts the characteristics of ANN and fuzzy logic.

The objective of this paper is to develop the disease prediction model using feature selection and ANFIS. Feature selection is the one of the pre-processing technique which reduces the size of the dimensionality of the dataset. This paper use Cronbach's alpha [41] for optimal feature selection.

The significant findings in this paper as follows:

- A semi-centralized Blockchain-based digital healthcare network for the protection and sharing of patient data is introduced to ensure safe and effective data storage and data sharing.
- The rule-based clustering algorithm is used to group the diabetic and cardio disease patient records.
- After this clustering, diabetic and cardio disease is predicted using Feature selection based ANFIS.
- Finally, the model is created to evaluate the performance of the proposed work in terms of various metrics.

The remaining of the paper is organized as follows: The background of fog computing and blockchain explained in section II and Section III describes the reviews of the related work. Section IV explains the system and data model. The proposed methodology is defined in Section V. The experimental results are analyzed in Section VI and finally, Section VII, concludes the paper.

## II. BACKGROUND

### A. FOG COMPUTING

It is a distributed computing framework that expands the network's cloud infrastructure to the edge. It supports the operation and configuration of data center and end-user processing, networking, and storage facilities. Fog computing generally comprises specifications of the software that operates between sensors and the cloud, i.e., smart access points, routers or advanced fog devices, in both the cloud and edge applications. Fog computing embraces agility, computational power, networking protocols, the flexibility of the interface, cloud convergence, and disseminated data analytics to meet requirements of applications requiring short latency with large and compact geographic delivery [3].

Cisco initially coined the word fog computing [6]. Open Fog Consortium [7] describes fog computing as: 'a horizontal system-level architecture which distributes computation, storing, controlling and networking tools and services everywhere in the Cloud to Things spectrum.' The author in [8] defined as, "A situation in which a vast amount of heterogeneous, omnipresent and autonomous computers interacts and theoretically collaborate and with the network to execute storage and processing activities without third-party intervention. These activities may be to support simple network operations or new technologies and applications operating in a sandboxed environment".

The structure of fog computing is shown in Fig. 1. The cloud layer, which is the cornerstone of fog computing, conducts data virtualization, analysis, deep learning, and in the proxies of the fog layer updates laws and patterns. The proxy server acts as a web service and is more manageable. A centralized data collection enables creditworthiness and convenient data access through storing power within a cloud. A data store situated in the center of the fog computing system can be reached from both the computer layer and the fog layer [4].

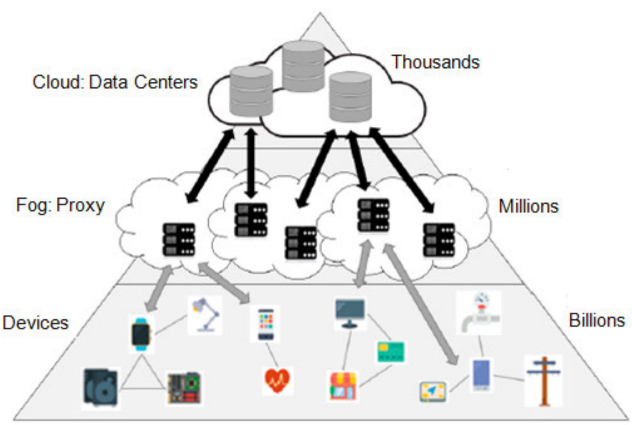


**FIGURE 1.** Fog computing structure [4].

Fog computing's characteristics include location recognition and low latency, spatial reach, scalability, accessibility support, real-time communications, convergence, interoperability, web analytics support, and cloud interplay. Reduced network load, automatic connectivity assistance, context awareness, no single fault point, improved market resilience, low latency, local and large-scale delivery, reduced running costs, versatility and heterogeneity are the benefits of fog computing [5].

The Fog computing network has a wide range of applications. Fig 2 shows the applications supported by fog computing.

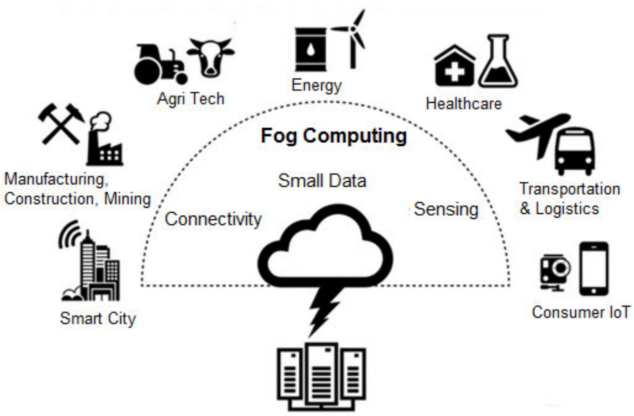


**FIGURE 2.** Fog computing applications.

In this paper, fog computing is used in the healthcare domain. Fog computing is a vital aspect of healthcare. It provides responses that are crucial in healthcare monitoring and incidents in real-time. Furthermore, the integration with a vast number with healthcare systems for remote collection, distribution, and cloud retrieval of medical data involves a secure network link that is not accessible.

### B. BLOCKCHAIN

Blockchain is one of the most innovative technologies and a digital wallet which retains track of transactions and events occurring across the network, and whose integrity is ensured via a peer-to-peer computing network, not by any centralized entity that might eliminate the risk of a single central point. It is composed of structured documents organized in a block structure that includes transaction batches and previous key hash. Every block is chronologically linked, and the data on the Blockchain network is unchallengeable [9].

Any users have individual access rights in a blockchain network to allow transactions that are modified throughout the framework, known as consensus protocol [10]. For inserting transactions, a blockchain uses SHA256 hash. The NSA creates that, which is 64 characters large. All transactions are registered in a blockchain network though not modifying or manipulating the public ledger; Both transfers are distributed to various users across the network to transfer and update the data; a blockchain network may be duplicated to a separate venue, for example, within the same ability or healthcare distribution network, or as part of a regional or global data exchange system.

The Blockchain's data structure is a hierarchical set of blocks shown in fig 3. Blocks are linked in the form of a tuple, while the current block stores such values as previous block hash, previous block Blockchain address etc. in its header. Every block is composed of two components: header and body. The header contains block number, previous block hash value to preserve chain reliability, current block body hash to protect transaction data integrity, timestamp, nonce, blockchain block creator address and other requested detail. Block bodies contain one or more transactions.

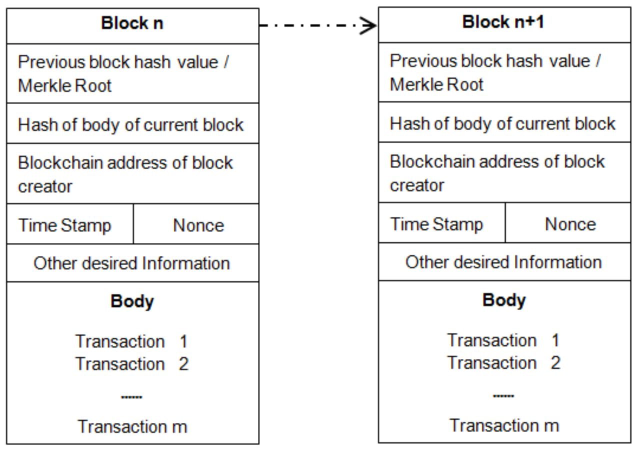


**FIGURE 3.** Block structure.

Decentralization, Durability, Transparency and Auditability are primary aspects of Blockchain. Public, private and consortium are kinds of Blockchain [11]. All archives are available to the public in the public Blockchain so that anyone may engage in the consensus process. Despite this, the consensus mechanism of a cooperative network will require only a collection of pre-selected nodes. As for private Blockchain, only those nodes originating from a single entity will be permitted to join the consensus process.

## III. RELATED WORK

### A. FOG COMPUTING IN HEALTHCARE

Health Fog framework is proposed in [15]. Fog computing is used as an intermediate layer among the cloud and end-users. Authors primarily focused on developing and addressing data protection problems in healthcare systems in a scalable way. Cloud access authentication agent is combined with Health Fog to enhance the security of the network. Besides, cryptographic features also specified to enhance Health Fog efficacy. The remote control of the patient's healthcare in smart homes is introduced in [16] based on the principle of fog computing at the intelligent access point. For handle the patient's real-time data at the fog layer, an event-based approach is adopted for initiating data transmission. The theory of immediate mining is used to evaluate incident difficulties by calculating the index of the temporal health of the victim.

