Computers and Electrical Engineering 101 (2022) 108086

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

# A novel blockchain-enabled heart disease prediction mechanism using machine learning[☆]

Huru Hasanova [a], Muhammad Tufail [b], Ui-Jun Baek [a], Jee-Tae Park [a], Myung-Sup Kim [a,*]

[a] Department of Computer and Information Science, Korea University - Sejong Campus, Sejong, South Korea
[b] School of Design, The Hong Kong Polytechnic University, Hong Hom, Kowloon, Hong Kong, China

ARTICLE INFO

*Keywords:*
Blockchain
Sine cosine algorithm
K-Nearest Neighbour
Weighted K-Nearest Neighbour
Heart disease

ABSTRACT

Heart disease is one of the leading causes for death in men and women across the globe. Several characteristics that can be monitored to predict the heart disease in the earlier stage are blood pressure, cholesterol level, blood sugar level and body weight. The technology is revolutionizing the existing healthcare infrastructure. With the inclusion of Internet of Things (IoT), now we can monitor patients remotely, store their data, and process it for further analysis. However, the need is to propose new and advanced secured algorithms for fast processing and efficient detection of events. In this article, a machine learning based Sine Cosine Weighted K-Nearest Neighbour (SCA_WKNN) algorithm is proposed for the heart disease prediction that learns from the data being stored in blockchain. Since the data stored in the blockchain are tamper resistant, it acts as an authentic source for learning data and also as a secure storage environment for patient information. The performance of proposed SCA_WKNN is assessed in comparison with other algorithms in terms of accuracy, precision, recall, F-score, and root mean square error. Our analysis indicates that SCA_WKNN achieves 4.59% and 15.61% maximum accuracy than W K-NN and K-NN, respectively. Also, blockchain-based storage is compared with peer-to-peer storage in terms of latency and throughput. The blockchain-based decentralized storage achieves 25.03% maximum throughput than peer-to-peer storage.

## 1. Introduction

Internet of Things (IoT) enables devices to get connected and exchange information. The IoT has become a boon to agriculture, smart home, and healthcare industries in recent days. IoT aims to help healthcare professionals to perform their routine optimally by reducing time [1]. IoT helps to perform several activities that include remote monitoring of patients and progression of treatment in the hospital ecosystem [2]. Tremendous improvement in the medical field includes disease prediction, monitoring the patient for a particular disease based on various symptoms gathered through multiple IoT devices. Thus, a colossal datastore is required to store such a vast amount of data generated. Also, when integrating the data across various devices, the interoperability concern arises.

IoT devices aim to gather health-sensitive information, process the data, and exchange high-priority sensitive information [3]. IoT

☆ This paper is for special section VSI-iomt. Reviews were processed by Guest Editor Dr. Ashish Singh and recommended for publication.

* Corresponding author.
*E-mail addresses:* hhuru@korea.ac.kr (H. Hasanova), tufail.tufail@polyu.edu.hk (M. Tufail), pb1069@korea.ac.kr (U.-J. Baek), pjj5846@korea.ac.kr (J.-T. Park), tmskim@korea.ac.kr (M.-S. Kim).

https://doi.org/10.1016/j.compeleceng.2022.108086
Received 7 January 2022; Received in revised form 1 May 2022; Accepted 10 May 2022
Available online 16 May 2022

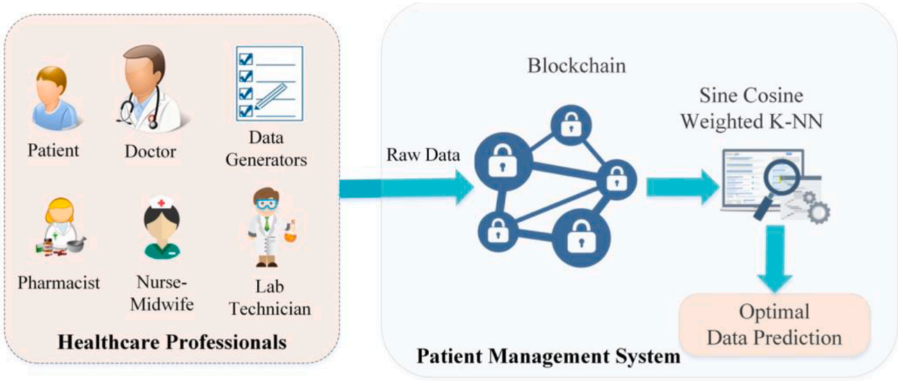


**Fig. 1.** Overall system design.

devices for healthcare applications deal with data that needs protection against confidentiality and privacy. The underlying architecture behind IoT is centralized, which leads to severe issues in the security and confidentiality of data. The conventional cryptographic way of offering protection brings threats to compassionate healthcare information. Thus, a means of providing security in a decentralized fashion is required.

The decentralized means of offering security and integration can be solved through Blockchain. The blockchain provides potential benefits by encrypting the data stored, and also each block is signed digitally to provide a high level of authenticity. Blockchain can be a suitable solution for the healthcare industry as it involves many actors and requires a high level of trust between the actors. In general, block chain is ideal for highly distributed applications, where the possibility to track the activities is high and where the reliability of the data is essential.

Blockchain helps in preserving the privacy of the numerous reports and it can run on billions of devices. The Healthcare industry is one of the leading areas in today's information and communication technology field. Electronic health records and remote patient monitoring are now made possible through the IoT in the healthcare industry. The healthcare data generated by various sources are extensive and are in different forms, which raises concerns about the quality of the data as well. Also, it is possible to use medical data in many applications, such as disease prediction. Thus, data quality should be ensured when integrating the data from various devices, and it is pretty challenging. Data Confidentiality problem arises when the healthcare data are shared across the network, and there is a possibility of single-point failure if the data are stored in a prominent centralized location. Also, the denial-of-service attack happens when the authorized storage is centralized. The solution that can address the issues mentioned earlier is made possible through the blockchain. Blockchain is a distributed storage that aims to synchronize information across healthcare providers. Each block contains sensitive health information, which is made available only to authorized persons. Also, blockchain is attracted by various features such as decentralized storage, consent, immutability, and increased capacity. Since the blockchain data is immutable, unauthorized persons cannot modify the data, and any disease prediction model can be built with machine learning algorithms. MeDShare is a blockchain based trustless medical data sharing system made for cloud service providers. It uses the blockchain technology mainly for replacing the traditional storge tier with a decentralised data processing unit. Medical service frameworks have been developed to protect the sensitive patient information with the help of blockchain [4]. MedBlock is a medical data sharing platform with the blockchain integration that can share medical records of the patients in a secure channel. MEdge-Chain is a medical data exchange framework that had been built with the intention to secure the data exchanged between the entities in the patient data management system. The predominant limitation that prevails in almost all the contribution is that, they involved blockchain to secure only the data and the communication, none of them integrated the learning paradigm with them.

