RESEARCH ARTICLE

MedAi: A Smartwatch-Based Application Framework for the Prediction of Common Diseases Using Machine Learning

SHINTHI TASNIM HIMI[ID][1]**, NATASHA TANZILA MONALISA**[ID][1]**,**
MD WHAIDUZZAMAN[ID][2]**, (Senior Member, IEEE), ALISTAIR BARROS**[2]**,**
AND MOHAMMAD SHORIF UDDIN[1]**, (Senior Member, IEEE)**
[1]Department of Computer Science and Engineering, Jahangirnagar University, Savar, Dhaka-1342, Bangladesh
[2]School of Information Systems, Queensland University of Technology, Brisbane City, QLD 4000, Australia
Corresponding author: Md Whaiduzzaman (wzaman@juniv.edu)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Biosafety, Biosecurity and Ethical Committee (BBEC), Jahangirnagar University, under Reference No. -BBEC/JU/M 2022/01(01).

ABSTRACT Health information technology is one of today's fastest-growing and most powerful technologies. This technology is used predominantly for predicting illness and obtaining medications quickly because visiting a doctor and performing pathological tests can be time-consuming and expensive. This has prompted many researchers to contribute by developing new disease prediction systems or improving existing ones. This paper presents a smartwatch-based prediction system named 'MedAi' for multiple diseases such as ischemic heart disease, hypertension, respiratory disease, hyperthyroidism, hypothyroidism, stroke, myocardial infarction, kidney failure, gallstones, diabetes, dyslipidemia using machine learning algorithms. It comprises three core modules: a prototype smartwatch 'Sense O'Clock' equipped with eleven sensors to collect bodily statistics, a machine learning model to analyze the data and make a prediction, and a mobile application to display the prediction result. A dataset consisting of patient bodily statistics was obtained from a local hospital according to ethical guidelines, such as obtaining the prior consent of both patients and doctors. We employ several machine learning algorithms, including Support Vector Machine (SVM), Support Vector Regression (SVR), K-Nearest Neighbor (KNN), Extreme Gradient Boosting (XGBoost), Long Short Term Memory (LSTM), and Random Forest (RF) to investigate the best performing algorithm. Experimentation using our dataset shows that the RF algorithm outperforms other machine learning algorithms such as SVM, KNN, XGBoost, etc., in predicting aforementioned diseases with an accuracy of 99.4%. The system provides full-time assistance to the user by reporting his or her body condition and suggesting requisite remedies. It is a notable addition to early disease prediction systems and can predict multiple disease vulnerabilities before they reach an irrecoverable stage. Finally, we compare our method with the related existing methods.

INDEX TERMS Healthcare, disease prediction, machine learning, smartwatch, mobile application.

I. INTRODUCTION

Medical professionals have worked continuously for decades to find treatments for deadly human diseases, but only

The associate editor coordinating the review of this manuscript and approving it for publication was Vivek Kumar Sehgal[ID].

approximately 500 prescribed treatments have been discovered for 10,000 diseases. However, estimates on the number of diseases have reached 30,000, according to a German government listing [1]. Some prominent diseases cause an alarming number of deaths worldwide every year. The IHME (Institute for Health Metrics and Evaluation) compared the

statistics on the global burden of diseases of the world's top 20 causes of death from 1990 to 2018. This report notes that the diseases that cause the most deaths are cardiovascular diseases, cancers, and respiratory diseases. Diabetes and kidney diseases are also on this list, and their prevalence is increasing [2]. People are spending massive amounts on healthcare to avoid death, affecting their expenditure on other essentials. The OECD (Organisation for Economic Co-operation and Development) states that the United States spends the most on healthcare per person (US$ 10,586). Switzerland (US$ 7,317), Norway (US$ 6,187), Germany (US$ 5,986), Sweden (US$ 5,447), and Austria (US$ 5,395) are also concerned about the health sector. In Asia, however, the amount of spending is less; India spends only US$ 257 per capita [3]. Other countries in Asia, such as Bangladesh, Pakistan, Nepal, and Bhutan are not capable of providing proper health coverage to their citizens. As a result, diseases such as malaria, dengue fever, Japanese encephalitis, chikungunya fever, typhus, and BurntOrange fever are causing concern in these regions [4]. Lack of money, knowledge, and time, as well as ignorance of body conditions, are deteriorating human health.

In the modern era, many solutions are related to technology. The same is true for healthcare and medication. According to 2022 statistics, downloads of health and fitness mobile applications have increased from 488 million in 2019 to 656 million in recent years [5]. As there are 5.31 billion unique mobile users worldwide, using mobile applications for medical care seems to be an efficient option [6]. It can save time and money and reduce the percentage of health issues and sudden deaths. Today, doctors and IT developers are helping improve the health conditions of residents of developed and underdeveloped countries. Machine learning, deep learning, the Internet of Things (IoT), Android applications, web development, electrical science, and other sectors of technology are making systems and machinery so that people can get health updates instantly and be informed of any bodily dysfunctions. Even the concern regarding data theft at the cost of these services has also been alleviated to a great extent by the use of several encryption mechanisms e.g. blockchain security mechanism is now being used in different types of networking, including fog-IoT [7].

The sudden deaths of those close to us are painful and unexpected. Undiagnosed illnesses such as advanced cancer; abrupt natural causes such as heart attack, brain hemorrhage, and cot death; and major illnesses such as epilepsy are some of the leading causes of such deaths [8]. To prevent these fatalities, routine checkups for healthy people are typically recommended once a year for people over the age of 50 and once every three years for those under 50. Someone diagnosed with a chronic disease or other ongoing health problems must see a doctor more often, regardless of age [9]. However, study or work pressure, family responsibilities, and sometimes even laziness hinder regular checkups. This problem can be solved by finding a system that is convenient and performs checkups in real-time. The rapid growth in intelligent wearables (predicted to reach more than one billion by 2022 [10]) has made this possible. However, not everyone can buy branded smartwatches and other wearables. Moreover, no wearables have just a few sensors that can be highly accurate in detecting multiple diseases. As a result, individuals greatly need affordable smart wearables that will include the majority of sensors needed to monitor bodily statistics and save them from sudden tragedy. Android health apps, frequently linked to such wearables, are another way for individuals to live a healthy lifestyle. These apps may show patients how healthy their bodies are, teach them how to manage their illnesses, and even anticipate sickness.

Furthermore, disease prediction applications currently use advanced technologies. In [11], Ramya developed a disease prediction system using the fuzzy c-means algorithm; in [12], Wang et al. detected disease using deep learning; in [13], Jackins et al. used a random forest classifier and naive Bayes for disease prediction. Therefore, machine learning has become a vast field of research to which many researchers are constantly contributing. The majority of these works concern only early heart disease prediction since ischemic heart disease (IHD) has the greatest death rate, according to the WHO [14]. However, this does not negate the possibility of an individual having risks for other diseases, the prediction of which could save lives. As a result, finding a system for predicting several of these diseases has become crucial.