Gia *et al.* [17] improve the health management program by leveraging the idea of fog computing at smart gateways offering specialized technologies and facilities such as distributed data processing, centralized storage and network-side monitoring. The author selects the Electrocardiogram (ECG) feature extraction as a case study. The ECG signals are analyzed with extracted features in smart gateways. Negash *et al.* [18] focus on developing an intelligent e-health interface being used in the Fog computing layer, linking a network of these gateways, both for home use and hospital use. Gateway technologies are addressed and tested when applying fog.

The idea of Fog Computing in Healthcare IoT systems is proposed in [19] via the creation of a Geo dispersed intermediate layer of information among sensor nodes and the cloud. A concept for the implementation of an intelligent e-health interface is being introduced. An IoT-based premature caution score safety screening is introduced to demonstrate the system's efficacy in a health case study. In [20], the authors propose a hierarchical computing model supported by fog for remote IoT-based patient management systems. The distributed computing system allows for the partitioning and distributing of analytics and decision-making among the fog and the cloud.

In a healthcare context, Alazeb and Panda [21] presented two separate frameworks for using fog computing. The two models are heterogeneous and homogeneous data from fog modules. They suggest a unique approach for each model to assess the harm done by malicious transactions so that actual data may be retrieved, and transactions marked for potential inquiry can be impacted. In [22], a novel architecture called Health Fog is proposed to incorporate deep learning ensemble into Edge computing devices and implemented it for real-life implementation of automated cardio disease detection. It offers healthcare as a fog service using IoT devices, handles heart patient information effectively, and comes as app requests.

### B. BLOCKCHAIN IN HEALTHCARE

Healthcare information-sharing network based on Blockchain is proposed in [23]. The author uses two liberally-coupled Blockchain to manage various forms of healthcare information and also incorporates off-chain storage and on-chain authentication to meet safety and authenticity criteria. Liang *et al.*, [24] suggest a revolutionary user-centric health data exchange approach through the use of a decentralized and approved Blockchain for guarding confidentiality using the channel creation method and improve individuality protection via the blockchain-based relationship program. Evidence of validity and authentication is indefinitely recoverable from the cloud database and embedded in the blockchain network to protect the confidentiality of health records inside each document.

A secure and privacy-conserving blockchain-based PHI networking scheme was proposed in [25] for improving diagnosis in e-Health scheme. Private and consortium Blockchain is developed through the creation of their information structures and consensus mechanisms. The private ledger manages the PHI while the ledger community keeps a database of the robust indexes of the PHI.

Griggs *et al.*, [26] propose smart, blockchain-based contracts to enable secure medical sensor research and management. The author built a network based on the Ethereum protocol using a private blockchain where the sensors connect with a mobile computer that calls smart agreement and mark logs of every activity on the Blockchain. In [27], a blockchain-based system is introduced for safe, interoperable, and proficient access by patients, clinicians, and third parties to medical data while maintaining the confidentiality of personal details of patients. Through an Ethereum-based blockchain, it makes use of smart agreement to boost access control and code obfuscation, using advanced cryptographic methods for enhanced protection.

In [28], a novel framework for the storage of medical data based on Blockchain was introduced. Users should retain valuable data in perpetuity, so where interference is alleged, the originality of the data may be checked. The author makes use of wise data management techniques and a number of cryptographic methods to protect user confidentiality. MedBlock, a blockchain-based information management program, was introduced in [29] for managing information from patients. The centralized MedBlock database in this system allows or secure entry and storage of medical information. The improved consensus process creates consensus on medical history without significant energy consumption and network congestion.

### C. DISEASE PREDICTION

A novel Optimistic Unlabeled learning strategy was introduced in [30], based on clustering and 1-class classification method. This method initially clusters positive data, studies 1-class classifier models using clusters, selects negative data intersection as the Stable Negative set, and finally uses binary SVM (Support Vector Machine) classification algorithm. In [31], a scheme called ensemble classification was investigated, which is employed by combining multiple classifiers to improve the precision of weak algorithms. The author applies the algorithm for a medical dataset, demonstrating its early utility in forecasting disease.

In [32], an appropriate segmentation and classification method is presented to discern the progression of Alzheimer's disease, moderate neurological dysfunction, and common objects of control correctly. A fusion segmentation method is invented to perform segmentation using K-means clustering and graph-cutting schemes. Depending on their characteristics, the clustered regions are given labels for the classification analysis. Nilashi *et al.* [33] are developing a new knowledge-based prediction method for diseases using clustering, noise reduction, and simulation methods. Classification and Regression Trees algorithm is used to produce the knowledge-based system's fuzzy rules.

An updated variant of K-Means based on density was introduced in [34], which provides an innovative and logical approach for choosing the early centroids. The algorithm's main concept is to pick data points that belong to dense regions and which are appropriately segregated as the initial centroids in feature space. This approach makes comparatively improved estimates of subtypes of cancer from evidence regarding gene expression. A classification algorithm for managing imbalanced datasets was introduced in [35] based on the principle of information granulation (IG). This algorithm assembles data from majority classes into granules to balance the class ratio inside the data. This algorithm first produces a collection of IGs using meta- heuristic methods and applies the data classification algorithm.

An edge-cloud-based healthcare infrastructure is proposed in [46] for real-time disease detection, monitoring, and recovery. This approach does not consider the blockchain concept. The proposed method uses blockchain for securing patient health record.

## IV. SYSTEM MODEL

This section explains the proposed system model and notations used in this model. In this model, the IoT medical sensors are used to collect, patient health related data. The fog nodes collect these data and send to medical analyzer for disease analysis and prediction. Fig 4 shows the system model. It contains five entities.

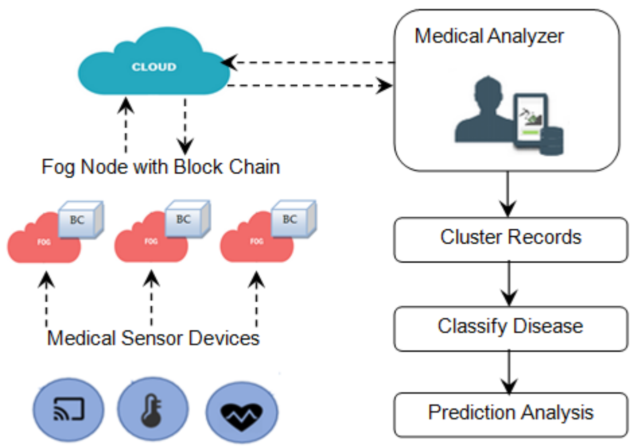


**FIGURE 4. Proposed system model.**

### A. MEDICAL SENSOR DEVICES

Sensor devices can track human health parameters of various sorts, whether wearable systems or embedded devices. Due to their restricted computing and storage capacities, these devices collect different types of health-related data and send data that will be well managed to fog nodes.

### B. FOG NODE

It is a simple platform for fog computing, which can be a network computer that manages underlying machines using processing resources, dedicated servers, or computational servers. It collects the data from the medical sensor devices and stores into a distributed ledger called Blockchain.

### C. BLOCKCHAIN

It is a cooperative network used to monitor patient health data and activity data status. No-one can access the network without authorization. This is composed of a sequence of blocks containing the previous hash block, status user health.

### D. CLOUD

It is used for storage purposes. It stores encrypted patient health information, and the authenticated medical analyzer can access these encrypted data for further process.

### E. MEDICAL ANALYZER

An authorized person who can access patient health information. The analyzer can group the information into two: normal patient and affected patient. The analyzer can also predict whether the patient contains diabetic or cardio diseases.

Table 1 show the notations used in clustering and classification process.

**TABLE 1. Notations.**

| Notation | Description |
|---|---|
| D | Dataset |
| $F_i$ | $i^{th}$feature in D |
| $DR_i$ | $i^{th}$data record in D |
| $A_{ij}$ | Attribute Value of $i^{th}$data record and $j^{th}$feature |
| RS | Rule Set |
| $R_i$ | $i^{th}$rule in RS |
| $Freq_{<R,C>}$ | Frequent Rule Set (R=rule, C=Count) |
| $R_{thr}$ | Rule Threshold |
| L+R=C | Left, Right and Class part (Rule) |
| $cand_+$ | Positive candidate rules |
| $cand_-$ | Negative candidate rules |
| Cls+ | Positive clustered data |
| Cls- | Negative Clustered data |
| $C\alpha$ | Cronbach's alpha |

## V. PROPOSED METHODOLOGY

This section explains the proposed Blockchain-based healthcare disease prediction with clustering and classification.