In this paper, sine cosine algorithm-based Weighted K-NN has been utilized to design a predictor model for predicting heart disease. In general, Coronary artery disease, arrhythmia, heart valve disease and heart failure are considered to be the major types of the heart diseases. The condition where any abnormality in heart is experienced, or any disturbance, that affects the heart can be referred as Heart disease. In this article, we primarily focused the disease that arise due to blood vessel disorder. Fig. 1 represents the overall architecture consisting of patients, Healthcare Professionals (HCP), admin, and patient management system. The problem with the conventional K-NN is that it suffers from majority voting, whereas the Weighted K-NN issue is that it is a function of distance. Thus, to find the optimal assignment of the class label for the test instance to predict heart disease, the sine cosine algorithm is integrated with Weighted K-NN.

In general, the contributions of the article are as follows:

1 A framework for patient data management based on Blockchain that ensures the integrity of patient records.
2 A novel heart disease prediction algorithm based on behavioural factors of human to mitigate the disease in its earlier stage using meta heuristic algorithm.
3 A comprehensive strategy to ensure the authenticity of the learning data through multi-tier blockchain based learning source.

The complete scenario depicted in the Fig. 1 has been simulated with several test cases under public blockchain eco system. The proposed model has been implemented in the test network of public chain (Ethereum in this case) through smart contracts deployed in solidity. The proposed heart disease prediction system has been trained and tested with the dataset available in UCI public repository. The performance of the system has been analysed in terms of accuracy, precision, recall, F-score, and root mean square error. It is inferred thorough the analysis that the proposed prediction system increases the accuracy of prediction by an average of 9.85% when compared to the competing algorithms like WK-NN and K-NN.

The rest of the article is organized as follows. Section 1 provided the introduction and motivation towards the need to use blockchain for storing healthcare records. Section 2 details the related contributions in the applications of blockchain and machine learning algorithms. Section 3 elaborates the proposed Blockchain based Learning Mechanism on Patient Data. Section 4 reveals the experimental results and comparison of the proposed work with other existing works. Finally, Section 5 concludes the contributions with possible leads for further extension.

## 2. Related works

This section provides an overview of the contributions that are closely related to this work. The role of block chain in the IoT healthcare sector and its related work are discussed and then the machine learning strategies for data processing in the healthcare sector are briefed.

### *2.1. Blockchain in healthcare*

Blockchain offers highly significant usage in healthcare data, including drug counterfeiting, medical research. Using the blockchain direct transaction is possible without relying on any third party [5]. Medical record management is possible with the help of blockchain together with the clinical research and claim process. The model chain which uses blockchain is used to increase security for distributed patient data across many organizations [6]. Though Blockchain is a boon for the healthcare industry, some potential vulnerabilities include scalability, privacy, and cost. Blockchain addresses challenges such as interoperability and enables to share the medical data between the medical professionals and patients in a secure way. The blockchain designed to handle healthcare applications must be public, and it has to address scalability, security, data privacy. Drug fraud is a more challenging task requiring close attention and can be addressed using blockchain.

The health chain is developed using the IBM blockchain to establish a secure contract between healthcare professionals and patients. Third parties' intervention to tamper with the data is impossible when using blockchain to store the patient data [7]. Blockchain uses cryptographic notations for its effective functioning. All the participants in the blockchain network use public key infrastructure for performing any transactions [8]. Healthcare data gateway architecture was designed using blockchain, which helps patients share their data without compromising privacy, thereby promoting intelligence [9].

A pervasive social network-based healthcare system was integrated with a high level of security using the protocols called IEEE 802.15.6 and blockchain. The blockchain was used to securely transfer the health data across the distributed network. The patient health record architecture comprises a management platform in which users upload their records, and the security was offered with the help of the SHA-256 algorithm and Rivest-Shamir-Adleman algorithm. The architecture uses blockchain where consensus protocol is used to provide the proof of authority [10]. mHealth communication framework was designed using blockchain to have secure and effective storage of health data. The framework offers functionalities such as remote diagnostics, tracking, patient management, and medical informatics [11]. Blockchain-enabled intelligent IoT architecture was designed to use artificial intelligence techniques over the data that are securely distributed in blockchain network.

Even though there are multiple platforms and frameworks available for medical data management using blockchain, none of them have integrated intelligence with them as it incurs additional cost. In this article, a metaheuristic algorithm is proposed to learn from the available data in the blockchain ecosystem. The motivation behind the blockchain integration is to have a verified learning source. In a dynamic environment, there is always a higher chance to have data manipulation attacks and those can be prevented by the use of blockchain eco system.

### *2.2. Metaheuristics and machine learning methodologies for processing data*

An optimal set of features has to be selected for maximizing classification accuracy. A nature-inspired metaheuristic algorithm [12] is used to determine the necessary attributes to train the model to predict disease [13]. Metaheuristic algorithms can select an optimal set of features to enhance the classifier and predictor accuracy. A metaheuristic algorithm called Particle Swarm Optimization is used with K-Means for grouping the patients with similar symptoms. When a new patient arrives with similar symptoms, the new patient will be predicted with the disease by computing the distance between the clusters [14]. XGBoost and logistic regression were used to predict cardiovascular heart disease using parameters such as the number of cigarettes smoked per day and BMI level [15]. Relevant features are identified to predict heart disease using a Hybrid Random Forest with Linear Model to produce an accuracy of 88.7% [16]. Appropriate feature identification and a suitable classifier model were designed to improve heart disease prediction accuracy, including hybridization of Naïve Bayes and Logistic Regression [17]. Gait prediction was made using multi-layer perceptron for the diagnosis of Parkinson's disease. The multi-layer perceptron performance was compared with logistic regression, and the former had greater accuracy than later [18]. Simulated Annealing with K-Means was used to detect the optimal feature subset. Having selected the relevant features, the support vector machine was used to predict cancer, and the model performed will have the necessary features [19]. The genetic algorithm was used to reduce the size of the dataset by selecting the optimal attribute set, and then the machine