The major contributions of the proposed system are as follows:

- We design a complete package smartwatch "Sense O'Clock" equipped with eleven sensors for health monitoring with optimal power.
- We select an optimum algorithm from among several widely used machine learning approaches.
- We propose an integrated application framework based on mobile applications and machine learning algorithms to notify the user of his or her health status.
- We present a machine learning model that provides noticeably higher accuracy (by at least 3-4%) compared to the existing works on common disease prediction, which predicts only a few common diseases.

The rest of this paper is divided into the following sections:

Section II presents a literature review that helps outline the previous studies' contributions and drawbacks, which helps improve our system. Section III outlines the system methodology by describing the work of each system component, such as the design of the prototype watch, development of the machine learning model, and deployment of the model as a mobile application. Section IV evaluates the result and discusses the findings. Section V concludes the paper by stating the actual contributions and discoveries made throughout the research and finishes the study by noting some future work.
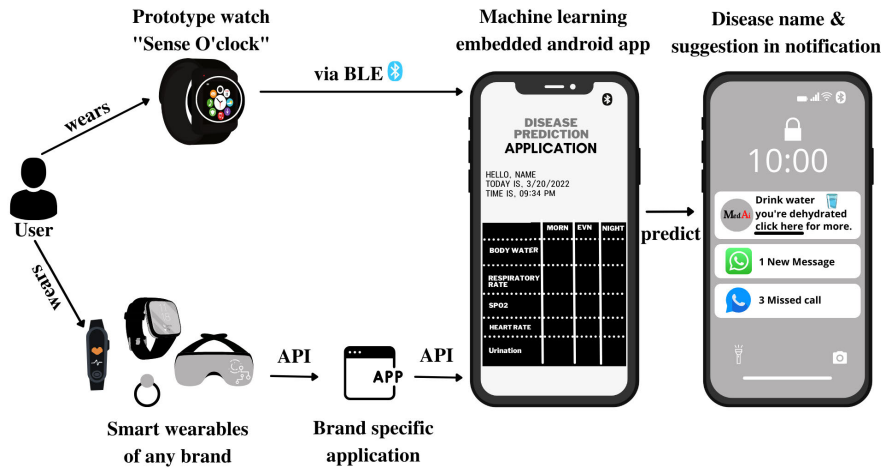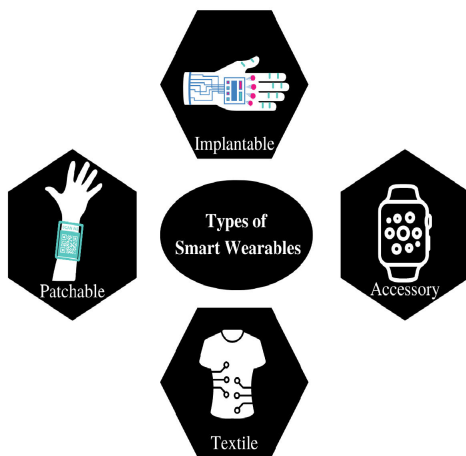
FIGURE 1. Schematic diagram of the system.

FIGURE 2. Types of smart wearable.

II. RELATED WORKS

A. SMART WEARABLES WITH VARIOUS SENSORS

This research involved studying the specifications of several branded and nonbranded smartwatches and wearables. Although a number of smartwatches had excellent performance, it seemed that the smartwatches were believed to be fitness trackers for no valid reason, as most of them did not have any features that could analyze bodily statistics or give feedback if any unusual results were found. The need for a smartwatch such as "Sense O'Clock" emerged after making this determination. This part of the article compares and improves the prototype watch' design by considering existing systems.

In [15], Fern 'andez-Caram'es et al. presented a review article on the past and present state of smart garments to guide developers of IoT-based fabrics, which will communicate with smartphones and operate on biometric information. Although these fabrics can process heart rate and breathing and may even obtain data on hormone levels, the internet of intelligent clothing is still a topic for the future. Our prototype "Sense O'Clock" is a smartwatch that is far more cost- and power-efficient than any smart garment wearable.

In [16], Sivaraj et al. proposed "Medibot", a medical chatbot for a smartwatch. In that peer-reviewed article, they claimed that there was no previous disease prediction app that included a smartwatch. In their system, they added prescription information and skin disease predictions and showed high accuracy. However, their smartwatch comprises only four sensors; it connects to Firebase and predicts disease based on it. The "Sense O'Clock" prototype watch has as many as 11 sensors for predicting disease, which seems feasible because using only four sensors could bias the prediction. Additionally, their watch data are stored in Firebase, which makes it difficult to provide assistance in the absence of an Internet connection.

In [17], Lutze et al. discussed health assistance for older people via smartwatches. They used a branded watch for their system, and by analyzing motion, notifications could be provided to family members; this could be a life-saving application. However, the cost of a high-functionality watch can exceed the budget of ordinary people, and the watch also includes a large number of features for an older adult that could distract from prediction; energy consumption is also a concern here. The "Sense O'Clock" watch, in addition to obtaining data from the sensors, has the minimal interface of a basic watch with necessary but not overly many features dedicated to prediction alone, with a low cost.

In [18], Qiu et al. surveyed smart wearables in fitness. Their research comprised all kinds of wearables, and they showed possible future directions of smart wearables. They stated that with the growth in lifestyle diversity across different ages, the device needs to be customizable according to taste. The prototype watch has a 3D enclosure and simple hardware implementation; people can customize it according to their needs.
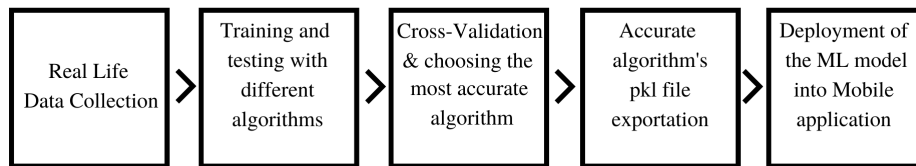
FIGURE 3. Complete workflow from data collection to deployment in ML application.
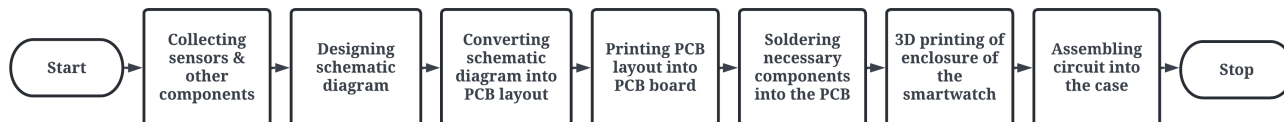

FIGURE 4. Hardware implementation flowchart.