### A. BLOCKCHAIN STORAGE

In the medical domain, control of access, validity, data confidentiality and integration are essential to protecting the identity of the patient and sharing data within the healthcare environment with other organizations. The traditional way to achieve control of access usually implies confidence among the data owner and the entities that store them. Such agencies are also entirely assigned servers for identifying and implementing policies on access management. Interoperability is the capability of dissimilar information systems, software or frameworks to link data between stakeholders in a synchronized way, within and across organizational borders, to improve individual safety. The provenance of data relates to the historical record of the data and its sources, e.g., provenance in health domain data may be to provide auditability and consistency in the health record and to attain trust in the electronic health record software framework. Data integrity is the concept of data validity that concerns with the consistency required of the information. That ensures the level to which the intended data quality is achieved or surpassed decides the validity of the report [36]. Blockchain technology has several enticing features that can be used to enhance and gain a higher degree of integration, sharing of knowledge, access security, validity, and data transparency between the stakeholders listed, while trying to move towards a novel trust-building and sustaining infrastructure.

Blockchain can be described as a blockchain, capable of storing stable and permanent transactions between parties. Each block contains many elements including, user submitted valid transactions, time-stamped batches, and the previous block hash. A hash function is a function that transforms the data, it is given into a fixed-length irregular form. The timestamp reveals there must have been data at the time. The previous block hash ties the blocks together and forbids modification of any block or addition of a block between two different blocks. Blockchains allow auditing and traceability by connecting a new block to the previous one by using the latter's hash, and thereby creating a blockchain. The block transactions are generated in a Merkle tree (Fig 5) where the known root can be verified for each value of the leaf (transaction). Any non-leaf node in the Merkle tree is the hash of the values of its infant nodes. Searching for a transaction becomes really quick through using Merkle tree. Instead of checking the transactions linearly, the Merkle tree will determine more quickly whether a transaction is found in the block or not.

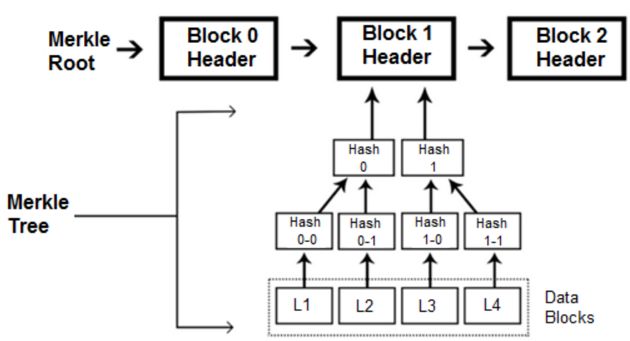


**FIGURE 5.** Merkle tree structure in blockchain.

This paper considers the consortium type of Blockchain, also known as semi-decentralized Blockchain. A consortium blockchain is not provided as a private blockchain to a single entity; it is conferred on a group of approved entities instead. Additionally, the blockchain consortium is a group of predefined nodes on the network. Consortium blockchain, therefore, provides security, inherited from public Blockchain. This gives a significant degree across the network. Consortium blockchains are most commonly associated with commercial use, as a consortium of the company's works together to use blockchain technologies to boost businesses. However, this kind of Blockchain may enable specific group members to access or adopt a hybrid method of access. The root hash and its Application Program Interface (API) may be publicly accessible. External entities can, therefore, use the API to conduct several inquiries and to obtain specific information relating to the blockchain status. Table 2 shows some properties [37] of consortium blockchain.

**TABLE 2.** Consortium blockchain property.

| Property | Value / Description |
|---|---|
| Consensus | Handled by set of nodes |
| Transaction Validation | Set of Authorized nodes |
| Transaction Reading | Any node or set of predefined node |
| Data Immutability | Yes |
| Transaction Throughput | High |
| Network Scalability | Low to Medium |
| Infrastructure | Decentralized |
| Features | ○ Applicable to tightly controlled business<br>○ Fee-free transaction<br>○ Laws on services are easier to manage<br>○ Effective defense from outside perturbations |
| Example | Hyperledger, Ethermint, Tendermint |

The authorized medical analyzer collects patient information and predicts whether the patient contains diabetic or heart-related diseases.

### B. DISEASE CLUSTERING

Clustering is one of the unsupervised techniques in data mining that deal with identifying groups inside a collection in unlabeled data. It is used to partition a set of data into different clusters, such that objects in the same group cluster are strongly related and distinct from objects in another cluster. Clustering technology has been widely accepted in many technologies such as pattern detection, image processing and pattern analysis of consumer transactions. It is essential during data analysis discovery and assessment, where researchers seek to find fundamental features that appear without previous knowledge of the data. However, the selection of appropriate clustering techniques and algorithms is determined by an interpretation of the data structure, the form of analysis to be carried out and the scale of the dataset.

Cluster classification in the medical domain provides a standardized, formalized approach for data discovery and identifying clinically related groupings. Efficient clustering methods are raising competition for costly health care services. It helps doctors deal with the influx of knowledge, and can assist with better facilities in strategic planning. The findings of the clustering are used to research patient independence or association and for more in-depth insight into evidence from medical surveys. All these advantages inspired the researcher to construct clustering models for grouping medical data.

Health data clustering raises a variety of new problems.

- Information overload – Developments in medical technology combined with high processing capacities are increasing the volume of data generated and processed in the healthcare sector. Discovery of knowledge and the retrieval of information from these large databases are difficult and prohibitively costly.
- Too many risk indicators are essential for decision-making and are heterogeneous.
- High consumer knowledge of medical treatment and improved life expectancy creates a rising demand for better health services. Yet misdiagnosis and imprecise care strategies arise with overworked and inexperienced doctors, challenging working environments etc.
- Choosing a suitable clustering approach and an adequate number of clusters in health care data can be challenging and often complicated.

To address this challenge, a novel rule-based clustering algorithm is proposed for the efficient cluster. This is a two-stage algorithm: in the first stage, the rules are generated based on patient information, and in the second stage, the clusters are generated based on the rules.

The pseudo-code of the rule generation algorithm has been given as follows.

This algorithm is suitable for a numerical data set. Initially, the numerical value is converted into discrete value (Low, Medium, and High) (steps 2- 12). Based on these values, the candidate rules (13-19) are generated for further process. This paper use frequency and threshold based rule generation. Based on the requirements, the candidate rules are extracted.

Consider the 15 patients fasting blood sugar level, 120, 90, 70, 45, 100, 130, 50, 35, 138, 82, 90, 50, 120, 58, 140. Table 3 shows the example.

Convert all the features values in the dataset. Count the frequencies of each record. If the record frequency is more than the $R_{thr}$ (initially set 5 – 10 depending on the requirements), then consider the record as candidate rule. The next stage is clustering. The pseudo-code of the clustering algorithm has been given as follows.

The candidate rules are divided into three parts ($L + R = C$), i.e. left, right and a class variable. Based on the C (class variable), $cand_{+}$ and $cand_{-}$ rules are generated. Positive and negative clusters are formed based on these candidate rules if any record not matched with candidate rules then it will be considered as an outlier record.

**Algorithm 1** Rule Generation

Input: D
Output: RS
1: $RS = \emptyset$
2: for each $F_i \in$ Feature do
3: $distF_i$ = get distinct value($F_i$)
4: Sort($distF_i$)
5: Group $distF_i$ values into Low, Medium and High
6: end for
7: for each $DR_i \in$ DataRecord do
8: for each $F_j \in$ Feature do
9: $newA_{ij}$ = convert $A_{ij}$ into Low, Medium, High based on Step 5
10: end for
11: end for
12: generate newDR based on $newA_{ij}$
13: $Freq_{<R,C>}$ = Find and Count Similar Records
14: candidate = $Freq_{<R,C>} \forall c > R_{thr}$
15: If candidate $\neq \emptyset$
16: RS = candidate
17: else
18: $RS = \emptyset$
19: end if

**TABLE 3.** Data conversion example.

| Steps | Value |
|---|---|
| Input Feature | 120, 90, 70, 45, 100, 130, 50, 35, 138, 82, 90, 50, 120, 58, 140 |
| Distinct Value | 120, 90, 70, 45, 100, 130, 50, 35, 138, 82, 58, 140 |
| Sort Value | 35, 45, 50, 58, 70, 82, 90, 100, 120, 130, 138, 140 |
| Group Values | (35, 45, 50, 58) = Low<br>(70, 82, 90, 100) = Medium<br>(120, 130, 138, 140) = High |
| Convert Feature | High, Medium, Medium, Low, Medium, High, Low, Low, High, Medium, Medium, Low, High, Low, High |

### C. DISEASE PREDICTION

Processing of medical data is a critical topic that needs to be accurate for disease prevention, diagnosis and processing. Maintaining health records has been a pivotal scientific mission. Patient data comprising of specific disease-related characteristics and symptoms will be reached with special caution to ensure professional treatment. Because the information stored in medical repository can include incomplete and redundant information, that medical data is inefficient [38]. Until implementing data mining algorithms, it is essential to contain effective data planning and reduction because this can impact the mining performance. Disease diagnosis is quicker and easier if the data is accurate, reliable and noise-free.

