Information Sciences 538 (2020) 486–502

Contents lists available at ScienceDirect

Information Sciences

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

# An accurate and dynamic predictive model for a smart M-Health system using machine learning

Kashif Naseer Qureshi [a], Sadia Din [b], Gwanggil Jeon [c,d,*], Francesco Piccialli [e,*]

[a] Department of Computer Science, Bahria University, Islamabad, Pakistan
[b] School of Computer Science and Engineering, Kyungpook National University, Daegu, Republic of Korea
[c] School of Electronic Engineering, Xidian University, Xi'an 710071, China
[d] Department of Embedded Systems Engineering, Incheon National University, Incheon 22012, Republic of Korea
[e] Department of Mathematics and Applications "R. Caccioppoli", University of Naples Federico II, Italy

ARTICLE INFO

*Article history:*
Received 27 November 2019
Received in revised form 5 June 2020
Accepted 8 June 2020
Available online 20 June 2020

*Keywords:*
Machine learning
Predictive
Models
M-Health
Classification
SVM
Decision tree
Accuracy

ABSTRACT

Nowadays, new highly-developed technologies are changing traditional processes related to medical and healthcare systems. Emerging Mobile Health (M-Health) systems are examples of novel technologies based on advanced data communication, deep learning, artificial intelligence, cloud computing, big data, and other machine learning methods. Data are collected from sensor nodes and forwarded to local databases through new technologies that enable cellular networks and then store the information in cloud storage systems. From cloud computing services or medical centres, the data are collected for further analysis. Furthermore, machine learning techniques are being used for accurate prediction of disease analysis and for purposes of classification. This paper presents a detailed overview of M-Health systems, their model and architecture, technologies and applications and also discusses statistical and machine learning approaches. We also propose a secure Android-based architecture to collect patient data, a reliable cloud-based model for data storage. Finally, a predictive model able to classify cardiovascular diseases according to their seriousness will be discussed. Moreover, the proposed prediction model has been compared with existing models in terms of accuracy, sensitivity, and specificity. The experimental results show encouraging results in terms of the proposed predictive model for an M-Health system. *Keywords:* Machine Learning, Predictive, Models, M-Health, Classification, SVM, Decision Tree, Accuracy



## 1. Introduction

The healthcare sector is encountering various issues related to disease diagnosis, and cost-effective service provision [1]. One of the critical requirements of a healthcare system is to provide opportune treatment to the patient as a result of an analysis of patient data, lifestyle habits, and any variability in molecular traits. With the rapid development of new technologies, these systems have been facing various challenges related to data collection, the association of information, data extraction and decision making. To address these challenges, several intelligent tools have been designed based on machine learning and data-driven approaches. These approaches have consisted in an attempt to link multiple data sources in order to establish collective knowledge for further discovery and predictive analysis. Additionally, various prediction-based meth-

* Corresponding authors.
*E-mail addresses:* ggjeon@gmail.com (G. Jeon), francesco.piccialli@unina.it (F. Piccialli).

https://doi.org/10.1016/j.ins.2020.06.025
0020-0255/

ods have been designed for specific diseases like brain tumours and cardiovascular heart diseases [2,3]. However, there are many challenges to face in connection with biomedical data due to their temporal dependencies, irregularities, sparsity, high-dimensionalities and heterogeneities [4].

New and integrated technology-enabled systems have been used for disease prediction, Electronic Health Records (EHR), Mobile Health Systems (M-Health), Electronic Medicine (E-Medicine) and other Body Networks (BN). EHR refers to the maintaining of digital records or information for administrative healthcare tasks. EHR systems have been adopted all over the world and represent one of the most significant requirements for physicians and hospitals. According to the National Coordinator for Health Information Technology, around 84% of hospitals have utilized EHR systems. In these smart systems, the data are related to the patients' demographic information, laboratory reports, image-based radiological records, prescriptions, medicine dosage records and, dietician appointment records. M-Health systems are providing health services by using handheld and wearable devices. In this filed, the patient information is collected in order to diagnose and track the disease and provide timely services, especially for those areas where immediate treatment is a challenge. E-Medicine is another online system containing a vast repository of medical contents based on multimedia materials.

Despite the presence of these new communication systems, the technological provision to the healthcare field does not yet meet all requirements, due to scalability and multiple technology problems. There is a need to design a more appropriate M-Health architecture to facilitate the service to providers, users and medical staff. All data communication should be secure and fast for an opportune data delivery. Security is one of the most important features, especially in connection with wearable sensors and hand-held devices. The patient data require security-based solutions from the application layer to the cloud level [5,6]. The massive amount of related medical data is stored in the cloud, so determining how we can use such information to obtain better predictions represents a significant challenge. Predictive models are advantageous to analyze the data and predict patient statistics. In this perspective, this paper presents and discusses three smart healthcare models, addressing the limitations mentioned above.

First, machine learning methods have been used for data regression and classification. Previous studies have focused on machine learning predictive methods for specific diseases like brain tumours and heart diseases. To the best of our knowledge, this is the first attempt to discuss and design a complete M-Health system, in terms of applications, the cloud, machine learning models and other statistical techniques, and to present predictive models for future analysis. Additional objectives of the paper are as follows:

- to review M-Health in detail, its technologies, models, architectures, applications and data collection processes;
- to propose an M-Health home and hospital services model;
- to propose an Android-based secure application architecture for an M-Health system;
- to propose a secure cloud-based model for an M-Health system;
- to discuss the machine learning methods employed for M-Health data analysis. and, finally,
- to propose a machine learning-based predictive model for an M-Health system to improve the efficiency of existing systems.

The rest of the paper is organized as follows: Section 2 presents the proposed M-Health system, including operational fundamentals, phases, applications and cloud models. Section 3 presents details about machine learning methods and their implementation in an M-Health system. Section 4 presents the proposed machine learning-based dynamic and accurate predictive model, and all its design phases. Section 5 presents the results and a discussion of the proposed model. Finally, Section 6 concludes the paper, outlining future research directions.

## 2. The proposed M-Health model

Information and communication technologies have led to the development of mobile systems, especially for emergency applications. Emergency applications are essential for severe patients suffering from cardiac diseases, serious brain injuries, accidents and spinal cord or other acute trauma. M-Health services are valuable, especially where there is a lack of medical health facilities, where hospitals are located far from groups of the population or where medical care is not affordable. M-Health systems are based on new and advanced wireless and wired technologies such as satellite communication, cloud bases systems, 4th and 5th Generation cellular systems and the Global System for Mobile Communications (GSM). Satellite links provide high-speed data transfer capabilities ranging from 2.4 kpps to $2 \times 64$ kbps and having a worldwide coverage [7]. These technologies are the best option for infrastructure-less areas or transmission across the sea. The GSM provides mobile communication services up to where 4G offers 100 Mbit/s for high mobility access to 1 Gbit/s for local wireless access. Wide Area Networks (WAN) provide flexible data communication using radio frequency technology. M-Health systems also have smart sensors, where 5G communication capabilities are integrated with Web 2.0, social networking platforms and cloud computing. Smart sensor-based systems like Wireless Body Area Networks (WBAN) also play a central role in M-Health systems where advanced wearable devices are used to provide monitoring and sensing features. The Internet of Things (IoT) is also used in association with M-Health systems, offering more flexible ecosystem capabilities [8,9]. These technologies provide unprecedented speed, high data rates, a better packet delivery, latency, throughput, and a maximum bandwidth.

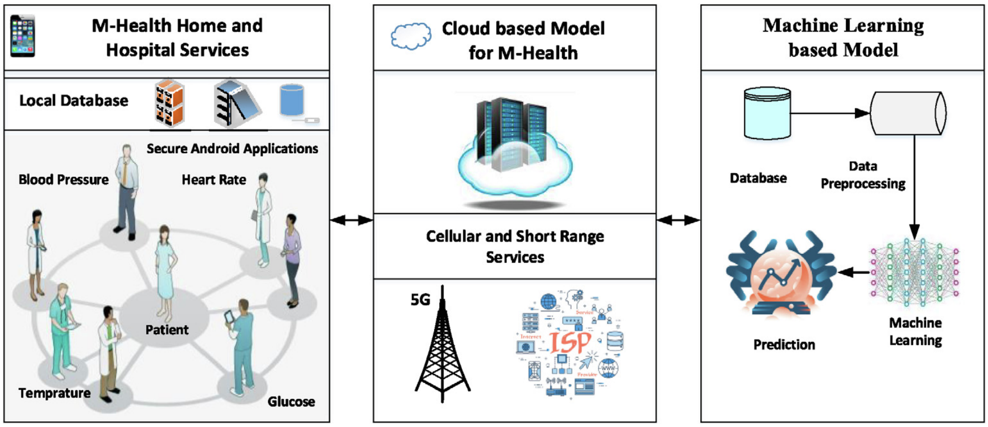


**Fig. 1.** The proposed M-Health Model.

New advanced technologies also offer a high data rate up to 10Gbps, less latency (milliseconds), better energy resources, more coverage and better connections [10]. Moreover, most of M-Health enable sensors and devices transmit the data to other devices or endpoints using low power Bluetooth and ZigBee technology. Subsequently, the data are transmitted to remote systems or cloud data storage via cellular or WANs for further reprocessing and decision making. To fulfil all the requirements in an M-Health system, we proposed a complete M-Health model with all the modules, as shown in Fig. 1. The first phase is data collection through smart sensor nodes that are planted on or inside the patient body. These sensor nodes are further connected with access points or gateway or mobile devices. Mobile devices communicate between patients and medical staff located in the home, office or anywhere. The entire cellular network can handle these mobile communications. Due to the development of open networks, this activity has encountered significant security challenges. We propose a secure Android-based application for the data collection and further forwarding to the proposed reliable cloud-based model. From this model, the data are collected and applied machine learning techniques are implemented for the feature extraction and a dynamic and accurate predictive model is introduced in relation to cardiovascular disease. Fig. 1 shows the compete M-Health model and its three modules.