TABLE 1. Smartwatch prototype component description (sensors).

Sensor		
Name	Description	Quantity
Blood Pressure Sensor	Sensor for measurement of the force required by the heart to pump blood around the body.	1
Blood Glucose Sensor	Biosensor for detecting glucose levels.	1
Body Water Sensor	Sensor that shows the water level of the body in numbers.	1
ECG Sensor	Keep track of the electrical signals that cause the heart to beat.	1
Heart Rate Sensor	Sensor records heart rate in real-time.	1
Temperature Sensor	This sensor has the temperature of the body.	1
Respiratory Rate Sensor	Sensor measures breaths per minute.	1
SpO2 Sensor	Detects oxygen level changes.	1
Urination Sensor	Determine urination occurrences in a non-invasive manner.	1
Sleep Tracking Sensor	Keeps sleeping records by the sensor.	1
Stress Level Sensor	Keep records of stress levels.	1

TABLE 2. Smartwatch prototype component description (other components).

Other		
Name	Description	Quantity
ESP32 MCU	It is riched with features. Also used to read data and supports Bluetooth, WiFi, and USB.	1
ST7789 TFT LCD	240x240 Wide Angle Thin Film Transistor LCD Displays	1
ST7789 Controller	Controller for the LCD.	1
CR123 Lithium Battery	ESP32 operates well with one 3V CR123 battery	1
TP4056 LiPo Charger	It's a single-cell lithium-ion battery charger	1
MCP1700T-3302E/TT Voltage Regulator	Regulators with dual outputs, high voltage, and low dropout.	1
Capacitor	A capacitor is a device that employs storage of electrical energy in an electric field.	31 (approx.)
USB Port	A USB2 Type-C port is used.	1
5050 RGB LED	Tri-color LED diodes.	1
Push Switches	Used to cause a temporary change in the state.	4
Buzzer	Device for audio signalling.	1
Resistor	It provides electrical resistance.	18 (approx.)

B. DISEASE PREDICTION USING MACHINE LEARNING METHODS

Machine learning is now being used in many disease prediction studies. After reviewing a number of these studies, we discovered certain common disease prediction systems, particularly those for cardiovascular diseases, diabetes, and other conditions. It is uncommon to find a system that can predict several diseases using real-time data. However, the machine learning algorithms' performance in such systems can be evaluated and compared to the studied system.

In [19], Arumugam et al. devised a machine learning approach for predicting multiple diseases. Although they stated that this was a multiple-disease prediction system, it focused only on predicting heart disease in a diabetic patient. They applied a decision tree, naive Bayes, and support vector machine to their dataset. In contrast, ''MedAi'' can predict twelve different diseases, including diabetes and heart disease, with the deployment of eight popular machine learning algorithms, each providing very high performance.

In [20], Alfifian et al. proposed a health monitoring system for diabetic patients that predicts diabetes and blood glucose levels using machine learning. They used Bluetooth low energy (BLE)-based sensors to collect bodily statistics and performed data processing on them. The performance of classification algorithms such as random forest, naive Bayes, and support vector machine was decent, but in comparison to ''MedAi'', it is not satisfactory. Furthermore, the proposed system only predicts diabetes, in contrast to ''MedAi'', which predicts multiple diseases, making it much more practical.
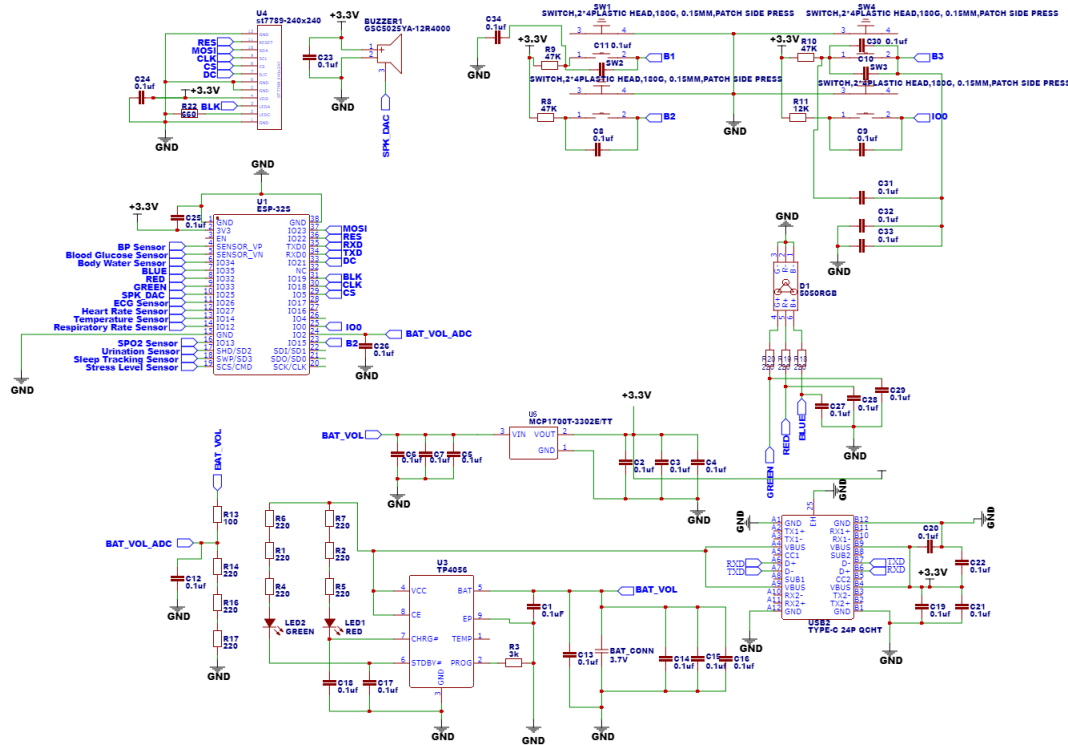
FIGURE 5. Design diagram of "Sense O'clock".

In [21], Kim proposed a cardiovascular disease prediction model using smartwatch data. As mentioned in the article, the initial dataset for machine learning was collected from the Korea National Health and Nutrition Examination Survey 2019, and three machine learning methods were applied. In contrast, "MedAi" employed eight popular machine learning methods on a multivariate dataset and obtained a higher accuracy score than the model proposed by Kim.

In [22], Ayon et al. presented a comparative analysis of seven machine learning techniques for heart disease prediction. According to the article, the dataset was collected from the UCI machine learning repository database and subjected to all seven learning techniques; the deep neural network was shown to produce the best results. In comparison, "MedAi" used eight machine learning approaches to determine the best model for predicting multiple diseases, and the performances of the learning algorithms were far better than those described in the above article.