Selecting a feature is an effective pre-processing method in data mining designed to reduce data dimensionality.

**Algorithm 2** Clustering

Input: D, RS
Output: $Cls^{+}$, $Cls^{-}$
$cand_{+} = \emptyset$, $cand_{-} = \emptyset$
for each $R_i \in RS$ do
    Split $R_i$ into three parts ($L + R = C$)
    $cand_{+} = L + R = C$ (rule with positive patients)
    $cand_{-} = L + R = C$ (rule with negative patients)
end for
for each rec $\in$ newDR
    if (rec match with $cand_{+}$) then
        Cls +.add(rec)
    else (rec match with cand-) then
        Cls-.add(rec)
    else
        Out.add(rec)
    end if
end for

Identifying the most severe disease-related risk factors is very important in medical diagnosis. Specific recognition of features helps delete unwanted, unnecessary features from the dataset of the disease, resulting in a simple and improved outcome. Classification and prediction is a technique of data mining that initially utilize training data to create a training model and then applies the resulting model to test data to achieve predictive results. Diverse recognition systems have been applied to disease data sets for diabetes and cardiovascular disease treatment. This paper proposes a Feature Selection and use Adaptive Neuro-Fuzzy Inference System [39], which adopts the characteristic of ANN and Fuzzy Logic for disease prediction. Fig 6 shows the prediction model workflow.

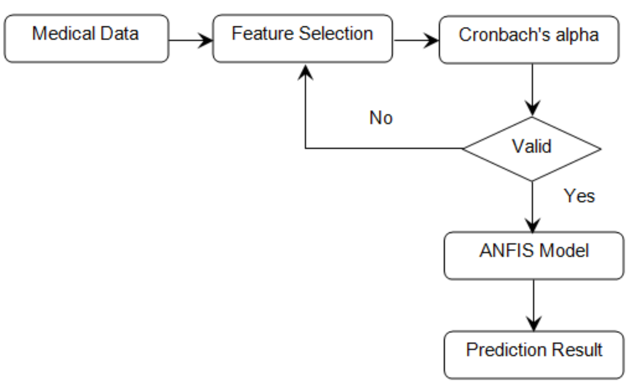


**FIGURE 6.** Prediction work flow.

Feature selection is a commonly used data pre-processing method in data mining that is essentially used to reduce data by removing irrelevant and redundant features from the dataset [40]. In addition, this method increases data interpretation, improves information analysis, decreases learning algorithm training times and increases prediction efficiency.

To collect more useful knowledge, different feature collection methods have been applied to the healthcare datasets. The use of feature selection methods is performed on clinical databases to predict various diseases. Different learning algorithms operate effectively and provide more reliable outcomes if there are more important and non-redundant attributes in the details. Given the vast number of redundant and unnecessary features in the medical datasets, an effective feature extraction strategy is required to mine fascinating attributes specific to the disease.

This paper proposes an optimal feature selection algorithm which uses Cronbach's alpha [41]. The Cronbach alpha measures the consistency of features in a test, i.e. the test's internal consistency. It can be measured by,

$$\boldsymbol{C\alpha} = \frac{|\boldsymbol{F}| \cdot \boldsymbol{CV_{avg}}}{\boldsymbol{V_{avg}} + (|\boldsymbol{F}| - \mathbf{1}) \cdot \boldsymbol{CV_{avg}}} \tag{1}$$

Where $|F|$ = number of features, $CV_{avg}$ = average of covariance, $V_{avg}$ = average variance.

The pseudo-code of the feature selection algorithm has beengiven as follows.

**Algorithm 3** Feature Selection

Input: D
Output: SF (Selected Features)
1: pc =10, global_C $\alpha = 0$, maxIter = 100
2: for i = 1 to pc do
3: $pop_i^j = Random\{0, 1\}$, $j \in F_j$
4:     $C\alpha_i = 0$
5: end for
6: for iter = 1 to maxIter do
7:     for i = 1 to pc
8:         compute Cronbach's alpha (Ca) using (1)
9:         if ($Ca > C\alpha_i$) then
10:             $C\alpha_i = Ca$
11:         end if
12:     end for
13:     maxCa = max($C\alpha_i$)
14:     if (maxCa > global_ $C\alpha$) then
15:         global_$C\alpha$ = maxCa
16:         SF = $pop_i$(index of maxCa)
17:     end if
18:     Replace the pop which contain lowest Ca
19: end for

Randomly generate the population using the random function and assign alpha as zero (steps 2 – 5). An iterative process is used to select optimal features (6-19). The maximum iteration is set as 100. Compute Cronbach's alpha (using (1)) for each randomly generated population. Select the maximum alpha value (step 13) and population if it is more than global alpha then set selected features as population (step 16). Change the population, which contains the lowest alpha (step 17) repeat steps (6-19) until maximum iteration reached. The selected features are used in the ANFIS model to predict the disease.

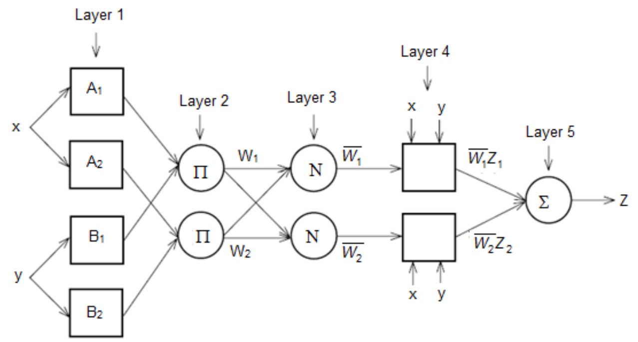


**FIGURE 7.** ANFIS architecture.

The ANFIS network is a neuro-fuzzy network developed by Jang in 1993 [42]. Because of ANFIS 'adaptive property, some nodes obtain the same property, and after that, the output comes based on the constraints that belong to those nodes. For efficient optimization, two learning methods are used to adjust constraints. Of convenience, the above-suggested method should have 2-inputs and 1-output, and its rule base includes two fuzzy if-then TSK [43] fuzzy model rules. This TSK model generates fuzzy rules from the dataset input-output. If x = A and y = B, z = f(x, y). Here, f(x, y) = flat function that typically denotes a polynomial.

The ANFIS architecture is depicted in fig 7. The function of each layer is defined below.

*Layer 1:* This layer is the membership layer which contains adaptive nodes with node functions defined as

$$L_i^1 = \mu_{Ai}(x)\ (i = 1, 2) \quad (2)$$

$$L_i^1 = \mu_{B(i-2)}(y)\ (i = 3, 4) \quad (3)$$

where $x$ and $y$ denote input nodes, $A$ and $B$ are linguistic labels, $\mu(x)$, and $\mu(y)$ refer to membership functions.

*Layer 2:* This layer adopts the 'set node' property and each node is labeled with a ring symbol and named with multiplying the node function to act as output through input. Consider

$$L_i^2 = \omega_i = \mu_{Ai}(x)\,\mu_{Bi}(x)\ (i = 1, 2) \quad (4)$$

The output $\omega_i$ represents the rules firing strength.

*Layer 3:* Each node in this layer is labeled with a ring symbol and called N, with the node function to regulates the firing force by measuring the proportion of the firing force of the ith node to the sum of the firing power of all laws. In fact,

$$L_i^3 = \bar{\omega}_i = \frac{\omega_i}{\sum \omega_i} = \frac{\omega_1}{\omega_1 + \omega_2}, \quad (i = 1, 2) \quad (5)$$

The outputs of that layer are called to as standardized firing ability for ease.