NB “Temprature” should be “temperature”

### 2.1. The proposed M-Health home and hospital services model

New communication technologies have been transforming traditional healthcare services. Smart tiny wearable devices are installed inside or outside the patient’s body to monitor their vital signs and subsequently send the data to their mobile phones and their medical assistants’ mobile phones using 4G and 5G cellular technologies [5,11]. The power of these new technologies is improving healthcare services and saving human life by enabling a timely diagnosis and transmission of alerts to medical centres. The collection of sensor nodes is attached directly on the patient’s skin or inside his/her body and is able to collect the data and forward them to access points or other devices. These sensor nodes have low frequencies and operate between 402 and 405 MHz for data communication [12]. The data are stored in a local database or forwarded to a hospital or medical centre and at the same time saved in cloud storage. The sensors have a limited radio range from 2 to 5 m, with two to four networks per m2. Around 256 devices per network are possible for the communication. The data rate is approximately 10 kb/s to 10 Mb/s. Smart phones are also connected with sensor nodes to collect the data directly or through access points [13,14]. Fig. 2 presents the M-Health home and hospital model.

#### 2.1.1. Applications

M-Health applications are principally designed for smart devices to provide health-related services [15]. These applications are user-friendly and offer greater convenience, attracting the interest of patients and reducing healthcare costs. There are various types of applications which have been designed for medical purposes for patients and physicians. Physicians can access the patient data and analyze them for further treatment decisions. This framework, rich in applications, offers a great platform to patients as it has four significant dimensions, including understandable language, visual supports, user-friendly functions, and a two-way fast communication mechanism. Smart devices, such as smart phones, tablets, and smart watches, are used to play these applications. M-Health applications have been designed for brain and heart diseases, dementia and other disorders. In the table below, most of the applications presented have been developed for different conditions, such as cardiovascular disorders and dementia where the patient suffers from heart problems or deficits in memory, thinking skills and problem-solving capabilities [14,16] (see Table 1).

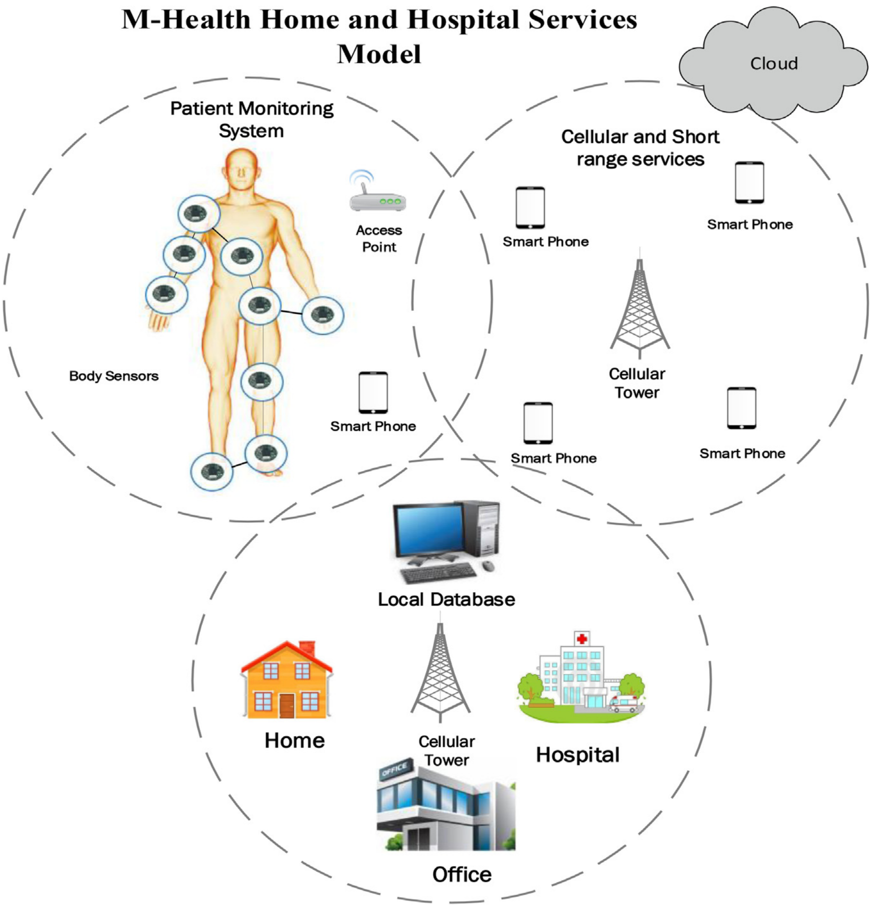


**Fig. 2.** The M-Health Home and Hospital Services Model.

*2.1.2. The proposed secure Android-based application architecture for an M-Health system*

The main objective in the design of a more secure Android-based application architecture for an M-Health system is to protect the patient's private data. The applications mentioned above are open and do not address the threat of a security attack. Android platform is a Linux-based operating system which offers a more flexible and user-friendly platform for mobile users. The proposed secure Android- based application architecture allows a better design of an M-Health application for any device. This model will help developers design new and integrated Android-based applications, as shown in Fig. 3.

Fig. 3 shows the proposed Android-based secure M-Health application architecture on receiving data from sensor nodes and forwarding them to cloud computing.

The proposed Android application has different phases which are explained as follows:

1. The User Interface: this is a user interface screen where the user interacts with medical centres and with other users by using a touch screen and keypad. This main screen also provides the link to the other pages and screens, such as those used for sending emails and notifications to the medical centre, and for storing previous records.
2. The Content Storage: This component provides a database platform where users can share and store the data in the file systems or forward the data to the web or any other storage device. This component supports standards including SQL-like queries INSERT, UPDATE and SELECT for data storage, retrieval and modification.
3. The Security Component: This component provides security by using a sandbox and permission framework which enforces the access control mechanism and other calls from the core functionalities. The permission system consists in a protected system using protected APIs for sensitive resources like GPS, Bluetooth, SMS, and camera. Additionally, the sandbox provides a unique user ID for each application. Native code also exists in the sandbox mechanism.
4. The User Authentication Interface: In this component, there is provision for user signing to ensure the application sandbox mechanism. Certificates are used to provide the user ID, which is also associated with different user IDs. When an application is installed, the system automatically verifies the application certification.

**Table 1**
M-Health applications.

| S# | M-Health Application | Features | Platform | Description | Limitations |
|---|---|---|---|---|---|
| 1 | Pain Rating Scales (2015) | To assess the pain | Android | This application is for health and safety monitoring, | No privacy and accuracy in this application. |
| 2 | Care Tracker (2015) | Tracking patients | Android | This is a navigation- based application for tracking purposes. | Localization issues. Inadequate functioning in indoor scenarios. |
| 3 | Dementia Risk Tool (2016) | Risk analysis tool to assess future risks. | Android & iOS | This application is for screening purposes | Provides privacy but is not particularly accurate in future risk analysis. |
| 4 | BarinScore (2016) | Designed to check the patient's attention skills | iOS | This is a memory training application. | Not very attractive in terms of gaining the patient's attention. |
| 5 | CogniCare-Enabled personalized dementia (2017) | Designed to monitor symptoms and communicate better with care workers. | Android | The main purpose of this application is to make connections between co-workers and to monitor the patient's symptoms. | Neglects various important features in the recording of symptoms. |
| 6 | Cognitive Rehab in Dementia (2017) | Provides cognitive, impairment and rehabilitation information. | Android & iOS | Educational purposes to guide a user with early stage dementia. based on cognitive rehabilitation | Not user friendly in that it is difficult to handle. |
| 7 | Dementia Respond (2018) | To obtain information in relation to dementia. | Android | This application provides carer support in order to provide the patient with information about dementia | Like the previous application,it lacks enhancing features. |
| 8 | Your Care Card (2018) | Provides connections among healthcare team members. | IOS | This application provides a carer support for a better understanding of a person with dementia. | Like the previous application,it lacks enhancing features |
| 9 | CareZone (2019) | Provides an easy way to manage medications and doctor instructions | Android | To facilitate patient treatment by managing difficult medication lists. | Difficult to manage and modify the medication dosages. |
| 10 | Fig. 1 – Medical Images (2019) | Provides connections among healthcare professionals for instant feedback. | Android | Used for diagnosis purposes and provides collaboration services among professionals. | The lack of any response from the professionals is a significant drawback of this application. |

5. The Data Communication: The inter-process communication mechanism is used for data communication which is based on the UNIX-style method. The inter-process communication also has a binder, services, and contents. The binder is used to ensure a high performance in or across process calls by using a remote procedure call mechanism.