In [23], Terrada et al. presented a support system, MDSS, for supervised machine learning medical diagnosis of atherosclerosis prediction. The system used machine learning methods such as ANN, AdaBoost, and DT algorithms on a dataset acquired from UCI. Their highest accuracy was 94 percent after running 10-fold cross-validation using the ANN approach. "MedAi", on the other hand, predicts multiple diseases using the best among eight machine learning algorithms, and the accuracy obtained by using the RF algorithm was 99.4%, significantly greater than that of MDSS.

In [24], Tyagi et al. proposed a machine learning-based interactive thyroid disease prediction system. This study discussed the analysis and classification methods used in thyroid disorders based on UCI machine learning repository data. The prediction system involved the application of SVM, KNN, and decision trees and showed a good accuracy score. Thyroid disease (including hyperthyroidism and hypothyroidism) is one of several diseases that "MedAi" predicts; the learning methods RF, SVR, GB, XGB, and LSTM are used in conjunction with those mentioned above to achieve a higher score.

In [25], Wang et al. proposed a chronic kidney disease prediction system using machine learning and an associative classification technique. The study examined chronic kidney disease utilizing machine learning approaches based on a chronic kidney condition. The findings were evaluated across several classification methods, including ZeroR, OneR, naive Bayes, J48, and IBk. It was claimed that the system had a high level of accuracy using the a priori approach; however, the model may not be suitable for other diseases. Conversely, "MedAi" uses eight classifications, regressions, and ensembling approaches to forecast kidney disease and other conditions with a high level of accuracy.

In [26], Ganesan et al. proposed a model for heart disease prediction and diagnosis based on IoT and machine learning. The idea was quite simple; it included collecting health data from different data sources, training the machine learning models, and testing the models with real-life data gathered using IoT devices. The employed machine learning methods

were J48, LR, MLP, and SVM. The highest accuracy obtained was that of J48, which was commendable, but in comparison to "MedAi", the system and accuracy are relatively average.

C. MOBILE APPLICATIONS FOR HEALTH CARE

Various mobile applications developed in different parts of the world were evaluated during this research. Although a few could be found that perform prediction using a pretrained machine learning model, most other applications provide basic health monitoring functionalities. The functional features and feasibility of these applications were tested and compared with those of the "MedAi" application.

In [27], Rahman et al. performed a pilot study to create a smart-health framework that analyzes biomarkers collected from wearables at a low cost. The framework was a smartphone application. Although the application used ML, the patient needed to provide their disease and symptoms, and the application would only measure the severity of the disease and visualize it. The "MedAi" application will be able to predict the disease for which the person is at risk with only smartwatch data instead of using multiple components, such as IJP sensors, scanners, and commercial wearables in the framework, with less time and cost.

In [28], Johari et al. proposed a self–checkup mobile app that predicts the early stages of diseases. It extracts information from existing datasets and has a few diseases in its prediction list that are not very prevalent. The prediction uses the user's manual input, from which it is not currently feasible to predict diseases. The "MedAi" application gathers data in real-time through a smartwatch and assists people with their health.

In [29], Rathi et al. proposed a mobile healthcare application with a disease prediction and recommendation system. Although they claimed that this application could predict all kinds of diseases, in reality, it was successful with only five diseases. Moreover, the system requires manual input every time it predicts, which is inconvenient. "MedAi", on the other hand, predicts diseases without user intervention.

In [30], Sethi et al. wrote an article on a mobile application using symptoms for disease prediction. They used manual input and selection-based prediction, which is very broad and helpful. However, the selection process requires good English proficiency and knowledge of medical terms for operating the application. The "MedAi" application is fairly straightforward and has a user-friendly UI to help people use it comfortably without the need to have domain knowledge of other fields.

Reviewing and analysing all the research revealed the inadequacies in the earlier works. Even though many disease prediction systems use wearables, machine learning algorithms, and mobile applications, none offer a comprehensive, simple, and satisfying solution within a single framework. The number of diseases predicted was reasonably low, even if wearable technology and machine learning had been used. Additionally, some efforts have been discovered to create mobile or online applications that offer predicted results yet call for user manual input. "MedAi" solves all these problems by proposing a wearable design with a plethora of sensors dedicated to multiple disease prediction, a machine learning model developed after carefully analysing several widely used learning algorithms and choosing the best performing of them. A mobile application which is self-directed and provides real-time prediction results and suggestions connecting with the wearable and ML model.

III. METHODOLOGY

A. SYSTEM ARCHITECTURE

The system is named "MedAi" because it involves the integration of "Artificial Intelligence" and "Medical Technology". The system requires the user to wear the prototype watch "Sense O' clock"", which includes all the necessary sensors. The smartwatch will then connect to a smartphone that has a machine learning-enabled app installed. After registering on the "MedAi" system by signing into the application with basic information, the watch will gather the user's body readings and transfer serial data through BLE to the machine learning embedded Android application. This application will make a request to the Flask Rest API for a prediction and fetch response, which will be displayed on the screen and in the notification bar.

As illustrated in Fig. 2, there are various types of smart wearables. Although the mobile application was created for use with the prototype watch, it can be used with any other wearable device. However, these include third-party APIs, which can lead to security breaches and data theft. Additionally, implantable and patchable smart wearables raise skin health concerns and are inconvenient. Even the price of smart cloth is out of reach for most people. Our system includes the design and blueprint for making the "Sense O'Clock" smartwatch.

As the system is based on disease prediction, building a reliable and high-performing machine learning model is the core of the system. The optimal deployment of the mobile application and the prediction of diseases will depend on the accuracy of the model and the speed with which the prediction is delivered to the application. Taking all these into consideration, a dataset consisting of records for twelve distinct diseases partitioned for 5-fold cross-validation is used for training and testing all eight learning algorithms, which include random forest, support vector machine, and K-nearest neighbors. Finally, the model built with the best-performing algorithm is embedded into a mobile application for performing all tasks, from fetching smartwatch data to connecting with the API to make predictions. The application will then notify and display the prediction result and suggestions, as described above.

B. "SENSE O'CLOCK" PROTOTYPE WATCH FRAMEWORK

Currently, there are many smartwatch brands that provide health or fitness tracking features, but only a few of them

provide accurate results. Those that give higher accuracy are much more expensive than typical ones, which makes it difficult for most people to afford them. Not only is the expense a drawback, but the current features available in these smartwatches are not sufficient to predict several acute health problems in an individual. There is a need for a smartwatch that contains all the sensors needed to predict common health problems to potentially save someone from sudden death. The newly designed smartwatch will have nearly all the sensors available to detect diseases, which is the goal of the system. It will have a large number of sensors and the basic features of a clock, which is why it is named "Sense O'Clock". The smartwatch architecture will include 3 phases, as described below.