*Layer 4:* In this layer, each node is in nature, flexible, and is noticeable with a square. Node role is specified by

$$L_i^4 = \bar{\omega}_i \cdot f_i = \bar{\omega}_i\,(p_i x + q_i y + r_i), \quad (i = 1, 2) \quad (6)$$

where $\bar{\omega}$ is the output of layer 3 and $\{p_i, q_i, r_i\}$ is the set of parameters.

*Layer 5:* Each node within this layer is a constant node, and the overall result can be expressed as a linear mixture of the following parameters. Two parameter sets can be modified, $\{a_i, b_i, c_i\}$ marked as parameters of the assumption and $\{p_i, q_i, r_i\}$ marked as the subsequent parameters. The training process must harmonize the two parameters that are set to predict successful outcomes.

## VI. EXPERIMENTAL RESULT

In this section, the performance of the proposed work was analyzed. The proposed work was implemented using Java (version 1.8), and the experiments are performed on an Intel(R) Pentium machine with a speed 2.13 GHz and 4.0 GB RAM using Windows 7 32-bit Operating System.

### A. DATA SET

The two dataset diabetes and heart disease data set is used for the experimental result. The diabetes data set contains 768 instances, with eight numeric features. Table 4 shows the data set information.

**TABLE 4.** Diabetes data set information.

| Feature Name | Description | Range |
|---|---|---|
| Pregnancies Count | Number of pregnancies | 0 – 17 |
| Glucose Level | Plasma glucose level | 44 - 199 |
| BP Level | Diastolic hypertension | 24 - 122 |
| Skin Thickness | The thickness of Triceps skin fold | 7 – 99 |
| Insulin | Insulin serum for 2-hours | 14 – 846 |
| BMI | Body mass index | 18.2 - 67.1 |
| Pedigree Function | A pedigree function of diabetes | 0.078 – 2.42 |
| Age | Age in Years | 21 – 81 |
| Class Label | The patient has diabetes or not | 0 or 1 |

The heart disease data set contains 800 instances, with six numeric features and eight categorical attributes. Table 5 shows the data set information.

### B. EVALUATION METRICS

This section explains the evaluation metrics for the experimental result.

#### 1) PURITY

This measure evaluates the clustering consistency. The pureness of the final clusters can be seen when opposed to the

**TABLE 5. Heart disease data set details.**

| Feature Name | Description | Range |
|---|---|---|
| Age | Age in years | 29 – 77 |
| Sex | Patient Gender | 0, 1 |
| CPT | Chest pain type | 1, 2, 3, 4 |
| Trest_bps | Resting BP | 94 – 200 |
| Chol | Cholesterol in Serum | 126 – 546 |
| FBS | Fasting Blood Sugar | 0, 1 |
| RestECG | Resting Electrocardiographic | 0, 1, 2 |
| Thalach | Maximum Heart rate achieved | 71 – 202 |
| Exang | Exercise-Induced Angina | 0, 1 |
| OldPeak | ST depression induced by exercise relative to rest | 0 – 6.2 |
| Slope | Slope of the peak exercise | 1, 2, 3 |
| CA | No of major vessels | 0, 1, 2, 3 |
| Thal | Defect value | 3, 6, 7 |
| Class Label | Patient have heart disease or not | 0 or 1 |

ground truth groups. It can be calculated as,

$$Purity = \frac{\sum_{i=1}^{|C|} \frac{n_i^d}{n_i}}{|C|} \tag{7}$$

where |C| is the total number of clusters, $n_i^d$ is the number of instances with the leading class label in Cluster $C_i$ and $n_i$ indicates the number of the instances in the cluster $C_i$

### 2) NMI (NORMALIZED MUTUAL INFORMATION)

It measures the mutual experience, followed by a normalization process, between the resulting cluster labels and ground truth labels. It can be calculated as

$$NMI = \frac{\sum_{i,j} n_{ij} \log \frac{n*n_{ij}}{n_i*n_j}}{\sqrt{(\sum_i n_i + \log \frac{n_i}{n})(\sum_j n_j + \log \frac{n_j}{n})}} \tag{8}$$

where $n_{ij}$ is the number of instances belonging to the class i found in the cluster j and $n_i(n_j)$ is the number of instances in the cluster i (j)

### 3) ACCURACY

Overall prediction result

$$Acc = \frac{TP + TN}{TP + FP + TN + FN} \tag{9}$$

Where TP = true positive i.e. properly predicted disease as normal. FP = false positive i.e. wrongly predicted disease as affected TN = true negative i.e. properly predicted disease as affected. FN = false negative i.e. wrongly predicted disease as normal.

## C. EXECUTION TIME COMPARISON

This section compares the execution time of blockchain hash generation, rule generation and cluster formation for diabetic and heart disease data.

Fig. 8 shows the blockchain hash generation for diabetic and heart disease data set.

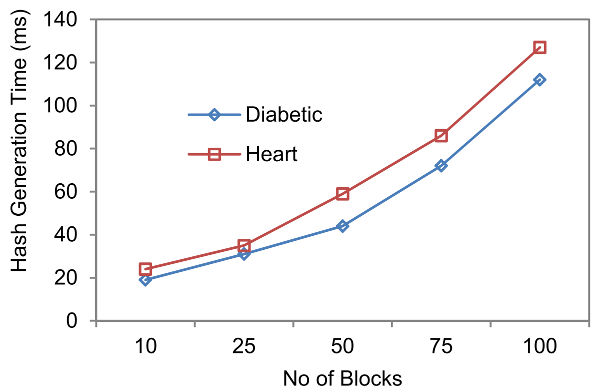


**FIGURE 8. Blockchain hash generation time.**

Fig. 9 shows the transaction creation time for two data set. It is the time taken to create a transaction for a given block. This paper use blockchain for secure storage purpose. The other parameters of the blockchain (latency, throughput and bandwidth) are out of scope.

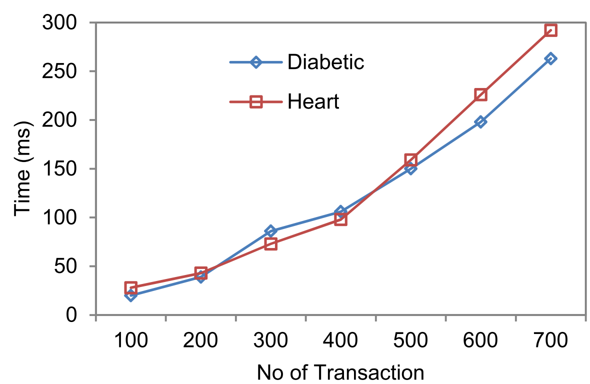


**FIGURE 9. Transaction creation time.**

Fig. 10 shows the execution time for rule generation and cluster formation for diabetic and heart disease data set. For two data sets, the cluster formation time is less than compared to the rule generation. The rule generation takes more time because it converts all the original data set into low, medium, high value to generate the candidate rules.

Fig. 11 shows the running time for the feature selection process. When increasing the number of iterations, the running time also increases. The proposed feature selection algorithm is compared with binary cuckoo search (BCS) [45]

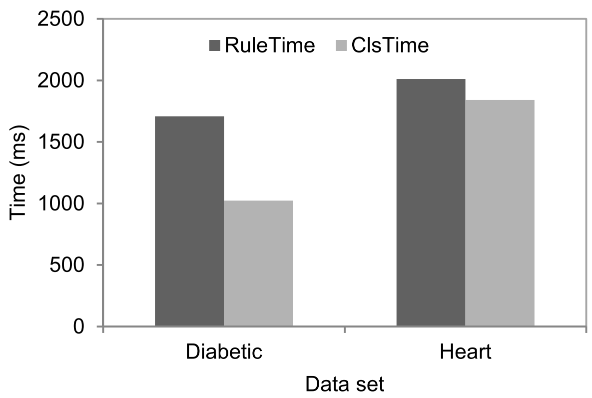


**FIGURE 10.** Execution time for rule and cluster formation.

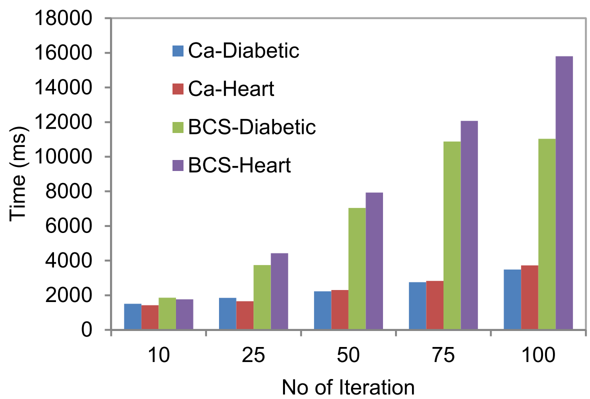


**FIGURE 11.** Feature selection running time.

algorithm. The BCS algorithm takes more execution time for feature selection.