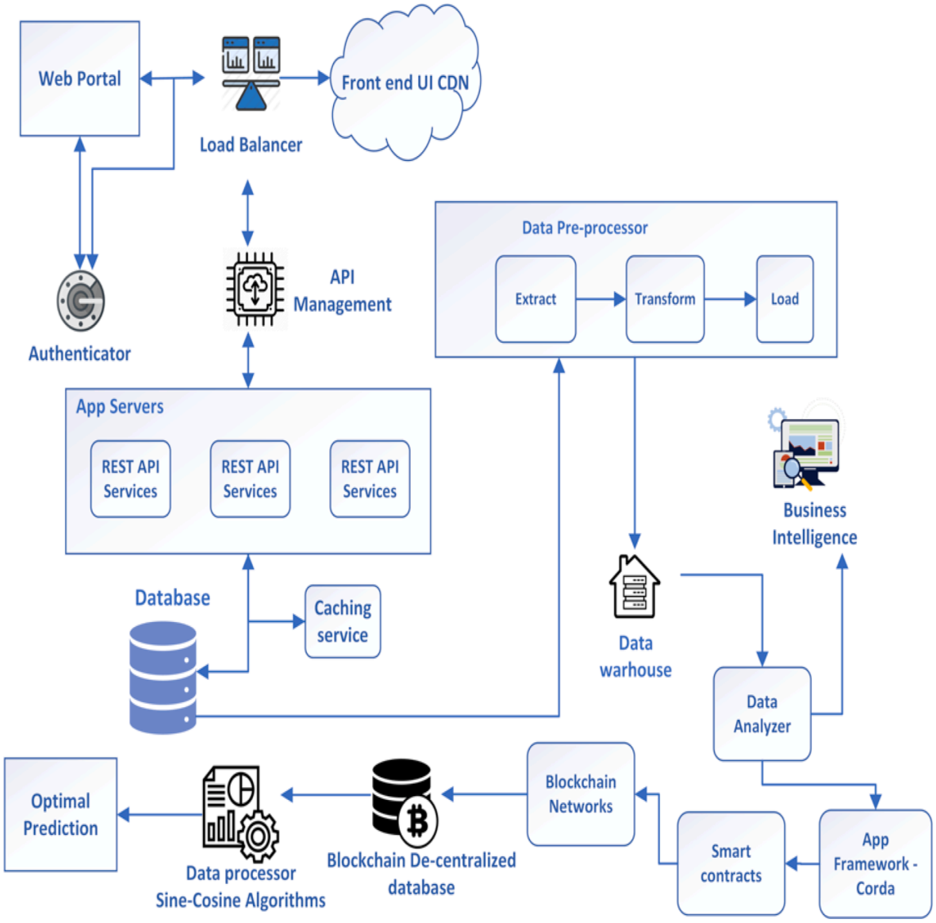


**Fig. 2.** Proposed BLMPD system architecture.

learning algorithms such as decision tree and Bayesian classifier were used to predict heart disease. The accuracy of the classifier was improved after selecting the relevant attributes [20]. A Coactive neuro-fuzzy inference system (CANFIS) was designed to predict heart disease, which integrated fuzzy inputs with a neural network. The best rules and other control parameters such as learning rate and momentum coefficient were selected using a genetic algorithm. Also, the essential features were determined using a genetic algorithm. The heart Disease Prediction System was designed using an artificial neural network. The designed prediction system had produced an accuracy of 80%.

Thus, machine learning and deep learning frameworks had been consistently used to learn the behaviour of the patient information and disease prediction. However, the authenticity of the information that is used in the learning process had not been audited in any aspect. A typical system should involve an authentic data source and it should be able to handle patient records seamlessly. In this article, the proposed system learns from the immutable source, which ensures the authenticity of the data and returns a greater prediction accuracy.

## 3. Blockchain based learning mechanism on patient data (BLMPD)

The overall system architecture and the process concerning the operations involved in the proposed BLMPD for Patient Data Management (PDM) using a blockchain network is discussed in this section.

### 3.1. Patient data management using blockchain

A shared symmetric key and private key manages the personal data and distributes the PD over the blockchain so that it can be accessible to healthcare professionals in the network. The algorithm for managing the patient data is given in Algorithm 1. It describes the patient data formation process by taking the patient information and the health care professional information as input.

**Table 1**
Notations used.

| Notation | Description |
|---|---|
| $P_{id}$ | Patient id |
| $PK_{P_{id}}$ | Private key of patient |
| $HCP_{id}$ | Id of Health care professional |
| $BCN$ | Block Chain Network |
| $AttackerA$ | One who tries to steal health information |
| $Req_{HCP_{id}}$ | Request from Health care professional |
| $Acc_i$ | Represents accuracy of i$^{th}$ solution |
| $A_i$ | Represents i$^{th}$ agent (solution) |
| $A_{i,d}^{t+1}$ | Represents value of the d$^{th}$ dimension of i$^{th}$ solution at iteration $t+1$ |
| $A_{Best}$ | Best Solution |

### 3.2. System architecture

The blockchain based PDM system architecture is detailed in this section. Fig. 2 represents the proposed BLMPD system architecture which consists of a data processing section, application servers to receive and transfer PD, Blockchain-based decentralized database, Application Programming Interface (API) management component, and data analytics unit. Some of the process that are performed in the system are Create_Patient_Data, Grant_Access_ToHCP, and Revoke_Access.

The structured data of the patient is recorded in the traditional databases, and the unstructured data will be dumped into the Data warehouse. The data should be cleaned enough before applying the machine learning algorithms to ensure higher accuracy in learning. The pre-processed cleaned data is stored in the distributed blockchain using a desirable smart contract written as per the use case. Once the data gets stored in the allocated blockchain, sine cosine algorithm integrated with weighted K-NN is used for analysing it to make an optimal prediction. The novelty in the proposal resides in the section of learning from the blockchain. The cleaned data is loaded into the decentralized blockchain, and the proposed SCA_WKNN is applied to the data fetched from the decentralized network.

The proposed PDM works as follows. Patients register through the user interface to get a private key from the certificate authority. Ethereum blockchain network is used to distribute patient data. Each actor performs different actions in the network and can only have access to the data for which they are provided access. Any patient can create their patient data and can commit the transaction. Once the transaction is committed in the blockchain network, the modified transactions are distributed across the web, ensuring every PD present in the network is made available to all the actors in the network and cannot be altered by unintended persons. When a new transaction is included in the blockchain, it is appended to the previous record together with the required information and timestamp.