1) Hardware
2) Software
3) Communication

1) HARDWARE IMPLEMENTATION PHASE
Smartwatches are electronic devices, so hardware implementation will take up to 80% of the implementation phase. As the system description suggests, there are some strong requirements for the hardware. Fig. 4 illustrates the steps to be followed in the implementation of the hardware.

a: SENSORS AND OTHER COMPONENTS
As shown in Fig. 1, the "MedAi" system depends on sensory data from "Sense O'Clock" for a variety of disease predictions. ESP32 MCU allows the integration of several sensors with a high level of compatibility. This feature has enabled the MCU to gain popularity in health monitoring wearable development [31]. 34 GPIO (General Purpose Input/Output) pins on the ESP32 can be used for various purposes by programming the relevant registers. Several of these pins have been utilized to place the 11 sensors required for reading values of body vitals. The machine-learning model's features and attributes will take those values as input, so all the sensors must be in the watch.

All of these are easily collectable from markets, and those will be much less expensive than other branded smartwatch components.

b: DESIGNING SCHEMATIC DIAGRAM
Schematics can give a clear view of electron source and path as well as circuit components. Fig. 5 depicts a schematic diagram of "Sense O'clock".

c: CONVERTING THE SCHEMATIC DIAGRAM INTO A PCB LAYOUT
The printed circuit board (PCB) is the foundation of any smartwatch. All the components will be integrated and then soldered on this board. The semitransparent layer view of the PCB design gives a clear view of the connections, and the 3D view illustrates the component placement more precisely in Fig. 6.
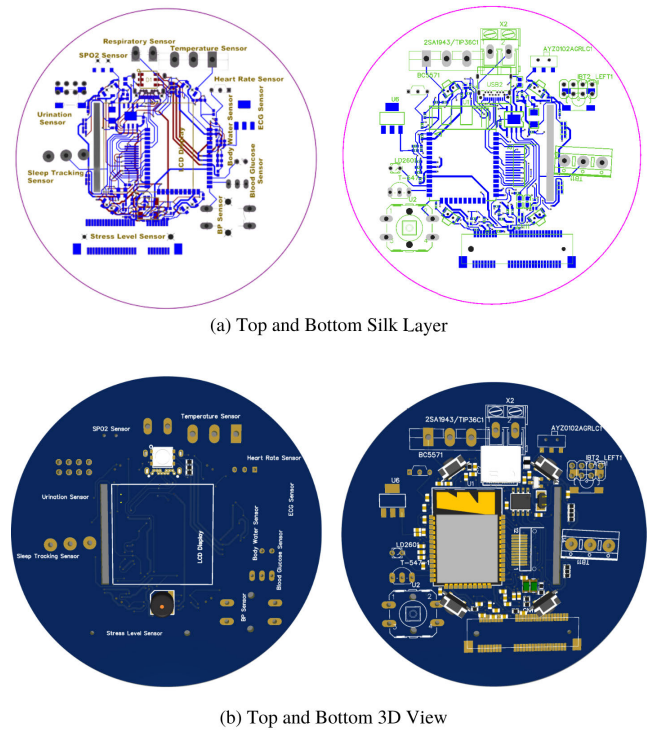

(a) Top and Bottom Silk Layer

(b) Top and Bottom 3D View

FIGURE 6. PCB design of "Sense O'clock".

d: PRINTING THE PCB LAYOUT ONTO A PCB BOARD
The previous phase concerned designing the PCB layout; in this phase, the design will be printed on an actual PCB (a nonconductive substance with printed or etched conductive lines).

e: SOLDERING THE NECESSARY COMPONENTS ONTO THE PCB
All the components mentioned in phase.1 are soldered onto the board.

f: 3D PRINTING OF THE ENCLOSURE OF THE SMARTWATCH
Each part of the watch body is 3D printed using a 3D printer. Thus, we can customize the parts to the shape and size of the PCB.

g: ASSEMBLING THE CIRCUIT IN THE CASE
Finally, the circuit is fixed in the 3D enclosure. The hardware part of Sense O'Clock is then finished.

2) SOFTWARE IMPLEMENTATION PHASE
A smartwatch is considered an IoT device. That is why integrating the hardware with the software is essential. Fig. 7 displays the framework for the software implementation of "Sense O'Clock". The working of the software is as such, it collects data from the sensors and checks for the connection of the mobile device. If the connection is found to be established, it calls the BLE component embedded in the system and sends serial raw data of each sensor to the device.
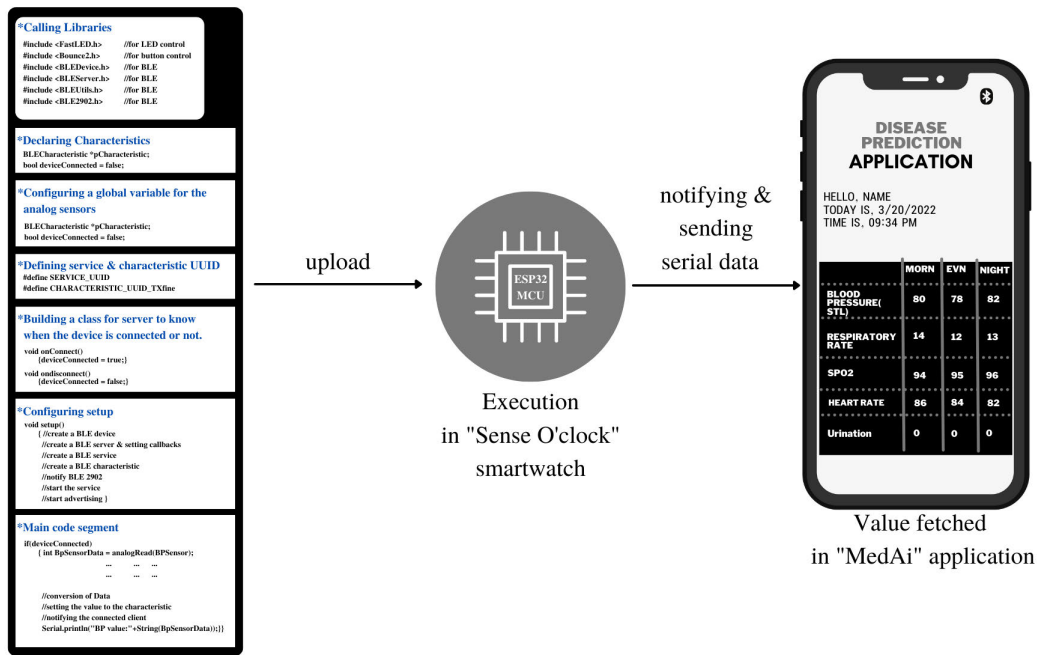
FIGURE 7. Framework for the software implementation part of "Sense O'Clock".