### *D. CLUSTERING RESULT*

This section explains the rule-based clustering performance result.

Fig. 12 and 13 show the rule count for diabetic and heart data. The rules are increased when the number of instances

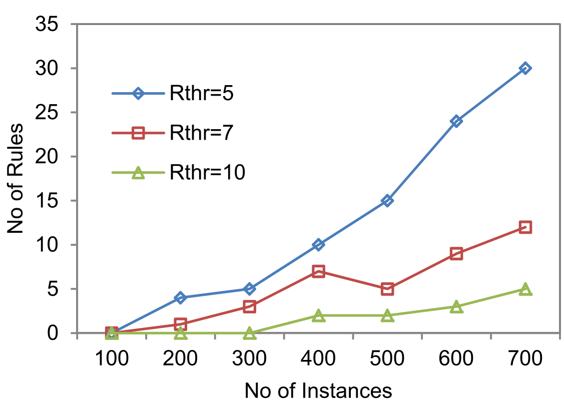


**FIGURE 12.** Instances vs rules for diabetic data.

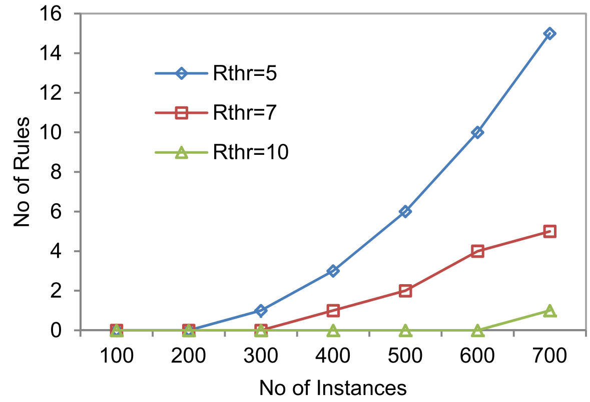


**FIGURE 13.** Instances vs rules for heart data.

is increased. Three threshold values (5, 7, 9) are used for experiments. More rules are generated for the threshold value $R_{thr} = 5$ for both diabetic and heart data set.

Fig 14 shows the candidate rule count with positive and negative rules for diabetic and heart disease for Rthr = 5.

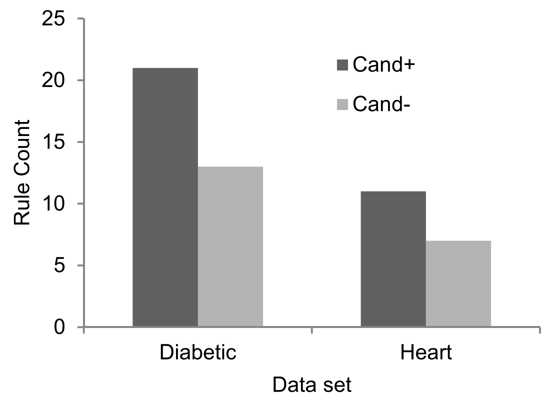


**FIGURE 14.** Candidate rule count Rthr = 5.

Fig 15 and 16 shows the purity and NMI result for diabetic and heart disease data set. For diabetic data set, the purity achieved 77%, and for heart disease 81%. The NMI value is more than 70% for both diabetic and heart disease data set when increasing the number of rules.

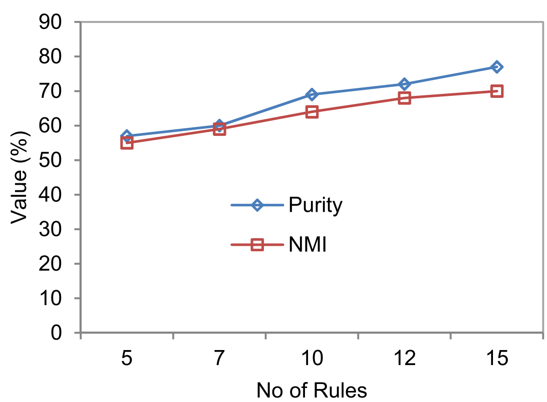


**FIGURE 15.** Purity and NMI for diabetic data.

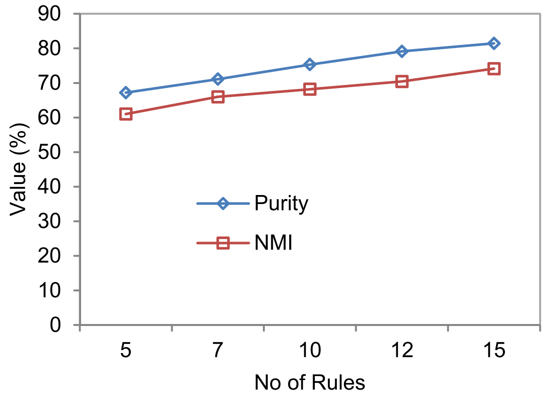


FIGURE 16. Purity and NMI for heart data.

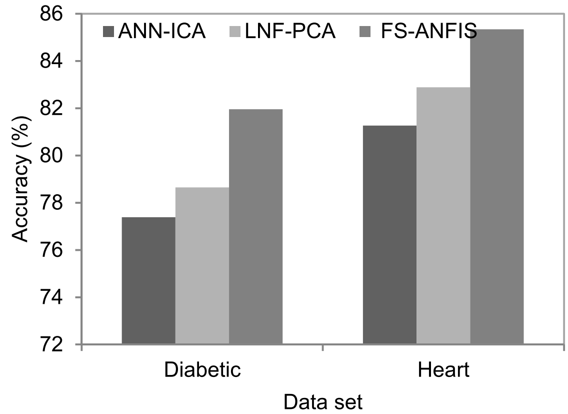


FIGURE 19. Accuracy comparison.

### E. PREDICTION RESULT

This section explains the FS-ANFIS prediction performance result.

Fig. 17 shows the Cronbach's alpha for a different population. The percentage of alpha value > 75 is acceptable consistency, and more than 90 is excellent consistency. Both the data set achieved good consistency.

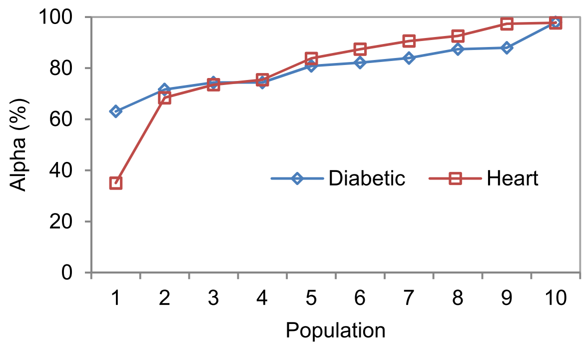


FIGURE 17. Alpha for different population.

Fig 18 shows the alpha value for 100 iterations.

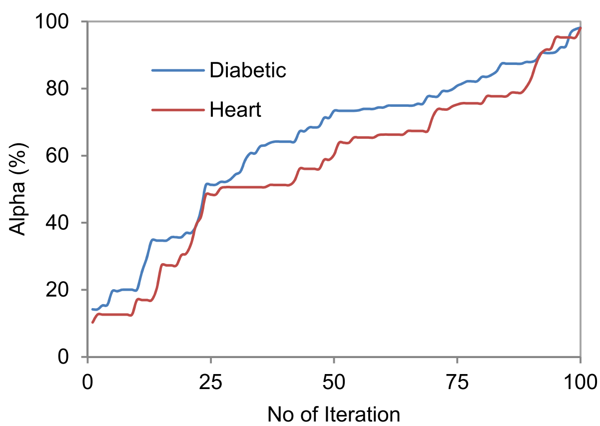


FIGURE 18. Alpha vs. no of iteration.

Fig 19 shows the accuracy comparison of 3 different algorithms. Compared to ANN-ICA (Integrated Component Analysis) and LNF-PCA [44], the proposed algorithm obtains higher accuracy.