PD's are updated, and it is made available to all actors in the blockchain network. The healthcare professionals shall be able to access the PD if and only if the corresponding patient gives access permission. Algorithm 2 represents the operations of the patient in the blockchain. The notations used are described in Table 1.

Healthcare professionals can read the patient's record and can diagnose the disease. The patient can create their medical record with the symptoms and upload it to the blockchain. The patient gives a request to the healthcare professional to whom they want to consult. When the doctor received the request, they will accept the request if they are available. Having received the request, the doctor sends the request to view the patient data to the intended patient. If the healthcare professional id is valid, the patient will accept the request to handle patient data. The healthcare professional then analyses the patient data to predict whether the patient is subjected to disease or not. Algorithm 3 represents the working of healthcare professionals.

### 3.3. Processing patient data using sine cosine algorithm weighted K-NN

The patient data stored in blockchain will have many dimensions. Machine learning can predict or diagnose diseasing by processing the data stored in the blockchain. Since the health data is sensitive, it demands higher accuracy, and in this work, the patient data are processed using K-NN. The conventional K-NN classifier algorithm suffers from majority voting, which degrades the performance of disease diagnosis. Therefore, to maximize disease diagnosis accuracy, a metaheuristic algorithm called Sine Cosine Algorithm is used. The motivation behind the selection of metaheuristic algorithm is to have a local optimal selection. Since, meta heuristic algorithms often does not search through the entire set of feasible solution, local optimal solution is obtained with less computation when compared to general optimization algorithms. The Sine Cosine algorithm solves the persistent problem and is used to find the optimal weight that intends to maximize disease diagnosis accuracy in patient data. Each agent in the SCA represents a feasible solution that represents the weight for each K neighbour. The agent is evaluated using the fitness defined in Eq. (1). The fitness function is estimated based on the accuracy and the number of attributes selected. Accuracy should be maximum on any excellent classifier algorithm. The fitness is made maximum, as shown in Eq. (1).

$$Maxf(A_i) = Acc_i \tag{1}$$

$Acc_i$ represents the accuracy of i$^{th}$ agent. At each iteration, the position of the agent will be computed using Eqs. (2) and (3).

$$A_{i,d}^{t+1} = A_{i,d}^{t} + r_1 * sin(r_2) + \left| r_3 * A_{Best,d}^{t} - A_{i,d}^{t} \right| \; if(r_4 < 0.5) \tag{2}$$

**Table 2**
Evaluation metrics.

| Metrics | Formulae |
|---|---|
| Accuracy | $\frac{TP+TN}{TP+TN+FP+FN}$ |
| Precision | $\frac{TP}{TP+FP}$ |
| Recall | $\frac{TP}{TP+FN}$ |
| F-Measure | $2*\frac{precision*recall}{precision+recall}$ |
| Root Mean Square Error | $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\widehat{x}_i)^2}$ |

$$A_{i,d}^{t+1}=A_{i,d}^{t}+r_1*cos(r_2)+\left|r_3*A_{Best,d}^{t}-A_{i,d}^{t}\right|\ if(r_4\geq 0.5) \tag{3}$$

Where $A_{i,d}^{t}$ represents the position of the agent $i$ for the dimension $d$at the iteration $t$. $A_{Best,d}^{t}$ represents the best agent for the dimension $d$at the iteration $t$. A random variable $r_1$ is used to control exploration and exploitation represented in Eq. (4).

$$r_1=a-t*\frac{a}{T} \tag{4}$$

Where $t$ represents the current iteration, $T$ represents the maximum number of iterations for fitness convergence. $a$ is a constant which is initialized to 2 and gradually decreased to 0. The random variable $r_2$ is used to check whether all the agents move towards the best agent [21]. The random variable $r_3$assigns weight to the best agent and the random variable $r_4$ is used to switch between sine and cosine function.

Algorithm 4 represents the illustration of the use of sine cosine algorithm for finding optimal weights for predicting disease using weighted K-Nearest Neighbour algorithm.

Algorithm 4, a simple K-NN algorithm for finding the K-neighbours, is invoked first which then invokes Algorithm 5 for finding the optimum class label for the patient data. Algorithm 5 illustrates the assignment of class label to the instance. K-Nearest Neighbour, a lazy learning that assigns a class label based on the instance's nearest neighbours. The Euclidean distance is computed between the new patient data with the instances in the dataset. K neighbours having minimum distance are chosen and that is given as input for SCA_WKNN. SCA_WKNN aims to find the optimal class label for the patient data by assigning weight to the distance between the patient data and its neighbours. The weight value is updated at each iteration according to the Eqs. (2) and (3). The agent with the best fitness is computed after the maximum number of iteration and the class label of that agent is assigned to new patient data. The class label actually depicts the presence of disease or not. Algorithm 5 illustrates the working of SCAWKNN for disease prediction.

## 4. Experimental results

The proposed data processing method combing SCA and K-NN has been implemented under the blockchain ecosystem to check the scalability and deployment feasibility. The proposed method is evaluated under specified scenario constituting different set of end users namely, Healthcare professionals, Doctors and the patients.

### *4.1. Datasets used*

The data set used for the purpose of training as well as testing the proposed SCA_WKNN was taken from the UCI repository [22]. It consists of 303 instances and 76 attributes. amongst the 76 attributes, only 14 key attributes are utilised in this research. Some of the important attributes that are considered for analysis are chest pain type, resting blood pressure, serum cholesterol, and fasting blood sugar. The objective of the dataset is to identify whether a patient had suffered with heart disease or not.

### *4.2. Performance evaluation*

Various experimentation was performed to show the efficiency of the proposed data processing mechanism using SCA-WKNN. The patient stores their data in the blockchain. The blockchain provides secure access to the patient data, as the health data is sensitive, it is essential to protect the data. Thus, blockchain is used for storing the patient data. The data from the blockchain is accessed by the healthcare professional to predict the disease. The optimal assignment of class label for new test instance is intended for maximizing the accuracy of the prediction are selected using the metaheuristic algorithm called sine cosine optimization algorithm. The proposed SCA-WKNN for processing patient data is compared with other traditional machine learning methods such as Support vector machine, random forest, multi-layer perceptron, weighted K-NN, and K-NN. All the classifiers are compared before selecting the necessary features and also after choosing the features. To measure the efficiency, the dataset is divided into 70:30 i.e., 70% data is used for training, and 30% is used for testing. The model is evaluated 10 times, and the mean value is taken for analysis.