The "MedAi" application receives the serial data and sends them to API for processing.

3) COMMUNICATION PHASE

The ESP32 MCU is well known for its robust design and power savings. Its level of integration is very high, as it has a built-in power amplifier and various filters. It can perform as a standalone system or as a follower device and hosts an MCU. It enables Wifi and BLE features through its SPI/SDIO or I2C/UART interfaces. The "MedAi" system needs to send real-time data to a mobile application. This MCU can accomplish this task in 3 common ways.

a: WiFi

WiFi is typically faster than using mobile networks. This could be an excellent option for transferring data, as it can deliver data efficiently. However, some issues prevent the "MedAi" system from using this communication medium. First, one can obtain this connectivity at a certain cost, which may hinder its mass accessibility. Second, although some developed countries have WiFi connectivity in all places, there is no such facility in less developed countries, as internet connection may be available at home only and not in public places.

b: THIRD-PARTY API

Some APIs from well-known companies are free of cost and have many other benefits. These APIs may organize data and let people retain vast amounts of data on their servers, but there are undeniable security risks. As the system will deal
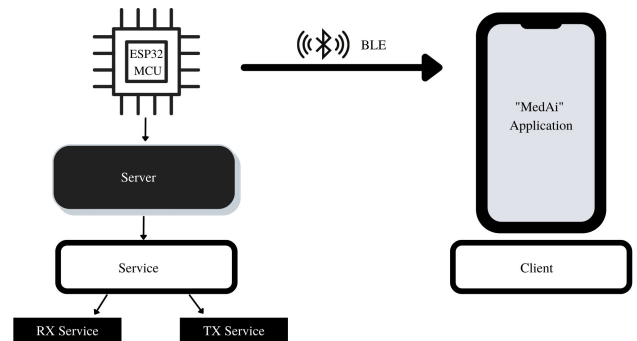

FIGURE 8. Transfer of data to "MedAi" Application via BLE.

with health-related information, privacy remains a concern. Therefore, this medium will not be used either.

c: BLE

ESP can send data over BLE, which is a slightly different protocol than typical Bluetooth. The main difference is that regular Bluetooth constantly sends data, but BLE does not. BLE uses servers that can notify clients when a change in data occurs. It is much more suitable for low-power IoT applications such as smartwatches. From Fig. 8, it can be observed that in BLE, both the server and client have universally unique identifiers (UUIDs) that help the system determine which server-client pair will be connected. Each server has a service; each service has characteristics for reading and transmitting data. The "MedAi" system will use BLE because of its power optimization, security, and availability.

Fig 9 presents the whole system in a nutshell, giving an overview of the components used and steps followed to obtain the output.
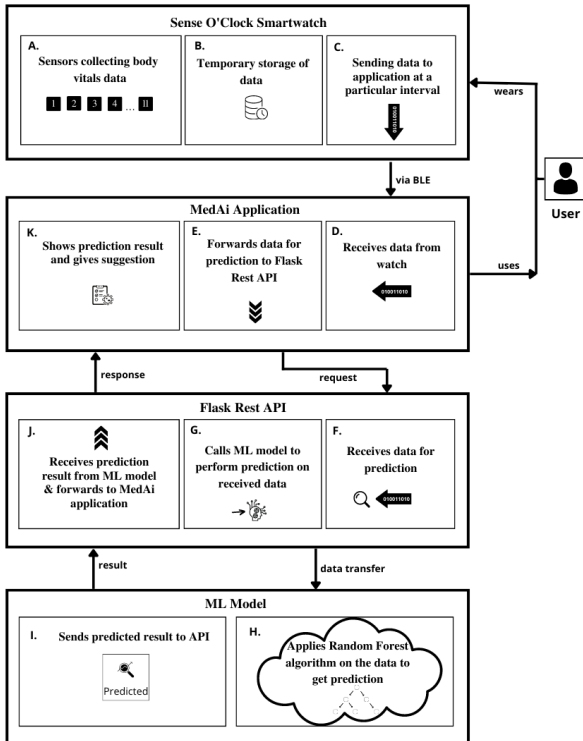
FIGURE 9. MedAi framework.

TABLE 3. Dataset description table.

Sl. No.	Attributes	Dtype	Values
1	Age	int64	Range:[9-73]
2	Sex	int64	Male=0 and Female=1
3	BMI	float64	BMI:[16.3-32.4]
4	ChainSmoker	int64	Yes=1 and No=0
5	Chest pain type	int64	No=0 and Type 1,2,3,4
6	D Blood Pressure(mmHg)	int64	Diastolic BP: [60-120]
7	C Blood Pressure(mmHg)	int64	Systolic BP: [100-183]
8	BloodGlucose	int64	Blood Glucose: [64-245]
9	Cholesterol(LDL)	int64	LDL Range: [73-233]
10	Cholesterol(HDL)	int64	HDL Range: [25-79]
11	Body water	int64	Normal=0 and Abnormal=1
12	ECG results	int64	Normal=0, STD=1, STE=2, QW=3 and T'=4
13	Max Heart Rate	int64	Range:[60-128]
14	Temp(F)	int64	Range: [96-102]
15	Respiratory Rate(per minute)	int64	Range: [12-37]
16	SpO2	int64	Range: [45-128]
17	Urination	int64	Normal=0 and Abnormal=1
18	Sleep hours	float64	Range: [4-9]
19	stress level	int64	Range: [21-79]
20	Predicted	int64	Healthy=0, IHD=1, COPD=2, DM=3, CVA=4, GSD=5, HTN=6, AKI=7, DLP=8, MI=9, HTR=10, HTO=11, DHN=12

C. MACHINE LEARNING MODEL IMPLEMENTATION

1) DATA COLLECTION

This part of the project was slightly challenging, as open-source datasets could not be found for all the diseases 'that we intended to predict. Moreover, the collection of observational data from patients wearing smartwatches would not have been feasible, as the system involves the use of many sensor data for the prediction of several diseases that current smartwatches do not provide. Because of this concern, records for approximately 150 patients with specific diagnoses were obtained from Hospital, with written consent from the patients. Each patient was provided with a copy of the ethical consent form. The reports were also checked and approved by the legal authority of the hospital. The health records and diagnoses from the report were taken as symptoms for specific diseases and used to train our model, and the test results obtained were further validated by expert doctors from the hospital.

The final validated dataset consists of around 260 data that have been used for training and testing machine learning models is discussed in the latter part of this section.

2) ATTRIBUTE SELECTION

Multiple disease prediction necessitates the assessment of various attributes, each with a different threshold value that indicates the presence or absence of certain diseases [32], [33]. A total of twenty attributes were evaluated in order to make the best possible prediction [34]. The attribute characteristics for various diseases are shown in Table 3.