   Fig. 4 shows the user interface of the M-Health application.

#### *2.1.3. Data handling*

Due to the large number of M-Health applications and their connection with IoT-based ecosystems, data collection is another great challenge. The clinical and non-clinical data sets are generated from M-Health areas such as diagnosis, treatment, human health, nutrition and accident applications. In the context of M-Health data collection, there are still different challenges in the handling of structured and unstructured data and the matching of features to motivate users and professionals to ensure better services. Furthermore, the new methods should be able to check the data accuracy, to sequence the data with other health-related data and to develop data analytics methods to be applied for early detection. The data collection is possible through smart M-Health sensing devices positioned inside and outside the patient and Internet-centred devices at medical centres. Some other traditional methods of data collection are medical and professional notes, home monitoring data repositories, medical robotics and WBAN and smart city data centres.

### *2.2. The proposed cloud based model for an M-Health system*

The proposed cloud-based model is based on two main layers including storage and access control layers and data annotation and analysis layers.

#### *2.2.1. The data storage and access control layers*

In the storage layer, all the M-Health data received from the smart mobile devices through internet service providers are stored. The data are related to the patient's' medical records, such as those for blood pressure, glucose level and heart rate. The cloud computing model is an attractive platform for users due to its pervasive data access capabilities. For M-Health data, the cloud computing solution is cost-effective, stable and efficient. The cloud-based model also helps in the integration and analysis of massive M-Health data and provides accuracy, prediction and decision making facilities [17,17]. Security is one of the important aspects, especially for M-Health data, where every patient needs privacy and security. To secure the data on cloud computing, we have adopted a multi-occupant data storage strategy. In this strategy, the data are protected

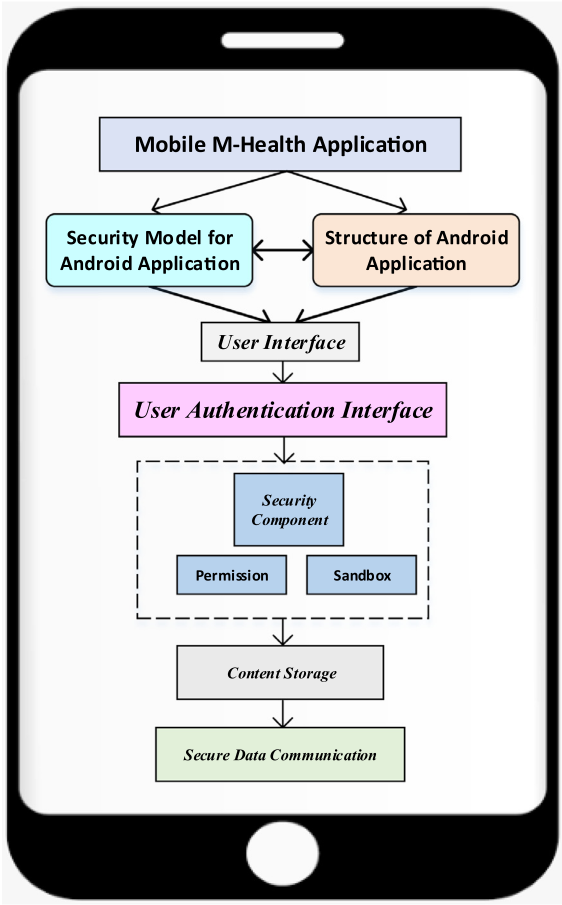


**Fig. 3.** The Android-based secure M-Health application architecture.

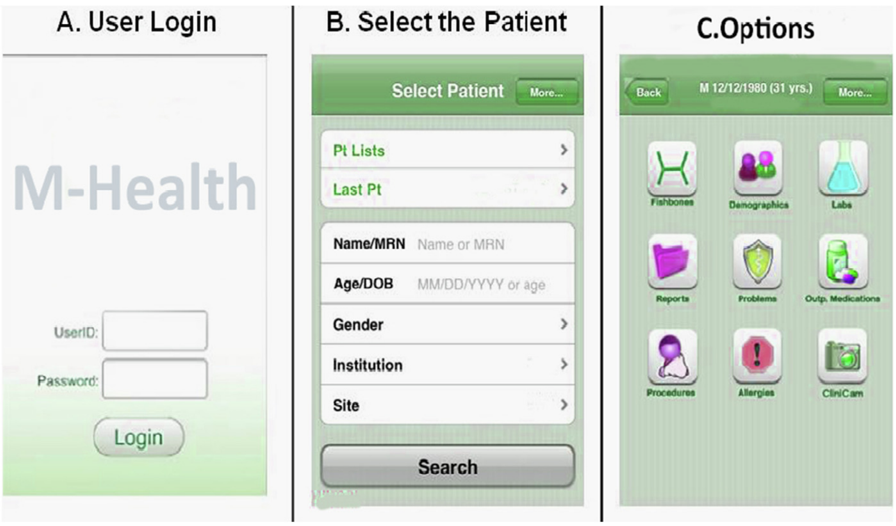


**Fig. 4.** The User Interface of the M-Health Application.

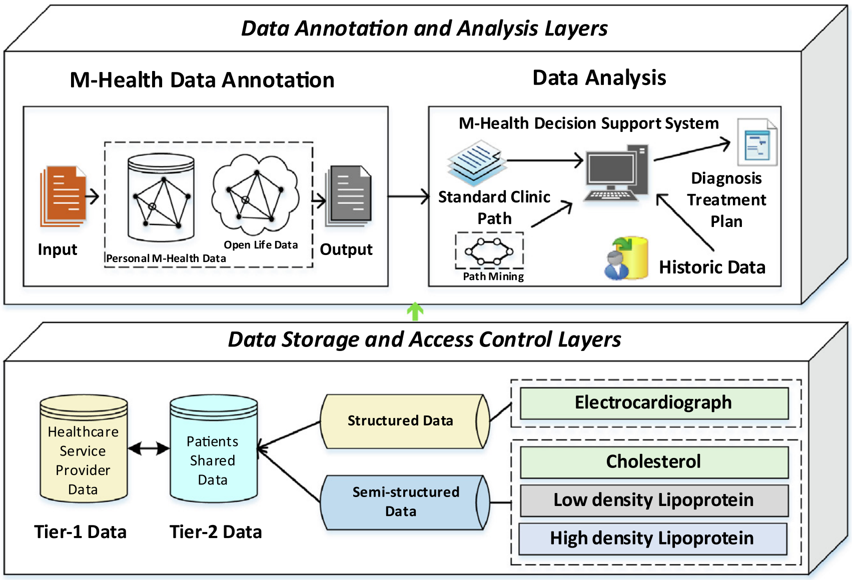


**Fig. 5.** The Cloud based Model for the M-Health Service NB "Clinic" should be "clinical" "Historic" should be "historical".

while providing storage spaces to medical centres, especially nursing homes, hospitals and research centres. The data isolation strategy is adopted for private and shared data by using data tables.

In the proposed cloud model, we have divided the received data into two main tiers where the first tier is assigned to healthcare service providers and the second to patients with a unique ID. In the first tier, isolated data are collected from a specific healthcare service provider and additionally from other healthcare service providers. Through isolated data, the security to access healthcare data is ensured. On the other hand, in the second tier, a shared table is used to store data where only authorized users can have access. By using this strategy, the patients can access their data quickly due to the shared tables, in contrast to isolated tables or data.

#### *2.2.2. The data annotation and analysis layers*

These layers are responsible for ensuring the data heterogeneity when using data from different medical service providers. Interoperability is one of the significant issues for medical service providers. Because of this limitation, there is a need to annotate the data semantically by using an ontology method where data are eliminated by data heterogeneity. The entire ontology model implementation is not a good option due to its complexity and large scale volume. In the proposed cloud-based M-Health model, we have adopted only a semantic search method to explore the data mining from the database as used in [19,19]. After annotation, data analysis is performed to discover the domain knowledge hidden in the historical data of patients related to diagnosis and treatment. A case searching algorithm is used to find any similarity among patients and then to compare the clinical path of similar patients [20]. If any differences are found, then a notification is sent for error correction. Fig. 5 shows the complete proposed secure cloud computing model and its processes.

#### *2.2.3. Big data handling*

Big data handling covers data redundancy, noise removal, and error correction. After this phase, the data are transmitted to the processing section where the data are processed as an offline and online data stream. The offline data are stored in the cloud for future analysis. In contrast, the real-time data are forwarded to the filtration and extraction section. Load balancing is also applied to the storage servers. Different distributed query methods have been adopted for query generation due to the complex nature of the data, Next, the data are sent for processing. The servers are used to process all the data intelligently, following which the data are forwarded to the next phase, where data analytics and data visualization methods are applied.