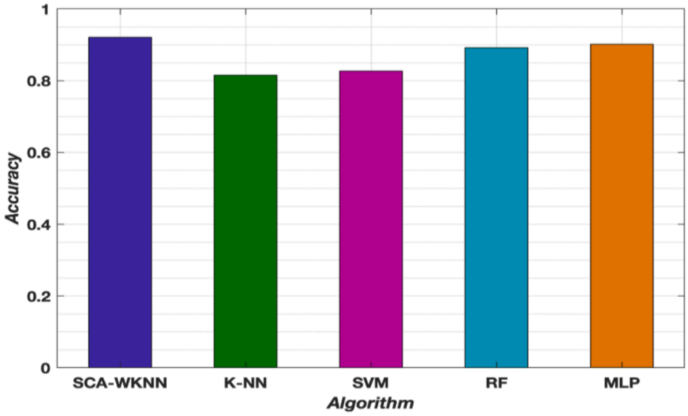


**Fig. 3.** Comparison of accuracy.

**Table 3**
Comparison of classifier.

| Metrics \ Classifier | SCA_WKNN | K-NN | SVM | RF | MLP |
|---|---|---|---|---|---|
| Accuracy | 0.9213 | 0.8149 | 0.827 | 0.8914 | 0.9014 |
| Precision | 0.8821 | 0.697 | 0.79 | 0.81 | 0.8713 |
| Recall | 0.9327 | 0.62 | 0.7103 | 0.8104 | 0.8513 |
| F-Measure | 0.9066 | 0.6562 | 0.7480 | 0.8102 | 0.8611 |
| RMSE | 0.1115 | 0.25 | 0.41 | 0.18 | 0.123 |

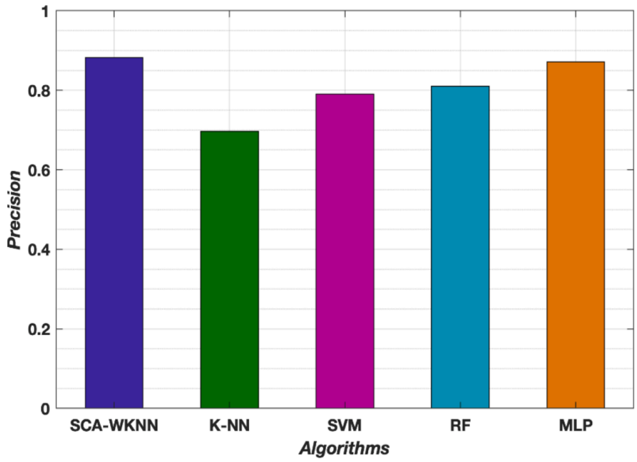


**Fig. 4.** Comparison of precision.

### *4.3. Evaluation metrics*

The performance metrics used to measure the efficiency of the proposed method is shown in the Table 2. Accuracy is defined as the correct prediction of the instances using the proposed SCA and K_NN. The positive instances that are correctly classified as positive are given as true positives (TP), False negative (FN) measures the number of positive instances which are identified as negative. False positive (FP) represents the instances that have no disease but predicted as positive. True negative (TN) represents the negative instances that are absence of disease is correctly identified as absence by the model. x represents the actual value and $\widehat{x}_i$ represents the predicted value of the classifier.

### 4.4. Results and discussions

The proposed method is compared with the algorithms specified in Section 4.2 for the metrics discussed in Section 4.3. Accuracy is measured and compared with all the classifiers and the result is shown in the Fig. 2. The data from the blockchain is analysed for

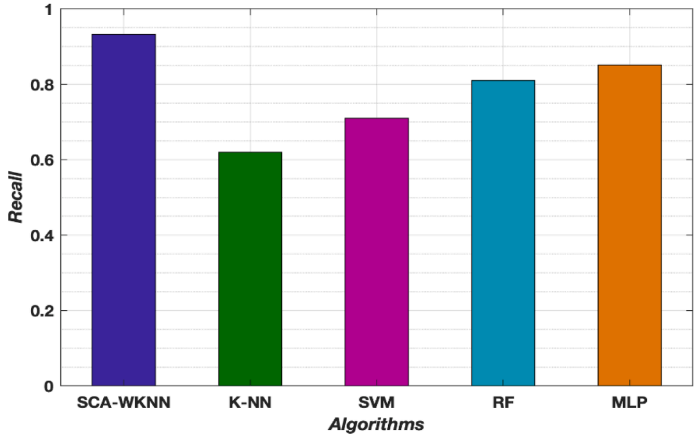


**Fig. 5.** Comparison of recall.

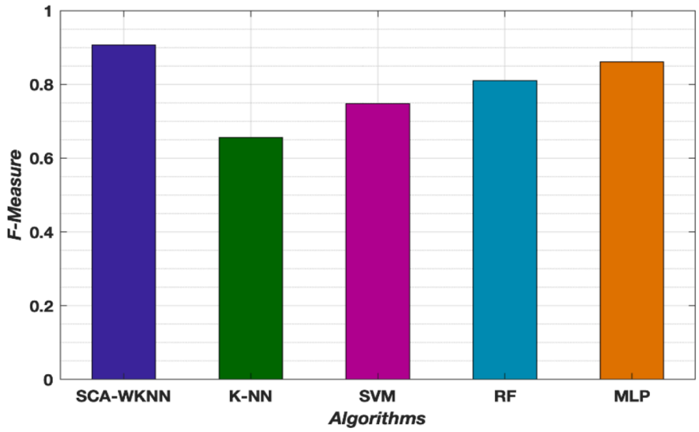


**Fig. 6.** Comparison of F – measure.

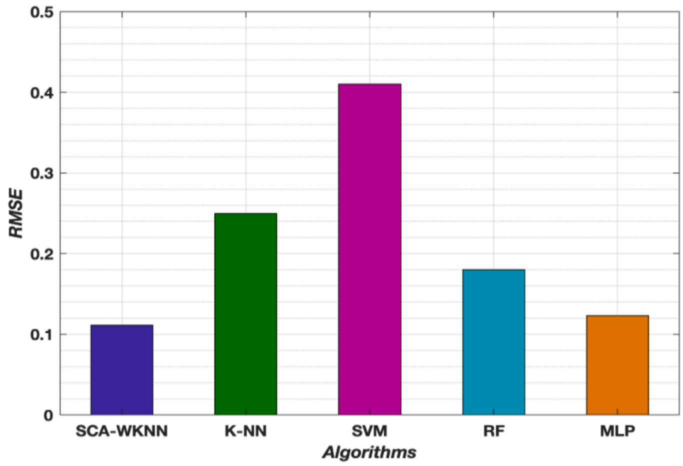


**Fig. 7.** Comparison of RMSE.

prediction of disease. From Fig. 3 it is evident that accuracy is maximum for W K-NN where $K = 1$. Table 3 shows the accuracy of various classifiers such as SCA_WKNN, K-NN, SVM, RF, MLP. Since medical data are highly sensitive, with the goal to further maximize the accuracy, SCA-WKNN is proposed. Figs. 3–7 shows the graphical representation of a comparison of performance metrics such as