3) DATA PREPROCESSING

Data preprocessing is a crucial aspect of the learning process. If this step is skipped, the learning process may suffer major consequences. The preprocessing section includes-
- Data acquisition,
- Library import,
- Dataset import,
- Missing data handling,
- Data encoding,
- Dataset splitting,
- Feature scaling etc.

Following the loading of the essential libraries and dataset in this study, it was verified for missing values, and as none were discovered, the next steps were taken. Since all attribute values were numerical, the dataset did not require encoding. Some insignificant attributes that might affect the outcome were removed through feature selection to improve learning. Fig. 10 shows the heatmap of the final attributes that will proceed for further processes.

4) IMPLEMENTATION OF LEARNING ALGORITHMS

After splitting the train and test sets, several learning methods were applied to the dataset under the study. Different parameter values for each of the distinct algorithms were chosen based on the dataset characteristics. Implementation of several learning methods along with the parameters used are discussed as follows:

a: SUPPORT VECTOR REGRESSION

Kernel: This algorithm utilized 2 values of kernel functions for the dataset, 'rbf' and 'linear'.

Gamma parameter: This determines the reachability of a single training example, which is measured in terms of 'far'
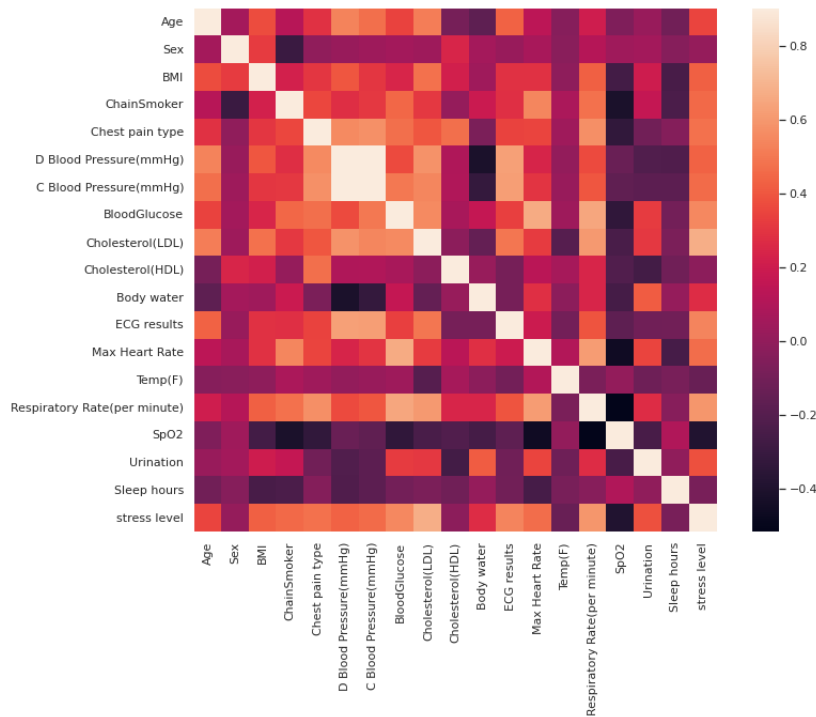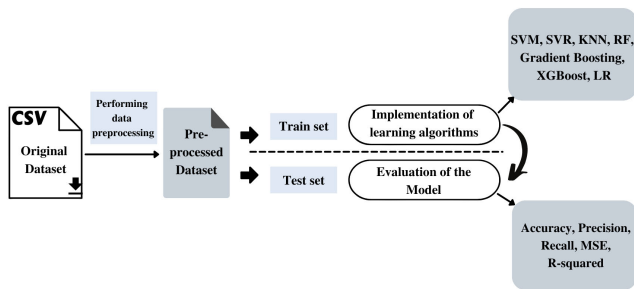
FIGURE 10. Attribute heatmap.

FIGURE 11. Machine learning workflow.

and 'near'. For the dataset, the gamma parameter has been set to 0.1.

Using the above-mentioned values for the parameters, SVR and Linear Regression methods have been applied to the train and test data. Finally, The Mean Square Error (MSE) and R-squared performance parameters reflected how well the training and testing set performed.

b: SUPPORT VECTOR MACHINE

SVM classifier used only a single kernel parameter for the dataset under consideration.

Kernel: 'linear' kernel has been used by SVM for classification analysis of the dataset.

The following table summarizes the results of the approaches on the test dataset:

The training model was evaluated using performance measures such as accuracy, precision, and recall, while the testing

model's performance was demonstrated using only accuracy, with the depiction of confusion matrices for both.

c: K-NEAREST NEIGHBORS

The algorithm requires the predefinition of the parameter K, which is the number of closest neighbors who will be involved in the majority voting process.

K parameter: For this study, the value of K has been defined as 3

After fitting the training and testing sets, the performance of the model has been measured in terms of accuracy, precision, and recall. The visualization of the confusion matrix also shows the number of correct classifications.

d: RANDOM FOREST CLASSIFIER

The most important of the Random Forest Classifier's parameters is n_estimators, which specifies the number of base models/decision trees that contribute to the final result via majority vote.

n_estimators: The number of estimators for this classification task has been set at 100.

As for performance analysis, the same metrics as the above classification methods have been determined after performing training and testing on the datasets.

e: XGBoost

XGBoost is applicable to both binary and multiclass classific problems. One of its most notable parameters is the loss function.
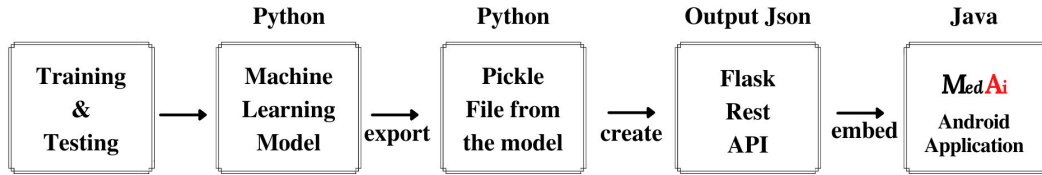
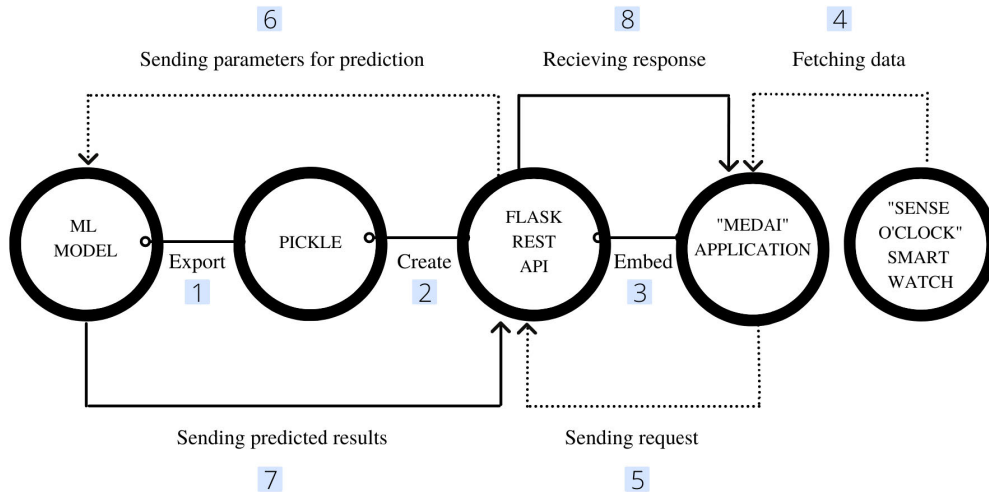**FIGURE 12.** Process of machine learning embedding in android application.