## 3. The proposed dynamic and accurate predictive model

Machine learning is one of the computational methods used to improve performance by mechanizing the acquisition of knowledge from experience. Machine learning is one of the most significant branches of Artificial Intelligence (AI). Machine learning also provides representation and generalizes the data by evaluating the instances [21]. Machine learning includes supervised and unsupervised methods. In the supervised methods, machine learning is deduced from labelled training data, which are composed of a set of training examples. Each training example has a supervised training dataset composed of a pair of input objectives, an input vector, and a preferred output value. Besides, the supervised methods also analyze the training data and produce an inferred function known as a classifier. Supervised method algorithms also generalize the training data to previously unobserved situations in a reasonable way.

On the other hand, the unsupervised machine learning method is used to find the hidden structure in unmarked data because the examples given to the learner are unlabelled. Unsupervised learning refers to density estimation in the area of statistics. The algorithms of unsupervised learning are present in Neural Networks (NN).

Supervised and unsupervised methods are different in terms of their structure. In the supervised type, one set of observations is the input, and other observations are the output. The output and input variables are at opposite ends of the casual chain. On the contrary, the unsupervised type is based on observations that are assumed by the latent variables. This type also leaves the probability that the inputs are undefined. Additionally, there is another type called semi-supervised machine learning which uses marked and unmarked data during the training period. In some cases, this type of learning is the optimal solution to solve a problem.

Reinforcement machine learning is another type that refers to those actions where an agent should take in an environment to capitalize on the idea of cumulative reward. This type has been studied primarily for control, games, information theories, operational research, statistics and genetic algorithms. Pattern recognition in machine learning is used to label a specific input. In the healthcare field, pattern recognition is well-known, where it needs to match the input values. Classification and prediction in machine learning are also used for classifying or prediction [22]. Clustering is used in machine learning to make groups from objects and to classify the object into different groups.

### *3.1. Predictive models*

The main objective of predictive models is to construct a model which is able to make predictions. Predictive models are based on combined outcomes and further connect with unique features. Predictive models are also based on machine learning techniques which are learned for certain variables from a training dataset to predict the results. Predictive models play a significant role, especially in the healthcare field, where we predict the patient's outcomes by using different biomarkers (features). Ding, et al. [23] discussed the ANN technique which is widely used with significant results. One of the best features of ANN is its wide acceptance and usage and its ability in mapping. The ensemble predictive model also handles a large number of systems [24]. Each copy shows the possible state from the real system. This technique uses multiple machine learning methods for a better predictive outcome. The machine learning ensemble is infinite and is used to consolidate the finite set and provide a flexible structure [25]. Bayes' law is another theorem to determine the probability of an event where the conditions are based on the event. This theorem is used with a probabilistic classifier with independent, strong assumptions [26]. The authors in [27] present the logistic regression predictive model, predictive ensemble mode, decision tree predictive model, ANN and support vector machine predictive model with fifteen datasets. All these models are used for prediction with different features and outcomes.

### *3.2. Machine learning methods*

Many machine learning methods have been designed for classification and prediction. The most commonly used machine learning methods are Artificial Neural Networks (ANN), Support Vector Machine (SVM), Naïve Bayes Method, Decision Trees (DT), Logistics Regression (LR) and K-Nearest Neighbour (K-NN) Algorithm. ANN is a computational model used for the connectivity of neurons to animate the nervous systems and is popular for prediction and classification. ANN has been used for the past thirty years and is considered to be an extremely significant method. One of the reasons behind this method's popularity is its mapping ability. It can analyze a group of inputs from the dataset to forecast the two inputs. An example is when a data set has a large collection of instances and each instance belongs to one or two classes. Next, SVM creates a model that allocates the instances in the form of points in a space that maps the instances into two classes. The classes are further analyzed by using their positions in terms of a gap. The Kerner trick is used to map the instances into high-dimensional space to provide non-linear prediction or classification [28].

NB, also known as the simple probabilistic method, is a method which assumes the absence or existence of any feature in a class which is independent and not related to other features. This method has been successfully adopted for many complex and real-world problems. It requires only a small amount of training data for categorization [29]. DT is another method for classification. It consists in dividing the input space to make the tree where the nodes are pure and contain points of a single class. This method starts from the first node and then moves to the root or terminal. The LR method is used to predict the probability of an occurrence by fitting the data to a logistics function. It addresses a discrete or categorical value. It is further divided into two types, including binomial and multinomial methods. The binomial method has only two states, true or false, whereas the multinomial method can have more than two states, such as cold, hot or normal states. The K-NN method is used to categorize objects based on the closest training. This is instance-based learning or lazy learning where the function is estimated locally, and the calculations are delayed until the prediction and classification. This is one of the simplest methods where the knowledge of its neighbours can categorize an entity.

### 3.3. The proposed predictive model for an M-Health system

New technologies have been transforming the trends of processes to make existing systems more efficient. For example, predictive models have been widely used to predict disease outcomes for future decisions [30]. In the medical field, the main objective of any predictive model is to provide accurate disease outcomes. Accurate prediction must also take into account any uncertain event in the identification of new disease outcomes from previous data. Prediction outcomes help professionals in the making of future decisions. Cardiovascular disease (CVD) is an extremely serious disease and causes death due to sudden heart attack, hypertension, and stroke. Early care and remote monitoring systems are the most effective means to treat the patient, using advanced scientific evidence-based guidelines. M-Health provides various remote monitoring applications for management and helps patients through self-reported outcomes and cost-effective solutions. Fig. 6 shows the proposed predictive dynamic and accurate model design phases and methodology in relation to CVD. Each step is discussed in detail in the next sub-sections.

#### 3.3.1. Data collection

For the proposed predictive model, the data are collected from clinical health records and physiological signals from wearable sensor nodes. The physiological signal data derive from electrocardiograms (ECG), accelerometers and Galvanic Skin Response (GSR) devices. The wearable sensor nodes monitor the patient data and, by means of a data acquisition microcontroller, transmit them to a mobile device. Some sensors, like an accelerometer, already exist in the mobile device. Existing data are extracted from cloud-based storage, where they provide a context for the physiological measurements. The cloud database contains all the patient-related information, such as medication history, age, body mass index, gender, ECG record,

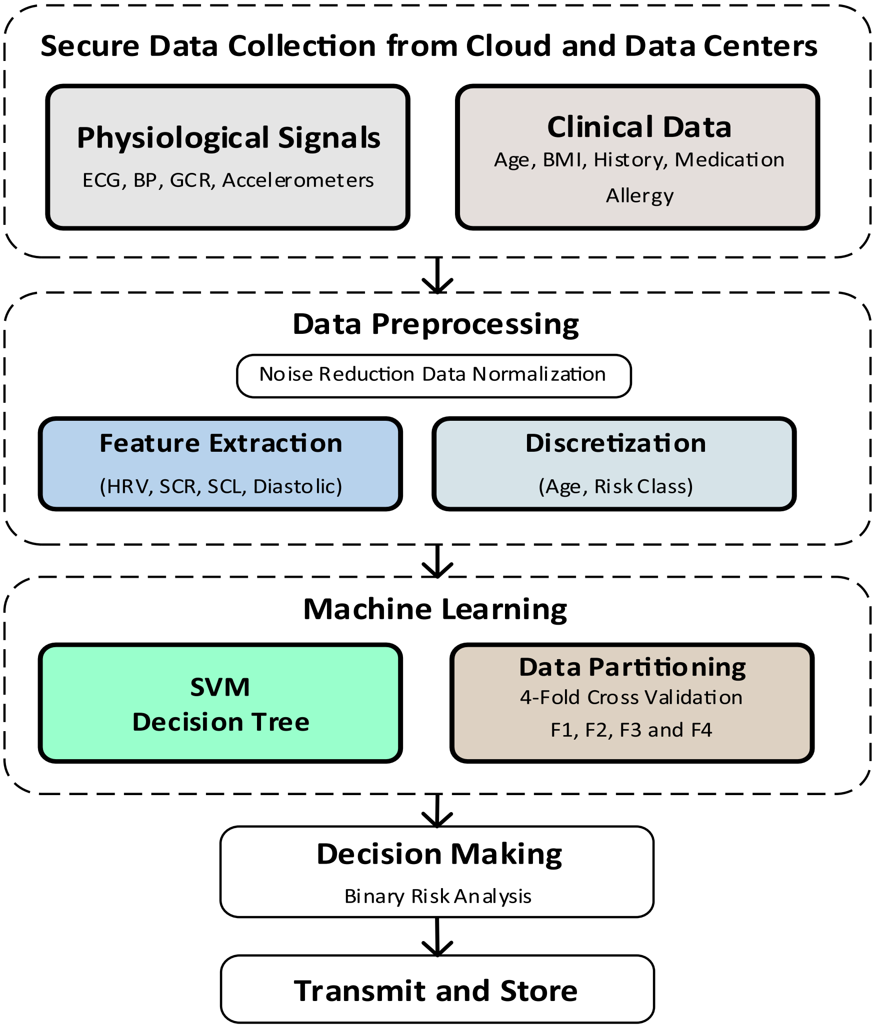


**Fig. 6.** The Proposed Predictive Dynamic and Accurate Model Design Phases and Methodology. NB "Centers" should be "centres". "Transmit and store" should be "Transmission and Storage".

SpO2, and BP. The clinical data is updated by medical centres. The database also contains a massive patient record and training set for SVM. Through cloud-based storage, mobile devices can rapidly measure the physiological events from the patient's history to check his/her risk of CVD.