## VII. CONCLUSION

In the current healthcare system, the use of Blockchain plays a crucial role. It can result in automated processes for collecting and verifying data, correcting and aggregating information from different resources that are indisputable, defiant to manipulation, and providing protected data, with condensed cybercrime chances and which also supports disseminated information, with system redundancy. This paper proposes efficient Blockchain-based secure healthcare services for disease prediction in fog computing. Diabetes and cardio diseases are considered for prediction. The proposed work efficiently clusters and predict the disease compared to other methods. In the future, the security and privacy for accessing patient medical data and some hybrid clustering and classification model can be added to enhance the performance of the prediction results.

## REFERENCES

[1] P. Sundaravadivel, E. Kougianos, S. P. Mohanty, and M. Ganapathiraju, "Everything you wanted to know about smart healthcare," *IEEE Consum. Electron. Mag.*, vol. 7, no. 1, pp. 18–28, Jan. 2018.

[2] M. A. Sayeed, S. P. Mohanty, E. Kougianos, and H. P. Zaveri, "Neuro-detect: A machine learning-based fast and accurate seizure detection system in the IoMT," *IEEE Trans. Consum. Electron.*, vol. 65, no. 3, pp. 359–368, Aug. 2019.

[3] A. V. Dastjerdi, H. Gupta, R. N. Calheiros, S. K. Ghosh, and R. Buyya, "Fog computing: Principles, architectures, and applications," 2016, *arXiv:1601.02752*. [Online]. Available: http://arxiv.org/abs/1601.02752

[4] H.-J. Cha, H.-K. Yang, and Y.-J. Song, "A study on the design of fog computing architecture using sensor networks," *Sensors*, vol. 18, no. 11, p. 3633, Oct. 2018.

[5] H. F. Atlam, R. J. Walters, and G. B. Wills, "Fog computing and the Internet of Things: A review," *Big Data Cogn. Comput.*, vol. 2, no. 10, pp. 1–18, Apr. 2018.

[6] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, "Fog computing and its role in the Internet of Things," in *Proc. 1st MCC Workshop Mobile Cloud Comput. - MCC*, Aug. 2012, pp. 13–15.

[7] *Definition of Fog Computing*. Accessed: Jan. 15, 2021. [Online]. Available: https://www.openfogconsortium.org/#definition-of-fogcomputing

[8] L. M. Vaquero and L. Rodero-Merino, "Finding your way in the fog: Towards a comprehensive definition of fog computing," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 44, no. 5, pp. 27–32, Oct. 2014.

[9] Z. Wang, N. Luo, and P. Zhou, "GuardHealth: Blockchain empowered secure data management and graph convolutional network enabled anomaly detection in smart healthcare," *J. Parallel Distrib. Comput.*, vol. 142, pp. 1–12, Aug. 2020.

[10] S. Tanwar, K. Parekh, and R. Evans, "Blockchain-based electronic healthcare record system for healthcare 4.0 applications," *J. Inf. Secur. Appl.*, vol. 50, Feb. 2020, Art. no. 102407.

[11] Z. Zheng, S. Xie, H. Dai, X. Chen, and H. Wang, "An overview of blockchain technology: Architecture, consensus, and future trends," in *Proc. IEEE Int. Congr. Big Data (BigData Congress)*, Jun. 2017, pp. 557–564.

[12] K. Hong, D. Lillethun, U. Ramachandran, B. Ottenwälder, and B. Koldehofe, "Mobile fog: A programming model for large-scale applications on the Internet of Things," in *Proc. 2nd ACM SIGCOMM Workshop Mobile Cloud Comput. - MCC*, 2013, pp. 15–20.

[13] S. Yi, Z. Hao, Z. Qin, and Q. Li, "Fog computing: Platform and applications," in *Proc. 3rd IEEE Workshop Hot Topics Web Syst. Technol. (HotWeb)*, Nov. 2015, pp. 73–78.

[14] N. Rifi, E. Rachkidi, N. Agoulmine, and N. C. Taher, "Towards using blockchain technology for eHealth data access management," in *Proc. 4th Int. Conf. Adv. Biomed. Eng. (ICABME)*, Oct. 2017, pp. 1–4.

[15] M. Ahmad, M. B. Amin, S. Hussain, B. H. Kang, T. Cheong, and S. Lee, "Health fog: A novel framework for health and wellness applications," *J. Supercomput.*, vol. 72, no. 10, pp. 3677–3695, Oct. 2016.

[16] P. Verma and S. K. Sood, "Fog assisted-IoT enabled patient health monitoring in smart homes," *IEEE Internet Things J.*, vol. 5, no. 3, pp. 1789–1796, Jun. 2018.

[17] T. N. Gia, M. Jiang, A.-M. Rahmani, T. Westerlund, P. Liljeberg, and H. Tenhunen, "Fog computing in healthcare Internet of Things: A case study on ECG feature extraction," in *Proc. IEEE Int. Conf. Comput. Inf. Technol.; Ubiquitous Comput. Commun.; Dependable, Autonomic Secure Comput.; Pervas. Intell. Comput.*, Oct. 2015, pp. 1–8.

[18] B. Negash, A. Anzanpour, I. Azimi, M. Jiang, T. Westerland, A. M. Rahmani, P. Liljeberg, and H. Tenhunen, "Leveraging fog computing for healthcare IoT," in *Fog computing in the Internet of Things Intelligence at the edge*. Cham, Switzerland: Springer, 2017, pp. 145–169.

[19] A. M. Rahmani, T. N. Gia, B. Negash, A. Anzanpour, I. Azimi, M. Jiang, and P. Liljeberg, "Exploiting smart e-health gateways at the edge of healthcare Internet-of-Things: A fog computing approach," *Future Gener. Comput. Syst.*, vol. 78, pp. 641–658, Jan. 2018.

[20] I. Azimi, A. Anzanpour, A. M. Rahmani, T. Pahikkala, M. Levorato, P. Liljeberg, and N. Dutt, "HiCH: Hierarchical fog-assisted computing architecture for healthcare IoT," *ACM Trans. Embedded Comput. Syst.*, vol. 16, no. 5s, pp. 1–20, Oct. 2017.

[21] A. Alazeb and B. Panda, "Ensuring data integrity in fog computing basedhealth-care systems," in *Proc. Int. Conf. Secur., Privacy Anonymity Comput., Commun. Storage*. Cham, Switzerland: Springer, 2019, pp. 63–77.

[22] S. Tuli, N. Basumatary, S. S. Gill, M. Kahani, R. C. Arya, G. S. Wander, and R. Buyya, "HealthFog: An ensemble deep learning based smart healthcare system for automatic diagnosis of heart diseases in integrated IoT and fog computing environments," *Future Gener. Comput. Syst.*, vol. 104, pp. 187–200, Mar. 2020.

[23] S. Jiang, J. Cao, H. Wu, Y. Yang, M. Ma, and J. He, "BlocHIE: A BLOCkchain-based platform for healthcare information exchange," in *Proc. IEEE Int. Conf. Smart Comput. (SMARTCOMP)*, Jun. 2018, pp. 49–56.

[24] X. Liang, J. Zhao, S. Shetty, J. Liu, and D. Li, "Integrating blockchain for data sharing and collaboration in mobile healthcare applications," in *Proc. IEEE 28th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Oct. 2017, pp. 1–5.

[25] A. Zhang and X. Lin, "Towards secure and privacy-preserving data sharing in e-Health systems via consortium blockchain," *J. Med. Syst.*, vol. 42, no. 8, pp. 1–18, Aug. 2018.

[26] K. N. Griggs, O. Ossipova, C. P. Kohlios, A. N. Baccarini, E. A. Howson, and T. Hayajneh, "Healthcare blockchain system using smart contracts for secure automated remote patient monitoring," *J. Med. Syst.*, vol. 42, no. 7, pp. 1–7, Jul. 2018.

[27] G. G. Dagher, J. Mohler, M. Milojkovic, and P. B. Marella, "Ancile: Privacy-preserving framework for access control and interoperability of electronic health records using blockchain technology," *Sustain. Cities Soc.*, vol. 39, pp. 283–297, May 2018.

[28] H. Li, L. Zhu, M. Shen, F. Gao, X. Tao, and S. Liu, "Blockchain-based data preservation system for medical data," *J. Med. Syst.*, vol. 42, no. 8, pp. 1–13, Aug. 2018.