**Table 4**
Comparison of classifier.

| Metrics \ Classifier | SCA_WKNN | W K-NN | K-NN |
|---|---|---|---|
| Accuracy | 0.9657 | 0.9213 | 0.8149 |
| Precision | 0.91 | 0.8821 | 0.697 |
| Recall | 0.9345 | 0.9327 | 0.62 |
| F-Measure | 0.941 | 0.9066 | 0.6562 |
| RMSE | 0.023 | 0.1115 | 0.25 |

**Table 5**
Time taken by the classifier.

| Classifier | Time Taken (ms) |
|---|---|
| SCA_WKNN | 2.05 |
| W K-NN | 1.735 |
| K-NN | 1.534 |

**Table 6**
Comparing the proposed approach with other existing approaches.

| Methodology | Classifier | Accuracy |
|---|---|---|
| Ali F et al. [1,36] | Deep Learning | 83.5 |
| Fitriyani NL et al. [37] | Density Based Spatial Clustering | 95.9 |
| Ayon SI et al. [3,38] | Deep Neural Network | 98.15 |
| Proposed Approach | SCA_WKNN | 96.57 |

**Algorithm 1**
Patient Data formation ($P_{id}$,$HCP_{id}$).

**Input:** Patient ID $P_{id}$, Health Care Professional ID $HCP_{id}$
**Output:** Access to Patient Data $PD$

**Begin**
**For** $\forall P_{id} \in P$
Include $P_{id}$ in Block chain $BCN \leftarrow BCN \bigcup \{ P\_\{id\}\}$
Give Access $(P_{id}, PK_{P_{id}})$
**EndFor**
**For** $\forall \mathrm{HCP}_{id} \in \mathrm{HCP}$
Include $HCP_{id}$ in Block chain $BCN \leftarrow BCN \bigcup \{HCP_{id}\}$
Give Access $(HCP_{id}, PK_{HCP_{id}})$
**End For**
**For** $\forall P_{id} \in P$
**If** $P_{id} \in AttackerA$
Remove $P_{id}$ from Block chain $BCN \leftarrow BCN - \{P_{id}\}$
**End If**
**End For**
**For** $\forall HCP_{id} \in P$
**If** $HCP_{id} \in AttackerA$
Remove $HCP_{id}$ from Block chain $BCN \leftarrow BCN - \{P_{id}\}$
**End If**
**End For**
**End**

accuracy, precision, recall, F-measure, RMSE.

From Table 4, it is evident that Accuracy, F-Measure, Precision and recall are increased for proposed SCA-WKNN when comparing to other classifiers. Root Mean Square Error is very low for SCA_WKNN than other classifiers. As the proposed SCA_WKNN achieves the highest accuracy, it validates that SCA assigns an optimal class label for the new instance. When the metaheuristic algorithm Sine cosine algorithm is applied to maximize the accuracy of weighted K-NN, the accuracy of SCA_WKNN is maximum than other classifiers such as W K-NN and K-NN.

Table 5 shows the time taken by various algorithms such as SCA_WKNN, W K-NN and K-NN. The time taken by SCA-WKNN is maximum when compared to W K-NN and K-NN. The K-NN finds the distance between the test instance and all the instances in the dataset to find the neighbours. W K-NN also tends to find more time in finding the neighbour for the test instance and thus time is high in W K-NN than K-NN also. SCA-WKNN is invoked after finding the neighbours using K-NN; thus, time is maximum is SCA_WKNN.

Table 6 shows the prediction of heart disease using the proposed approach with other existing approaches. The proposed SCA_WKNN achieves 13.53% greater accuracy than Ali F et al. [23] approach. Also, SCA_WKNN achieves 0.69% improved accuracy than Firtriyani et al. [24]approach. But the proposed SCA_WKNN is 1.6% minimum in accuracy than Ayon et al. [25]method which

**Algorithm 2**
Patient_Data_Operation($P_{id}$, $PK_{P_{id}}$, $Req_{HCP_{id}}$).

| | |
|---|---|
| **Input:** | Patient ID $P_{id}$, Private key of patient $PK_{P_{id}}$, HCP Request $Req_{HCP_{id}}$ |
| **Output:** | Access to Patient Data PD |
| | **Begin** |
| | **If** $P_{id} \in BCN$ |
| | **If** $PD_{P_{id}} \notin BCN$ |
| | CreatePatientData($P_{id}$, $PD_{P_{id}}$, $BCN$) |
| | **Else** |
| | UpdatePatientData($P_{id}$, $PD_{P_{id}}$, $BCN$) |
| | **End If** |
| | **Else** |
| | Print Invalid $P_{id}$ |
| | **End If** |
| | **If** $Req_{HCP_{id}}$ is received |
| | **If** $HCP_{id} \notin AttackerA$ |
| | Give Access($HCP_{id}$) |
| | **End If** |
| | **End If** |
| | **End** |

**Algorithm 3**
HealthCareProfessional_Operation($HCP_{id}$).

| | |
|---|---|
| **Input:** | Health Care Professional ID $HCP_{id}$ |
| **Output:** | Predict the disease for Patient Data $PD$ |
| | **Begin** |
| | **If** $Req_{P_{id}}$ is received |
| | **If** $P_{id} \notin AttackerA$ |
| | ViewPatientData($HCP_{id}$, $PD_{P_{id}}$, $BCN$) |
| | K-NN_Disease_Prediction() |
| | **End If** |
| | **End If** |
| | **End** |

**Algorithm 4**
K-NN_Disease_Prediction(Dataset, PD).

| | |
|---|---|
| **Input:** | *Dataset*, and *PatientDataPD* |
| **Output:** | K Neighbours |
| | **Begin** |
| | **For** each instance $X_i \in Dataset$ |
| | Compute the Euclidean distance as $dist(PD, X_i)$ |
| | **End For** |
| | φ = Find the K-Neighbors of *PD* |
| | Invoke SCA_WKNN (φ) |
| | **End** |

shows further research is needed to still maximize the accuracy of SCA_WKNN in optimal prediction of disease.