### *3.3.2. Data preprocessing*

Data processing is a significant step, involving processing the data into a set of features before sending it to the SVM. In this stage, the data should be free from noise, a task realized by using noise reduction and data normalization methods. The data from sensor nodes are in a raw form, and noise and environmental factors are present. Patient movement is another factor that corrupts the physiological signal and presents an obstacle to feature extraction. Low and high pass filters are used to reduce the noise from the data. Feature extraction is another stage of the preprocessing phase, where it extracts features from the received data [31]. Data preprocessing is a method in which the raw data are subjected to a further process, the traditional data preprocessing. Normalization is useful to change the values of the numeric columns in the dataset. We have used a signal processing method to extract additional features and convert the signals to discrete values; for example, from the ECG signal, we found the heart rate. To handle the discrete features issue, we have normalized the features from the range of 0, 1 or 1, 2, 3, 4 to eliminate the bias. Fig. 7 shows the proposed predictive model.

### *3.3.3. Data partitioning*

The preprocessed data are further partitioned into two parts, where 75% of the data is used for training the model and 25% for testing it. The 4-fold cross-validation method is used to test the model. The training and testing are performed four times. The four folds are performed and denoted as Fold-1, Fold-2, Fold-3, and Fold-4. Fold-1 is reserved for testing purposes and the remaining folds are used for training the model. This study presents the CVD data, where we have attempted to refine the data and convert fewer data folds. The proposed predictive model is fast and has less complexity due to the refining of the data, as compared to the bulk data. Different types of learning models have been designed and developed for disease prediction, such as in [32–36]. However, the existing models have some limitations in terms of unsatisfactory results due to the unavailability of multi-class prediction in such a case where this class has many positive features.

### *3.3.4. Adopted machine learning algorithms*

In this stage, the machine learning algorithm is trained for the calculation and analysis of the patient risk. Two machine learning algorithms are applied to examine the classification of CVD, namely SVM and DT. The classification of the CVD dataset consists of three classes, High Risk, Medium Risk, and Low Risk. There are various machine learning methods which have been applied for prediction, such as NNs, Fuzzy Model, Ensemble Model and KNN. Every method has its advantages and limitations. In our opinion, SVM is one of the best options when we do not have any idea about the data and is optimal for

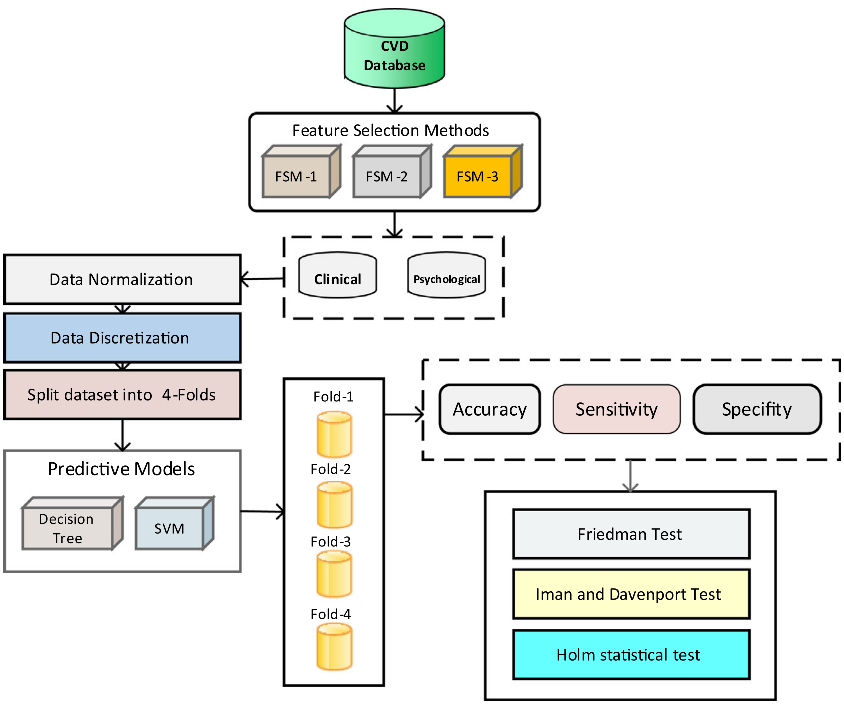


**Fig. 7.** The Proposed Predictive Model for CVD. NB "Specifity" should be "specificity".

unstructured and semi-structured data. It has a large scale in relation to high dimensional data. On the other hand, DT is another excellent option for possible decision outcomes. This method creates a comprehensive analysis of the consequences along each branch and identifies the decision nodes.

*3.3.4.1. Support Vector Machine.* SVM is a supervised learning model which is used to label one or two classes or strata. This model also trains the algorithm and classifies new samples into one or other category. This model presents the sample point in space and then arranges the samples into separate classes by using the hyper-plan. New samples are mapped into one of the classes [37]. SVM is used for the classification boundary to separate the input vectors into one or two classes. Patient risk is evaluated by training the machine learning algorithm. SVM creates a classification boundary to separate the vector ai and convert it into two classes bi € [0, 1]. SVM finds the c and d, denoting the error and classifier complexity. Structural Risk Minimization (SRM) is used for the trade-off. For the training phase, SVM is improved by using input vectors ai with a known class bi. For the testing phase, the data vectors and unknown classes are classified. Next, the well-known VC (Vapnik-Chervoenkis) theory is adopted to minimize the classifier margin and reduce the complexity of the classifiers. Furthermore, it minimizes the margin of the class boundary and then divides the hyper-surface with a bias b. Then, the Radial Basis Function (RBF) is adopted to train examples such as i is > than an (in the case of our data). SVM provides a mapping feature for high dimensional data by using a kernel function. It is the best choice for a small data set.

*3.3.4.2. Decision analysis.* For decision analysis, a DT is used as an analytical and pictorial tool in which the category label of alternatives is estimated. A DT is based on two types of shapes, ovals and rectangles. The oval shape represents the decision node and the rectangle represents the class label node. This tree also has decision rules in which the class label is linked with a decision node. Three conditions are applied for an outcome D1, D2, and D3. We have used the J48 algorithm for the DT, as used in Ref. [38]. The DT learning model is one of the best known models due to its tree strategy, where it assigns a specific value to each problem. This model also reduces any uncertainty and ambiguity among values [39]. The algorithm is based on supervised learning and works best where tmaximum information is available to the DT classifier. DT is a classification technique which is simple to apply and has a straightforward approach to solving the classification problem. In the healthcare sector, a series of test questions and conditions have been adopted to predict the onset of disease. The root and internal nodes contain attribute test conditions to separate the records which have different characteristics. After training the data, the DT is constructed and classifies a test record. The J48 classifier is used for the decision rule extraction from the tree and then applied to the proposed prediction model. The rules are extracted from human heuristics.

*3.3.4.3. KNN model.* K-Nearest Neighbor (KNN) is a simple and effective method for classification. This is a non-parametric classification approach where the data record is classified with its nearest neighbors. This classification is used without any consideration of distance-based weighting [40]. To apply KNN, we need to select an appropriate value for k and the success is based on this value. This method is case-based, where all the training data are kept for the classification. Although it has some benefits, this method is lazy and not useful for web mining for large repositories. It is more useful for text categorization to improve efficiency and classification accuracy [41].

*3.3.4.4. The Naïve Bayes model.* This method is another probabilistic classifier based on Bayes theorem. This is a robust naïve independent method due to its scalability to solve learning issues [42]. The Naïve Bayes classifiers are divided into different types including Multinomial Naïve Bayes, Bernoulli Naïve Bayes and Gaussian Naïve Bayes. The first types are used for document classification for feature prediction or for classifying the frequency of the words present in the document. The second type, Bernoulli, is used to predict the class variable with values yes or no. The Gaussian type is used for the prediction of a continuous value which is not discrete [43].

*3.3.5. Decision making*

The proposed model has been trained with two categories, at high (category 1) or low (category 0) risk levels. In this case, category 1 indicates a high risk and category 0 a low risk of CVD. Category 1 patients need attention and further clinical interventions. Category 0 patients also need further analysis to monitor future developments. The SVM model creates a high dimensional space. Those patients within the hypersurface fall into category 1 and those outside into category 0. These two categories have binary values, 0 and 1. The complete results are discussed in the next sections. All the results are stored in the remote server where they are reviewed by the healthcare professionals without any patient intervention.

## 4. Results and discussion

In this section, the complete results of the experiments with the SVM and DT learning models are discussed with a comparative analysis based on the classification of the CVD dataset extracted from the proposed secure cloud-based storage model. The results have been evaluated based on accuracy, sensitivity, and specificity.

First, the results of the proposed dynamic and predictive model using SVM, KNN and the Naïve Bayes Model were evaluated. We used a synthetic database collected from the cloud-based model with thirteen features.