[29] K. Fan, S. Wang, Y. Ren, H. Li, and Y. Yang, "MedBlock: Efficient and secure medical data sharing via blockchain," *J. Med. Syst.*, vol. 42, no. 8, pp. 1–11, Aug. 2018.

[30] A. Vasighizaker and S. Jalili, "C-PUGP: A cluster-based positive unlabeled learning method for disease gene prediction and prioritization," *Comput. Biol. Chem.*, vol. 76, pp. 23–31, Oct. 2018.

[31] C. B. C. Latha and S. C. Jeeva, "Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques," *Informat. Med. Unlocked*, vol. 16, 2019, Art. no. 100203.

[32] P. R. Kumar, T. Arunprasath, M. P. Rajasekaran, and G. Vishnuvarthanan, "Computer-aided automated discrimination of Alzheimer's disease and its clinical progression in magnetic resonance images using hybrid clustering and game theory-based classification strategies," *Comput. Electr. Eng.*, vol. 72, pp. 283–295, Nov. 2018.

[33] M. Nilashi, O. B. Ibrahim, H. Ahmadi, and L. Shahmoradi, "An analytical method for diseases prediction using machine learning techniques," *Comput. Chem. Eng.*, vol. 106, pp. 212–223, Nov. 2017.

[34] N. Nidheesh, K. A. A. Nazeer, and P. M. Ameer, "An enhanced deterministic K-means clustering algorithm for cancer subtype prediction from gene expression data," *Comput. Biol. Med.*, vol. 91, pp. 213–221, Dec. 2017.

[35] R. J. Kuo, P. Y. Su, F. E. Zulvia, and C. C. Lin, "Integrating cluster analysis with granular computing for imbalanced data classification problem—A case study on prostate cancer prognosis," *Comput. Ind. Eng.*, vol. 125, pp. 319–332, Nov. 2018.

[36] A. Hasselgren, K. Kralevska, D. Gligoroski, S. A. Pedersen, and A. Faxvaag, "Blockchain in healthcare and health sciences—A scoping review," *Int. J. Med. Informat.*, vol. 134, Feb. 2020, Art. no. 104040.

[37] O. Dib, K.-L. Brousmiche, A. Durand, E. Thea, and E. B. Hamida, "Consortium blockchains: Overview applications and challenges," *Int. J. Adv. Telecommun.*, vol. 11, no. 1, pp. 51–64, 2018.

[38] D. Jain and V. Singh, "Feature selection and classification systems for chronic disease prediction: A review," *Egyptian Informat. J.*, vol. 19, no. 3, pp. 179–189, Nov. 2018.

[39] E. D. Übeyli, "Adaptive neuro-fuzzy inference system for classification of ECG signals using Lyapunov exponents," *Comput. Methods Programs Biomed.*, vol. 93, no. 3, pp. 313–321, Mar. 2009.

[40] J. Tang, S. Alelyani, and H. Liu, "Feature selection for classification: A review," *Data Classif, AlgorAppl*, vol. 97, no. 7, pp. 1660–1674, Aug. 2006.

[41] A. Christmann and S. Van Aelst, "Robust estimation of Cronbach's alpha," *J. Multivariate Anal.*, vol. 97, no. 7, pp. 1660–1674, Aug. 2006.

[42] J.-S. R. Jang, "ANFIS: Adaptive-network-based fuzzy inference system," *IEEE Trans. Syst., Man, Cybern.*, vol. 23, no. 3, pp. 665–685, May/Jun. 1993.

[43] M. Sugeno and G. T. Kang, "Structure identification of fuzzy model," *Fuzzy Sets Syst.*, vol. 28, no. 1, pp. 15–33, Oct. 1988.

[44] H. Das, B. Naik, and H. S. Behera, "Medical disease analysis using neuro-fuzzy with feature extraction model for classification," *Informat. Med. Unlocked*, vol. 18, 2020, Art. no. 100288.

[45] D. Rodrigues, L. A. M. Pereira, T. N. S. Almeida, J. P. Papa, A. N. Souza, C. C. O. Ramos, and X.-S. Yang, "BCS: A binary cuckoo search algorithm for feature selection," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2013, pp. 465–468.

[46] R. Jayaram and S. Prabakaran, "Onboard disease prediction and rehabilitation monitoring on secure edge-cloud integrated privacy preserving healthcare system," *Egyptian Informat. J.*, to be published, doi: 10.1016/j.eij.2020.12.003.

**P. G. SHYNU** (Member, IEEE) received the M.E. degree in computer science and engineering from the College of Engineering, Anna University, Chennai, India, and the Ph.D. degree in Computer Science from the Vellore Institute of Technology (VIT), Vellore, India. He is currently working as an Associate Professor with the School of Information Technology and Engineering, VIT. He has published more than 30 research papers in refereed international conferences and journals. His research interests include machine learning, cloud security and privacy, ad-hoc networks, and big data.

**VARUN G. MENON** (Senior Member, IEEE) is currently an Associate Professor with the Department of Computer Science and Engineering, SCMS School of Engineering and Technology, India. His research interests include the Internet of Things, fog computing and networking, underwater acoustic sensor networks, cyberpsychology, hijacked journals, ad-hoc networks, and wireless sensor networks. He is also a Distinguished Speaker of ACM Distinguished Speaker. He is also a Guest Editor of the IEEE Transactions on Industrial Informatics, the IEEE Sensors Journal, the *IEEE Internet of Things Magazine*, and the *Journal of Supercomputing*. He is an Associate Editor of *IET Quantum Communications*. He is also an Editorial Board Member of the IEEE Future Directions: Technology Policy and Ethics.

**R. LAKSHMANA KUMAR** (Member, IEEE) is currently associated with the Hindusthan College of Engineering and Technology, Coimbatore, Tamil Nadu. He is also the Director-Research and Development (AI) for a Canadian-based company (ASIQC) in Vancouver region of British Columbia, Canada. He is also the Founding Member of IEEE SIG of Big Data for Cyber Security and Privacy, IEEE. He serves as a Core Member in the Editorial Advisor Board of Artificial Intelligence Group in Cambridge Scholars Publishing, U.K., *Trends in Renewable Energy Journal*, USA, *Frontiers in Communications and Networks*,-Switzerland, AI Forum(The world's leading forum for AI). He is an IEEE Brand Ambassador. He was invited as a Keynote Speaker of AVIS' 2020 (Asia Artificial Intelligence Virtual Summit 2020) which is the Asia's first biggest Virtual Summit on Artificial Intelligence held at Malaysia, in June 2020. He is a global chapter Lead of Machine Learning for Cyber Security (MLCS). He himself involves in research and expertise in AI and Blockchain technologies. He holds the certification in Data Science from John Hopkins University, USA. He also holds the Amazon Cloud Architect certification from Amazon Web Services. He is also an ACM Distinguished Speaker.

**SEIFEDINE KADRY** (Senior Member, IEEE) received the bachelor's degree from Lebanese University, in 1999, the M.S. degree from Reims University, France, in 2002, the EPFL (Lausanne) and Ph.D. degrees from Blaise Pascal University, France, in 2007, and the HDR degree from Rouen University, in 2017. His current research interests include data science, education using technology, system prognostics, stochastic systems, and applied mathematics. He is an ABET Program Evaluator of computing, and an ABET Program Evaluator of Engineering Tech. He is a Fellow of IET, IETE, and IACSIT. He is a Distinguish Speaker of IEEE Computer Society.

**YUNYOUNG NAM** (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in computer engineering from Ajou University, South Korea, in 2001, 2003, and 2007, respectively. He was a Senior Researcher with the Center of Excellence in Ubiquitous System, Stony Brook University, Stony Brook, NY, USA, from 2007 to 2010, where he was a Postdoctoral Researcher, from 2009 to 2013. He was a Research Professor with Ajou University, from 2010 to 2011. He was a Postdoctoral Fellow with the Worcester Polytechnic Institute, Worcester, MA, USA, from 2013 to 2014. He was the Director of the ICT Convergence Rehabilitation Engineering Research Center, Soonchunhyang University, from 2017 to 2020. He has been the Director of the ICT Convergence Research Center, Soonchunhyang University, since 2020, where he is currently an Assistant Professor with the Department of Computer Science and Engineering. His research interests include multimedia database, ubiquitous computing, image processing, pattern recognition, context-awareness, conflict resolution, wearable computing, intelligent video surveillance, cloud computing, biomedical signal processing, rehabilitation, and healthcare systems.