The next level of experimentation is addressed using blockchain to store the patient data rather than using centralized storage. From Fig. 8, it is evident that when the number of patients is increasing in the network, the latency is also increased. But the latency is maintained optimal in all the cases for blockchain based decentralized storage compared to peer-to-peer storage. When the number of patients reaches 100, blockchain-based storage achieves 34.63% minimum accuracy than centralized storage. When the number of patients is reached to 500, blockchain-based storage is 19.45% minimum latency than peer-to-peer storage. In all the cases, from minimum patients to maximum patients, the blockchain-based decentralized storage achieves optimal latency than other storage schemes.

The next experimentation is carried out to measure throughput for peer-to-peer and decentralized block-chain based storage. From Fig. 9, it is evident that blockchain-based storage achieves high throughput than peer to peer storage. When the number of patients is 200, blockchain-based storage achieves 43.52% higher throughput than peer to peer storage. The throughput for blockchain-based storage is 15.53%, 12.66%, 25.03% higher throughput than peer to peer storage.

***Limitations:***

- The proposed system has proven to have an authentic learning source but the cost of the operation depends on the number of transactions we do in the system and it is often expensive.

**Algorithm 5**
SCA_WKNN (φ, PD).

| | |
|---|---|
| **Input:** | φ neighbors, Patient Data *PD* |
| **Output:** | Class Label i.e. Disease prediction for *PD* |
| | **Begin** |
| | Initialize the agents *A* with φ neighbors |
| | **while** ($t! = T$) |
| | **For** $\forall A_i$ |
| | Compute fitness using Eq. (1) |
| | **End For** |
| | $A_{Best} = (A_i) \mid f(A_i)$ |
| | **For** $\forall A_i$ |
| | **If** fitness of Agent $A_i$is better than $A_{Best}$ |
| | $A_{Best} = A_i$ |
| | **End If** |
| | **End For** |
| | Compute $r_1$using Eq. (4) |
| | $r_2 = GenerateRand(\$)$ |
| | $r_3 = GenerateRand(\$)$ |
| | $r_4 = GenerateRand(\$)$ |
| | **For** $\forall A_i$ |
| | **If** $r_4 < 0.5$ |
| | Compute position using Eq. (2) |
| | **Else** |
| | Compute position using Eq. (3) |
| | **End If** |
| | **End For** |
| | **End while** |
| | **Assign** class label for PD as $A_{Best}$ |
| | **End** |

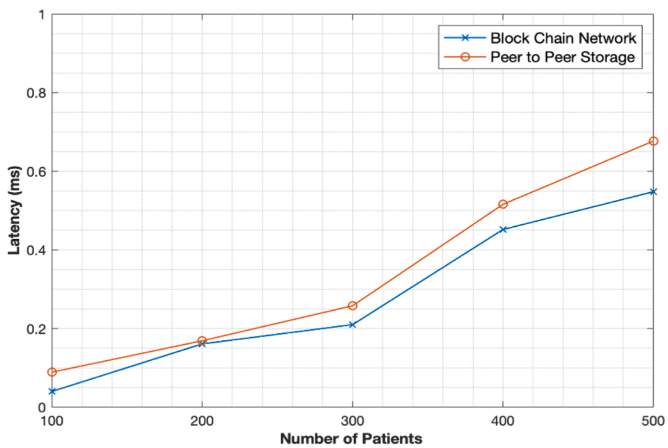


**Fig. 8.** Comparison of latency.

- Due to the decentralized nature of the system, the time incurred for the transactions will have a slight delay, and hence, this mechanism is not advisable to the low latency scenarios.
- As the data grows, the system needs to be restricted to learn from a limited data source, else it will incur additional costs to the system.

## 5. Conclusion

The blockchain plays a vital role in securing the health care data electronically with a great level of privacy. In this paper, the importance of using the blockchain in healthcare records is studied and the problem has been well addressed by storing the medical data in a blockchain network. The second important thing that is required in healthcare industry is optimal prediction. Thus, a metaheuristic algorithm called sine cosine based weighted K-NN is used for the optimal prediction of heart disease from the data stored in blockchain. The proposed SCA_WKNN achieves maximum accuracy, precision, recall, F-measure and minimum root mean square error when compared to the traditional algorithms. When SCA_WKNN is compared with W K-NN and K-NN, SCA_WKNN gives optimal assignment of class label. The proposed blockchain-based storage aims to maximize throughput and accuracy when compared with

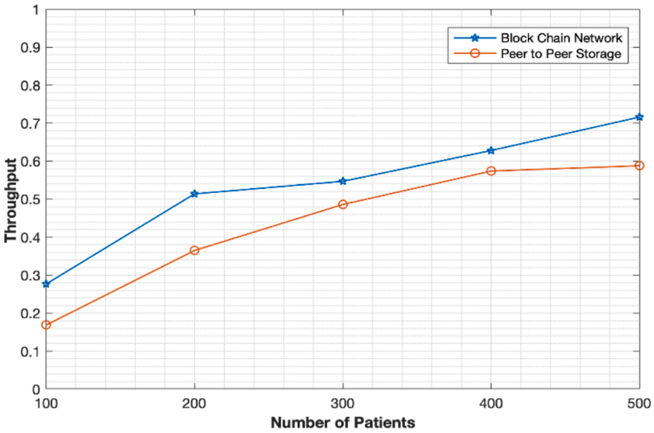


**Fig. 9.** Comparison of throughput.

centralized storage. As the blockchain is used for distributed storage, single point failure has been tackled to a great extent. Further the use of SCA_WKNN in turn maximize the accuracy in prediction by 13.05% than conventional WKNN. Also it is inferred through the analysis that, the proposed SCA_WKNN shows 16% improvement in the detection accuracy due to reliable data source. The future scope of this work can be further extended to address the latency issue that arise due to blockchain integration and also to predict several chronic diseases that affects the internal organs like kidney and liver.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References