**Table 2**
Selected Features for the Prediction Model.

| S# | Feature | Diagnosis range |
|---|---|---|
| *Clinical details* | | |
| 1 | Age | 30 to 80 years |
| 2 | Gender | Male/Female |
| 3 | BMI | Normal (24 kg/m$^2$), Overweight (25–30 kg/m$^2$), Obese (Range 1 – 30–40 kg/m$^2$, Range-2 >40 kg/m$^2$) |
| 4 | Smoking Habit | Yes/No |
| 5 | Hypertension or Stress | Yes/No |
| 6 | Exercise Habit | Yes/No |
| *Physiological features* | | |
| 7 | Systolic Blood Pressure | L < 120 mmHg, M (121–140 mmHg), H > 141 mmHg |
| 8 | Diastolic Blood Pressure | L < 120 mmHg, M (121–140 mmHg), H > 141 mmHg |
| 9 | Cholesterol | L < 5.2 mmol/L, H > 5.2 mmol/L |
| 10 | High Density Lipoproteins | L < 3.5 mmol/L, H > 3.5 mmol/L |
| 11 | Low Density Lipoproteins | L < 1 mmol/L, H > 1 mmol/L |
| 12 | R-R Interval | Normal 0.4–1.5 s, Abnormal > 1.5 s |
| 13 | Resting Heart Rate | Normal 60–100 100 beats/min, Abnormal > 100 beats/min |

The records of approximately 250 patients with CVD from a wide range of training sets for SVM were analyzed. Noise reduction and data normalization methods had already been applied to the database to avoid noise and missing values. The features were selected after a detailed investigation of CVD and consultations with professionals in this field. Risk factors were selected as features and 13 main features were identified for further investigation. All the features were based on discrete or vital signs of patients. According to the Canadian Heart and Stroke Foundation [44–48], every feature has a range and level in terms of increasing the risk factor for CVD. Cholesterol level is one of the indications of a higher risk of CVD, 5.2 mmol/L being a dangerous sign. SVM was trained and configured based on these features. Binary values were used in the database to evaluate and check the patients' vital signs levels, such as the cholesterol level, with a threshold value for a suitable SVM input type. Other risk factors like R-R and HR, collected from sensor nodes, have binary values and remain continuous.

Different types of risk factors were combined to train the SVM algorithm to achieve better results. All the features were populated through the MATLAB simulation via a random generator library. Age and BMI (Body Mass Index) are two main factors that are randomly distributed over physiological ranges (these two factors are not sufficient to predict that the patient has a high or low risk of CVD). We also distributed the feature in a database where every feature has a value based on specific conditions. During an examination of the patient record, every feature distribution increased due to the wide range of the SVM training sets of the patient record. Around 200 patient records were classified in the database in which around 155 were on a continued CVD risk and 45 were considered to be no longer at risk. Table 2 shows the features, which have been categorized into clinical details and physiological features.

In the proposed predictive model, we used 13 features. Two predictive models were incorporated to predict the outcomes of CVD and 4-fold cross-validation was used to validate the model. The predictive models were DT and SVM. These models were used on the CVD dataset extracted from the proposed cloud-based model. For the feature selection, we used Joint Mutual Information (JMI), Conditional Mutual Info Maximization (CMIM) and Double Input Symmetrical Relevance (DISR). The proposed predictive model was evaluated based on accuracy, sensitivity and specificity by using the confusion matrix. To verify the results, we used the Friedman statistical test [42,42].

In the first experiment, we used the proposed predictive model with state-of-the-art predictive models, including SVM and DT, with the CVD dataset to predict the disease outcome and then 4-fold cross-validation to validate the model.

### 4.1. Accuracy analysis

In the first experiment, we used the SVM predictive model and proposed a predictive model with a CVD dataset to predict the outcomes and then used for 4-fold for the validation.

The proposed predictive model, denoted with P5, was given a problem instance to be classified using a combination of multiples from the two most famous predictive models. Mathematically, the proposed predictive models can be formulated as follows:

$P_j$ = is the predictive model *(criteria)* Where $j = \{1, \cdots, p\}$ and $p$ is the number of predictive models.

P1: SVM

P2: Decision Tree

P3: KNN Model

P4: Naïve Bayes Model

The accuracy performance was calculated using Eq. (1).

$Acc_j$ denotes the accuracy of the predictive models and $P_j$ based on Eq. (1)

$Acc_j = \left(\sum_{i=1}^{n} \mathrm{TP}(i)\right) + \left(\sum_{i=1}^{n} \mathrm{TN}(i)\right) / \left(\sum_{i=1}^{n} (\mathrm{TP}(i) + \mathrm{FN}(i) + \mathrm{FP}(i) + \mathrm{TN}(i)\right)$ (1)where $n$ is the number of the outcomes and $1 \le i \le n$

The accuracy results are tabulated in Table 3.

**Table 3**
Accuracy of the proposed predictive model based on the CVD dataset.

| Method | Fold-1 | Fold-2 | Fold-3 | Fold-4 | Average |
|---|---|---|---|---|---|
| P1 (SVM) | 80.00 | 79.00 | 78.00 | 73.00 | 77.5 |
| P2 (Decision Tree) | 66.67 | 67.77 | 73.33 | 76.00 | 70.94 |
| P3 (KNN Model) | 71.32 | 73.25 | 77.45 | 82.00 | 76 |
| P4 (Naïve Bayes Model) | 75.10 | 76.23 | 77.40 | 77.70 | 76.60 |
| P5 (Proposed Model) | 83.67 | 86.77 | 87.67 | 88.78 | 86.72 |

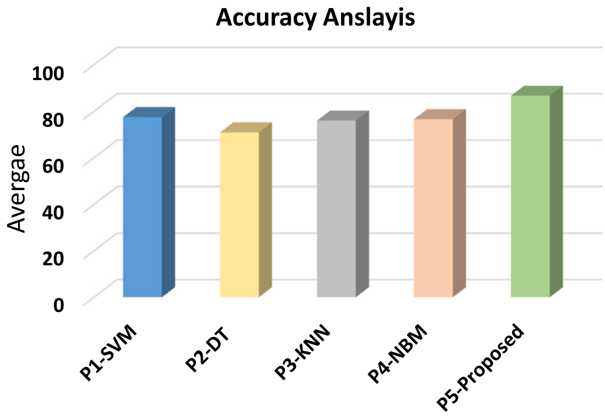


**Fig. 8.** Accuracy Analysis NB "ANSLAYIS" should be "Analysis".

The average ranking obtained by each predictive model in the Friedman test indicated that the proposed P5 predictive model has a significant difference with P1, P2, P3 and P4 based on accuracy. Additionally, this experiment rejected the null hypothesis where all the accuracy results are equal. On the other hand, the P2 predictive model has the worst average ranking among the predictive models. P3 and P4 are almost equal in average accuracy results. Fig. 8 shows the accuracy results.

### 4.2. Sensitivity analysis

In this experiment, we checked the sensitivity of the proposed model in comparison with the SVM, DT, KNN and Naïve Bayes predictive models. The CVD dataset was used to predict the outcomes and then the 4-fold was used. for the validation, For the sensitivity calculation, we used the predictive models based on Eq. (2).

$$Sens_j = \left(\sum\nolimits_{i=1}^{n} \mathrm{TP}(i)\right) / \left(\sum\nolimits_{i=1}^{n} (\mathrm{TP}(i) + \mathrm{FN}(i))\right) \tag{2}$$

Table 4 shows the sensitivity comparison of P1 (SVM), P2 (DT), P3 (KNN) and P4 (Naïve Bayes Model) with the proposed prediction model P5.

The sensitivity test indicated that, compared with the sensitivity of the predictive performance of the existing predictive models, the proposed predictive model is superior. P5 achieved the best average according to the Friedman statistics compared with the P1, P2, P3 and P4 models. In this test, P3 obtained the worst average, KNN being a lazy method with a low efficiency. Fig. 9 shows the sensitivity results.

### 4.3. Specificity analysis

In this experiment, we checked the specificity of the proposed model in comparison with the SVM, DT, KNN and Naïve Bayes predictive models. The CVD dataset was used to predict the outcomes and then the 4-fold was used for the validation. For the specificity calculation, we used predictive models based on Eq. (3) (see Table 5).

$Spec_j$=$(\sum_{i=1}^{n} \mathrm{TN}(i))/(\sum_{i=1}^{n} (\mathrm{FP}(i) + \mathrm{TN}(i)))$ (3)

The specificity test indicated that the specificity of the predictive performance of the proposed predictive model is superior to that of the existing predictive models. P5 achieved the best average according to the Friedman statistics compared with the P1, P2, P3 and P4 models. In this test, P3 obtained the worst average, KNN being a lazy method with a low efficiency. Fig. 10 shows the sensitivity results (Fig. 11).

The confusion matrix, also called the error matrix, is used in machine learning to solve the statistical classification. In this case, it was used to evaluate the predictive model performance, as shown in Table 6. True Positives, True Negatives, False Positives and False Negatives were used for the class instances. For a correctly classified instance, the term True Positive

**Table 4**
Sensitivity of the proposed predictive model based on the CVD dataset.

| Method | Fold-1 | Fold-2 | Fold-3 | Fold-4 | Average |
|---|---|---|---|---|---|
| P1 (SVM) | 0.67 | 0.67 | 0.67 | 0.50 | 0.62 |
| P2 (Decision Tree) | 0.58 | 0.50 | 0.67 | 0.50 | 0.56 |
| P3 (KNN Model) | 0.56 | 0.49 | 0.57 | 0.60 | 0.55 |
| P4 (Naïve Bayes Model) | 0.62 | 0.63 | 0.70 | 0.54 | 0.622 |
| P3 (Proposed Model) | 0.67 | 0.67 | 0.76 | 0.58 | 0.67 |

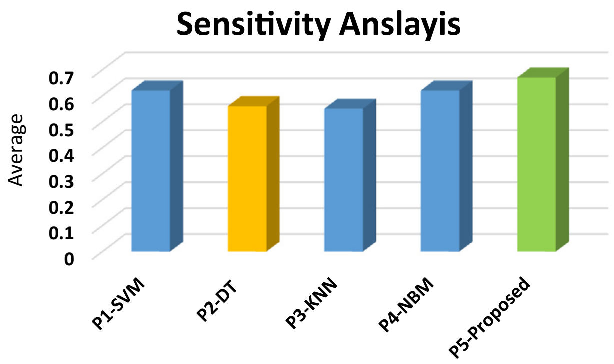


**Fig. 9.** Sensitivity Analysis NB “Anslayis” should be “Analysis”.

**Table 5**
Specificity of the proposed predictive model based on the CVD dataset.

| Method | Fold-1 | Fold-2 | Fold-3 | Fold-4 | Average |
|---|---|---|---|---|---|
| P1 (SVM) | 0.89 | 0.89 | 0.89 | 0.80 | 0.86 |
| P2 (Decision Tree) | 0.85 | 0.80 | 0.89 | 0.80 | 0.83 |
| P3 (KNN Model) | 0.81 | 0.78 | 0.87 | 0.78 | 0.81 |
| P4 (Naïve Bayes Model) | 0.83 | 0.81 | 0.89 | 0.80 | 0.83 |
| P5 (Proposed Model) | 0.89 | 0.89 | 0.93 | 0.85 | 0.89 |

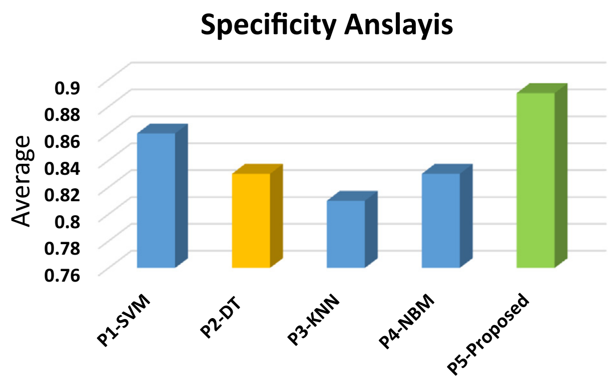


**Fig. 10.** Specificity Analysis NB “Anslayis” should be “Analysis”.

is used. True Negative is used where “No” class instances are correctly classified. False Positive is used where “No” class instances are incorrectly classified. False Negative is used for “Yes” class instances which are incorrectly classified.

Table 7 shows the confusion Matrix where the sensitivity refers to True Positives and “Yes” is the instance which is correctly identified. It is defined as: *Sensitivity = True Positive/Positive.*

Specificity refers to True Negatives with “No” instances identified. It is defined as Specificity = True Positive/Negative.

Table 7 shows the results for accuracy, sensitivity and specificity with the SVM, DT, KNN and Naïve Bayes models. The proposed predictive model has better results compared to the SVM and DT models.

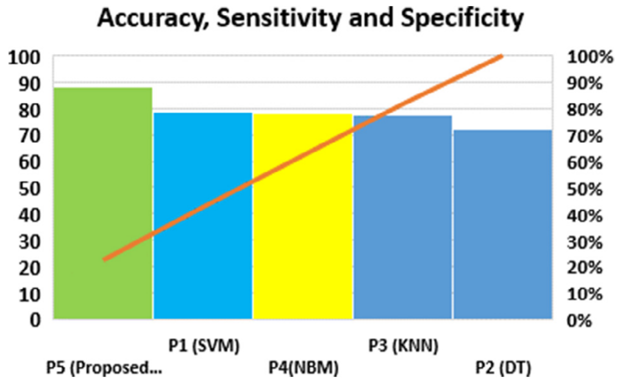


**Fig. 11.** Combined Performance of the Predictive Models.

**Table 6**
Comparative analysis of the methods.

| Methods | Comparative factors | | |
|---|---|---|---|
| | Accuracy | Sensitivity | Specificity |
| P1 (SVM) | 77.5 | 0.62 | 0.86 |
| P2 (Decision Tree) | 70.94 | 0.56 | 0.83 |
| P3 (KNN Model) | 76 | 0.55 | 0.81 |
| P4 (Naïve Bayes Model) | 76.60 | 0.62 | 0.83 |
| P5 (Proposed Model) | 86.72 | 0.67 | 0.89 |

**Table 7**
Confusion Matrix.

| | Yes | No | Total |
|---|---|---|---|
| Yes | True Positives | False Negatives | P |
| No | False Positives | True Negatives | N |
| Total | P' | N' | P + N |

**Table 8**
Combined performance of the proposed predictive model (P5) and benchmarks based on the Friedman test in terms of accuracy, sensitivity and specificity using the CVD dataset.

| Comparison | Accuracy | Sensitivity | Specificity | Benchmarks |
|---|---|---|---|---|
| P5 vs Benchmarks | P5 is better than the benchmarks based on the Friedman test in terms of accuracy with the CVD datasets considering a level of significance ($\alpha$) < 0.05 and ($\alpha$) < 0.01 | P5 is better than the benchmarks based on the Friedman test in terms of sensitivity with multiple datasets considering a level of significance ($\alpha$) < 0.05 and ($\alpha$) < 0.01 | P5 is better than the benchmarks based on the Friedman test in terms of specificity with multiple datasets considering a level of significance ($\alpha$) < 0.05 and ($\alpha$) < 0.01 | SVM (P1), Decision Tree (P2), KNN (P3) Naïve Bayes (P4) |

Fig. 10 shows the comparison in terms of accuracy, sensitivity and specificity using the four-fold dataset (Fold-1, Fold-2, Fold-3 and Fold-4) cross validation.

The comparison analysis shows that proposed predictive model is better than the SVM, DT, KNN and Naïve Bayes Models.

### 4.4. Discussion

The experiments are based on the SVM, DT, KNN and Naïve Bayes predictive models using the CVD dataset in terms of accuracy, sensitivity and specificity metrics. The CVD dataset was extracted from the proposed cloud-based models and then applied methods were used to extract the 13 features. The findings of this study can be summarized as follows:

The proposed predictive model (P5) achieved the best average and can be considered clearly the best predictive model among the whole multiple models in terms of accuracy, sensitivity, and specificity based on the Friedman test, as shown in Table 8.

The proposed predictive model (P3) significantly outperforms the benchmark predictive models in terms of accuracy, sensitivity and specificity with a level of significance ($\alpha$) < 0.05 and ($\alpha$) < 0.01 based on the CVD datasets.

The results indicate that the proposed predictive model achieves significant results and the best average ranking in term of specificity, accuracy and sensitivity. The proposed predictive model could be useful for healthcare systems in the prediction of CVD onset. Moreover, the predictive model has also been evaluated in comparison with existing models where it performs better in terms of analyzing patient features from a dataset.

## 5. Conclusion

New and smart technologies are able to offer advanced processes for healthcare services. M-Health systems are also being integrated with such technologies to monitor, process and store patient data. During the collection of patient vital signs data, security is a fundamental aspect, especially in relation to open systems. This paper proposes a complete M-Health model integrated with a secure application, secure cloud modes and machine learning-based predictive analysis for further diagnosis. The data are collected from sensor nodes and forwarded to local databases through new technologies that enable cellular networks and then store the data in cloud storage services. From cloud computing or medical centres, the data are collected for further analysis.

Moreover, machine learning techniques are used for their accuracy in the prediction of disease analysis and have also been adopted for classification. The proposed machine learning model seems to be an excellent option for CVD prediction, based on an analysis of patient features from a dataset. In a comparative evaluation. it outperformed existing state-of-the-art models in terms of accuracy, sensitivity and specificity. In the future, we aim to test the proposed predictive model with other predictive models and apply a 10-Fold validation. Furthermore, we will conduct the study with other diseases like Traumatic Brain Injury (TBI), which is another serious disease worldwide in terms of its impact on human life.

## CRediT authorship contribution statement

**Kashif Naseer Qureshi:** Conceptualization, Methodology. **Sadia Din:** Data curation, Writing - original draft. **Gwanggil Jeon:** Software, Validation. **Francesco Piccialli:** Writing - review & editing.

## Acknowledgements

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government (No. 2018045330).