[1] Arul R, Al-Otaibi YD, Alnumay WS, Tariq U, Shoaib U, Piran MD. Multi-modal secure healthcare data dissemination framework using blockchain in IoMT. Pers Ubiquitous Comput 2021:1–13. https://doi.org/10.1007/s00779-021-01527-2.
[2] Arul R, Alroobaea R, Tariq U, Almulihi AH, Alharithi FS, Shoaib U. IoT-enabled healthcare systems using block chain-dependent adaptable services. Pers Ubiquitous Comput 2021:1–15. https://doi.org/10.1007/s00779-021-01584-7.
[3] Alkurdi F, Elgendi I, Munasinghe KS, Sharma D, Jamalipour A. Blockchain in IoT security: a survey. In: Proceedings of the 28th international telecommunication networks and applications conference (ITNAC); 2018.
[4] Chen Y, Ding S, Xu Z, Zheng H, Yang S. Blockchain-based medical records secure storage and medical service framework. J Med Syst 2019;43:1–9.
[5] Zhuang Y, Sheets LR, Chen YW, Shae ZY, Tsai JJP, Shyu CR. A patient-centric health information exchange framework using blockchain technology. IEEE J Biomed Health Inform 2020;24:2169–76.
[6] Kuo TT, Kim HE, Ohno-Machado L. Blockchain distributed ledger technologies for biomedical and health care applications. J Am Med Inform Assoc 2017;24:1211–20.
[7] Jamil F, Ahmad S, Iqbal N, Kim DH. Towards a remote monitoring of patient vital signs based on IoT-based blockchain integrity management platforms in smart hospitals. Sensors 2020;20:2195.
[8] Rasool S, Saleem A, Iqbal M, Dagiuklas T, Bashir AK, Mumtaz S, Al Otaibi S. Blockchain-enabled reliable osmotic computing for cloud of things: applications and challenges. IEEE Internet Things Mag 2020;3:63–7.
[9] Abou-Nassar EM, Iliyasu AM, El-Kafrawy PM, Song O-Y, Bashir AK, Abd El-Latif AA. DITrust chain: towards blockchain-based trust models for sustainable healthcare IoT systems. IEEE Access 2020;8:111223–38.
[10] Lee HA, Kung HH, Udayasankaran JG, Kijsanayotin B, Marcelo AB, Chao LR, Hsu CY, et al. An architecture and management platform for blockchain-based personal health record exchange: development and usability study. J Med Internet Res 2020;22:e16748.
[11] Alam T. mHealth communication framework using blockchain and IoT technologies. Int J Sci Technol Res 2020;9(6):240–5.
[12] Iwendi C, Maddikunta PKR, Gadekallu TR, Lakshmanna K, Bashir AK, Piran MJ. A metaheuristic optimization approach for energy efficiency in the IoT networks. Softw Pract Exp 2021;51:2558–71.
[13] Sharma M, Kaur P. A comprehensive analysis of nature-inspired meta-heuristic techniques for feature selection problem. Arch Comput Methods Eng 2021;28:1103–27.
[14] Haq AU, Li JP, Khan J, Memon MH, Nazir S, Ahmad S, Khan GA, Ali A. Intelligent machine learning approach for effective recognition of diabetes in E-healthcare using clinical data. Sensors 2020;20:2649.
[15] Nalluri S, Vijaya Saraswathi R, Ramasubbareddy S, Govinda K, Swetha E. Chronic heart disease prediction using data mining techniques. Data engineering and communication technology. Springer; 2020. p. 903–12.
[16] Mohan S, Thirumalai C, Srivastava G. Effective heart disease prediction using hybrid machine learning techniques. IEEE Access 2019;7:81542–54.
[17] Diwakar M, Tripathi A, Joshi K, Memoria M, Singh P. others, "Latest trends on heart disease prediction using machine learning and image fusion. Mater Today Proc 2021;37:3213–8.
[18] Zahid L, Maqsood M, Durrani MY, Bakhtyar M, Baber J, Jamal H, Mehmood I, Song OY. A spectrogram-based deep feature assisted computer-aided diagnostic system for Parkinson's disease. IEEE Access 2020;8:35482–95.
[19] Huang S, Cai N, Pacheco PP, Narrandes S, Wang Y, Xu W. Applications of support vector machine (SVM) learning in cancer genomics. Cancer Genom Proteom 2018;15:41–51.
[20] Soni J, Ansari U, Sharma D, Soni S. Predictive data mining for medical diagnosis: an overview of heart disease prediction. Int J Comput Appl 2011;17:43–8.
[21] Zhang J, Zhou Y, Luo Q. An improved sine cosine water wave optimization algorithm for global optimization. J Intell Fuzzy Syst 2018;34:2129–41.
[22] Dua, D.; Graff, C. (2019). UCI Machine Learning Repository. Irvine, CA: University of California, School of Information and Computer Science. Retrieved from [http://archive.ics.uci.edu/ml], Accessed on 13/05/2022.

[23] Ali F, El-Sappagh S, Islam SR, Kwak D, Ali A, Imran M, Kwak KS. A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion. Inf Fusion 2020;63:208–22.
[24] Fitriyani NL, Syafrudin M, Alfian G, Rhee J. HDPM: an effective heart disease prediction model for a clinical decision support system. IEEE Access 2020;8:133034–50.
[25] Ayon SI, Islam MM, Hossain MR. Coronary artery heart disease prediction: a comparative study of computational intelligence techniques. IETE J Res 2020:1–20. https://doi.org/10.1080/03772063.2020.1713916.

**Huru Hasanova** received her BS and MS degrees in Computer Science from Azerbaijan Technical University in 2008 and 2010, respectively. Now, she is a PhD student at the Department of Computer and Information Science, Korea University-Sejong Campus. Her research focuses on internet security and Distributive Ledger Technologies.

**Muhammad Tufail** is a research assistant professor at the Hong Kong Polytechnic University. He has worked as a postdoctoral research associate at Ulsan National Institute of Science and Technology (UNIST). He obtained his master's degree from Korea Advanced Institute of Science and Technology (KAIST), and Ph.D. degree from UNIST. His-areas of interest include healthcare, ergonomic design, and service design.

**Ui-Jun Baek** is a PhD student in the Department of Computer and Information Science, Korea University, Korea. He received his BS in Computer and Information Science from Korea University, Korea in 2018. His research interest include Blokchain Network Analysis, Blockchain Network Design, and Network Security, Network Privacy.

**Jee-Tae Park** is a PhD student in the Department of Computer and Information Science, Korea University, Korea. He received his BS in Computer and Information Science from Korea University, Korea in 2017. His research interests include Internet traffic monitoring, Internet traffic analysis, network management, network services and Internet security.

**Myung-Sup Kim** is a professor in the Department of Computer and Information Science, Korea University, Korea. He received his BS, MS, and PhD degree in Computer Science and Engineering from POSTECH, Korea, in 1998, 2000, and 2004, respectively. He joined Korea University in September 2006. His-research interests include Internet traffic monitoring, Internet traffic analysis, network management, and Internet security.