## References

[1] K.M. Awan, A. Ali F. Aadil, K.N. Qureshi, Energy efficient cluster based routing algorithm for wireless sensors networks, in: 2018 International Conference on Advancements in Computational Sciences (ICACS), IEEE, 2018, pp. 1–7.
[2] H. Alanazi, A. Abdullah, K. Qureshi, A. Ismail, Accurate and dynamic predictive model for better prediction in medicine and healthcare, Irish Journal of Medical Science (1971) 187 (2) (2018) 501–513.
[3] H.O. Alanazi, A.H. Abdullah, K.N. Qureshi, M. Larbani, M. Al Jumah, Predicting the outcomes of traumatic brain injury using accurate and dynamic predictive model, Journal of Theoretical Applied Information Technology, 93 (2) (2016).
[4] K.N. Qureshi, A.H. Abdullah, O. Kaiwartya, S. Iqbal, R.A. Butt, F. Bashir, A dynamic congestion control scheme for safety applications in vehicular ad hoc networks, Computers Electrical Engineering 72 (2018) 774–788.
[5] M. Anwar et al, Green communication for wireless body area networks: energy aware link efficient routing approach, Sensors 18 (10) (2018) 3237.
[6] Y. Zia, A. Farhad, F. Bashir, K.N. Qureshi, G. Ahmed, Content-based dynamic superframe adaptation for internet of medical things, International Journal of Distributed Sensor Networks, 16 (2) (2020) 1550147720907032.
[7] B.C. Helsel, J.E. Williams, K. Lawson, J. Liang, J. Markowitz, Telemedicine and mobile health technology are effective in the management of digestive diseases: A systematic review, Digestive Diseases Sciences 63 (6) (2018) 1392–1408.
[8] S.S. Sadegh, P.K. Saadat, M.M. Sepehri, V. Assadi, A framework for m-health service development and success evaluation, International Journal of Medical Informatics 112 (2018) 123–130.
[9] K.N. Qureshi, M.M. Idrees, J. Lloret, I. Bosch, Self-assessment based clustering data dissemination for sparse and dense traffic conditions for internet of vehicles, IEEE Access 8 (2020) 10363–10372.
[10] A. Osseiran, J.F. Monserrat, P. Marsch, 5G Mobile and Wireless Communications Technology, Cambridge University Press, 2016.
[11] Y. Qu, G. Zheng, H. Ma, X. Wang, B. Ji, H. Wu, A survey of routing protocols in WBAN for healthcare applications, Sensors 19 (7) (2019) 1638.
[12] K. Awan, K.N. Qureshi, M. Mehwish, Wireless body area networks routing protocols: a review, Indonesian Journal of Electrical Engineering Computer Science, 4 (2016).
[13] B. Malik, V. Singh, A survey of research in WBAN for biomedical and scientific applications, Health Technology 3 (3) (2013) 227–235.
[14] R. Gk, K. Baskaran, A survey on futuristic health care system: WBANs, Procedia Engineering 30 (2012) 889–896.
[15] M. Hussain et al, Conceptual framework for the security of mobile health applications on android platform, Telematics Informatics 35 (5) (2018) 1335–1354.
[16] A. Bahga, V.K. Madisetti, Healthcare data integration and informatics in the cloud, Computer 48 (2) (2015) 50–57.
[17] S. Garg, K. Kaur, S. Batra, G. Kaddoum, N. Kumar, A. Boukerche, A multi-stage anomaly detection scheme for augmenting the security in IoT-enabled applications, Future Generation Computer Systems 104 (2020) 105–118.
[18] H. Panetto, M. Dassisti, A. Tursi, ONTO-PDM: Product-driven ONTOlogy for Product Data Management interoperability within manufacturing process environment, Advanced Engineering Informatics 26 (2) (2012) 334–348.
[19] Y. Zhang, M. Qiu, C.-W. Tsai, M.M. Hassan, A. Alamri, Health-CPS: Healthcare cyber-physical system assisted by cloud and big data, IEEE Systems Journal 11 (1) (2015) 88–95.

[20] A.T. Lo'ai, R. Mehmood, E. Benkhlifa, H. Song, Mobile cloud computing model and big data analysis for healthcare applications, EEE Access 4 (2016) 6171–6180.
[21] K.P. Ferentinos, Deep learning models for plant disease detection and diagnosis, Computers Electronics in Agriculture 145 (2018) 311–318.
[22] M. Juhola, J. Laurikkala, Missing values: how many can they be to preserve classification reliability?, Artificial Intelligence Review 40 (3) (2013) 231–245.
[23] S. Ding, H. Li, C. Su, J. Yu, F. Jin, Evolutionary artificial neural networks: a review, Artificial Intelligence Review 39 (3) (2013) 251–260.
[24] L. Rokach, Ensemble-based classifiers, Artificial Intelligence Review 33 (1–2) (2010) 1–39.
[25] X. Li, L. Wang, E. Sung, AdaBoost with SVM-based component classifiers, Engineering Applications of Artificial Intelligence 21 (5) (2008) 785–795.
[26] J. Alwedyan, W.e.M. Hadi, M.a. Salam, H.Y. Mansour, Categorize arabic data sets using multi-class classification based on association rule approach, in: Proceedings of the 2011 International Conference on Intelligent Semantic Web-Services and Applications, 2011, pp. 1–8.
[27] T. van der Ploeg, D. Nieboer, E.W. Steyerberg, Modern modeling techniques had limited external validity in predicting mortality from traumatic brain injury, Journal of Clinical Epidemiology 78 (2016) 83–89.
[28] K.Q. Weinberger, F. Sha, L.K. Saul, Learning a kernel matrix for nonlinear dimensionality reduction, in: Proceedings of the Twenty-First International Conference on Machine Learning, ACM, 2004, p. 106.
[29] P. Bhargavi, S. Jyothi, Applying naive bayes data mining technique for classification of agricultural land soils, International Journal of Computer Science Network Security 9 (8) (2009) 117–122.
[30] N. Stephenson et al, Survey of machine learning techniques in drug discovery, Current Drug Metabolism 20 (3) (2019) 185–193.
[31] H. Banaee, M.U. Ahmed, A. Loutfi, Data mining for wearable sensors in health monitoring systems: a review of recent trends and challenges, Sensors 13 (12) (2013) 17472–17500.
[32] M. Chen, Y. Hao, K. Hwang, L. Wang, L. Wang, Disease prediction by machine learning over big data from healthcare communities, IEEE Access 5 (2017) 8869–8879.
[33] R. Kaundal, A.S. Kapoor, G.P. Raghava, Machine learning techniques in disease forecasting: a case study on rice blast prediction, BMC Bioinformatics 7 (1) (2006) 485.
[34] M. Motwani et al, Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis, European Heart Journal 38 (7) (2016) 500–507.
[35] A. Singh, B. Ganapathysubramanian, A.K. Singh, S. Sarkar, Machine learning for high-throughput stress phenotyping in plants, Trends in Plant Science 21 (2) (2016) 110–124.
[36] S. Garg, A. Singh, S. Batra, N. Kumar, M.S. Obaidat, EnClass: Ensemble-based classification model for network anomaly detection in massive datasets, in: GLOBECOM 2017–2017 IEEE Global Communications Conference, IEEE, 2017, pp. 1–7.
[37] M.F.B. Othman, N.B. Abdullah, N.F.B. Kamal, MRI brain classification using support vector machine, in: 2011 Fourth International Conference on Modeling, Simulation and Applied Optimization, 2011, IEEE, pp. 1–4.
[38] G. Nuti, L. A. J. Rugama, A.-I. Cross, A Bayesian Decision Tree Algorithm, stat, vol. 1050, 2019, p. 11.
[39] R.D.H. Devi, M.I. Devi, Outlier detection algorithm combined with decision tree classifier for early diagnosis of breast cancer, International Journal of Advanced Engineering Technology/Vol VII/Issue II/April–June 93 (2016) 98.
[40] Q. Liu, C. Liu, A, novel locally linear KNN model for visual recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1329–1337.
[41] G. Guo, H. Wang, D. Bell, Y. Bi, K. Greer, KNN model-based approach in classification, in: "OTM Confederated International Conferences" On the Move to Meaningful Internet Systems, Springer, 2003, pp. 986-996.
[42] L. Jiang, H. Zhang, Z. Cai, A novel Bayes model: Hidden naive Bayes, IEEE Transactions on knowledge and data engineering 21 (10) (2008) 1361–1371.
[43] K.P. Murphy, Naive bayes classifiers, University of British Columbia 18 (2006) 60.
[44] T. C. H. a. S. Foundation, Heart disease recovery road. August 2013, Available: http://www.heartandstroke.com.
[45] K.N. Qureshi, S. Din, G. Jeon, F. Piccialli, Link quality and energy utilization based preferable next hop selection routing for wireless body area networks, Computer Communications 149 (2020) 382–392.
[46] K. Qureshi, R. Hussain, G. Jeon, A distributed software defined networking model to improve the scalability and quality of services for flexible green energy internet for smart grid systems, Computers and Electrical Engineering 84 (2020) 106634.
[47] K.N. Qureshi, A. Ahmad, F. Piccialli, G. Casolla, G. Jeon, Nature inspired algorithms based secure data dissemination framework for smart cities networks, Online published in Neural Computing and Applications (April 10, 2020).
[48] S. Iqbal, K.N. Qureshi, N. Kanwal, G. Jeon, Collaborative energy efficient zone based routing protocol for multi-hop internet of things, Online published in Emerging Telecommunications Technologies (ISSN 2161-3915).