FIGURE 13. Workflow of embedding process.

TABLE 4. Classification algorithm's performance on ischemic heart disease dataset.

			IHD					
			Testing Performance using Classification Algorithm					
	SVM-Linear	**SVM-Poly**	**SVM-RBF**	**LR**	**RF**	**KNN**	**GB**	**XGB**
Accuracy	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	75.00%	75.00%
Precision	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	75.00%	75.00%
Recall	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	75.00%	75.00%
F1-Score	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	75.00%	75.00%

TABLE 5. Regression algorithm's performance on ischemic heart disease dataset.

	IHD		
	Testing Performance using Regression Algorithm		
	LSTM		
RMSE	2.39%		
	SVR		
	SVR-RBF	**SVR-Linear**	**Regression Linear**
MSE	22.07%	5.62%	4.88%
R-Squared	73.25%	98.32%	98.54%

binary:logistic: XGBoost uses this parameter to predict the probability for binary classification problems.

multi:softprob: It is used for multiclass classification.

Since the system consists of 12 datasets to forecast 12 distinct diseases, each dataset only requires binary classification, hence each dataset's loss function binary:logistic is used.

f: GRADIENT BOOSTING
This ensembling technique works by first fitting the model, tuning the parameters, and making predictions. The notable parameters are n_estimators, learning_rate, max_depth, etc. Cross-validation is also very crucial to reducing the variance in estimating the algorithm's performance. In this study, 10 fold cross-validation has been applied to improve the model's performance.

g: LOGISTIC REGRESSION
Logistic regression uses the sigmoid function to perform classification. It provides a very simple classifier for model training and testing and often omits the necessity of explicit declaration of any parameter within the function.

Among several optional parameters of logistic regression, a few are mentioned as follows

penalty: This determines whether or not there is regularization and, if so, which strategy should be used. The value is 'l2' by default.

dual: It determines whether to use the dual or primal formulation.

tol: This specifies the procedure's tolerance for halting. By default, it is set to 0.0001.

TABLE 6. Classification algorithm's performance on stroke dataset.

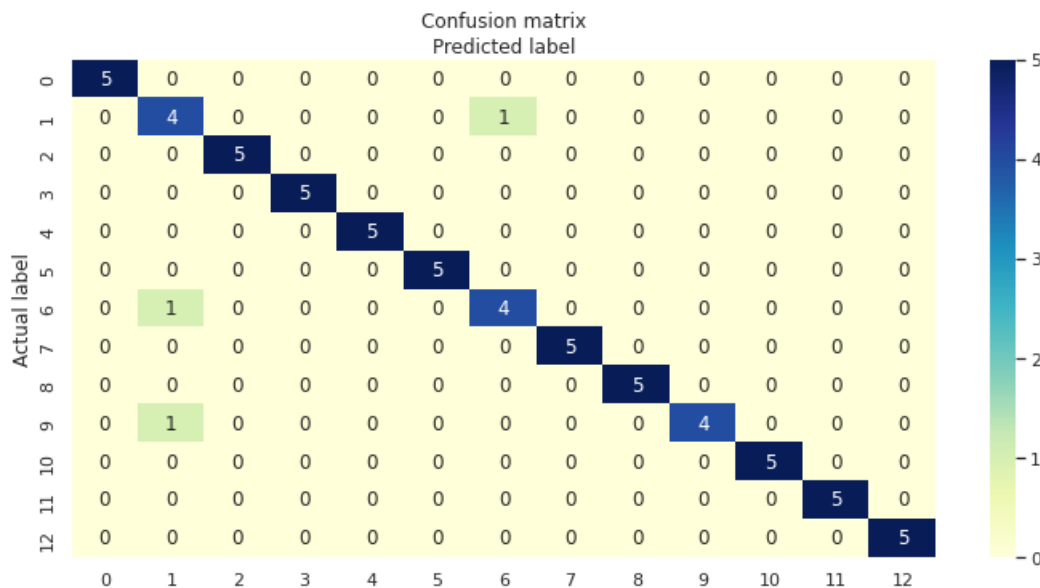|	CVA Testing Performance using Classification Algorithm							
	SVM-Linear	**SVM-Poly**	**SVM-RBF**	**LR**	**RF**	**KNN**	**GB**	**XGB**
Accuracy	96.67%	96.67%	96.67%	96.66%	96.67%	96.65%	75.00%	75.00%
Precision	96.85%	96.85%	96.85%	96.85%	96.85%	96.85%	75.00%	75.00%
Recall	96.67%	96.67%	96.67%	96.67%	96.67%	96.67%	75.00%	75.00%
F1-Score	96.64%	96.64%	96.64%	96.65%	96.64%	96.65%	75.00%	75.00%

FIGURE 14. Confusion matrix of entire dataset on the application of RF algorithm.

TABLE 7. Regression algorithm's performance on stroke dataset.

	CVA Testing Performance using Regression Algorithm		
	LSTM		
RMSE	2.61%		
	SVR		
	SVR-RBF	**SVR-Linear**	**Regression Linear**
MSE	93.99%	23.91%	15.78%
R-Squared	48.39%	92.39%	98.98%

Optional parameters have not been defined for our system. Performance has been measured by calculating accuracy after training and testing.

h: LONG SHORT TERM MEMORY

LSTM mostly works on sequence classification problems utilizing keras library for deep learning. LSTM has several parameters that take up different values based on the problem requirement. The following are the parameters that were employed in this research:

LSTM(4): Defines LSTM layer with 4 memory cells.

Dense(1): To create 0 or 1 predictions for the two classes, a dense output layer with a single neuron and a sigmoid activation function are used.

loss= 'mean_squared_error': The loss function is the mean absolute percentage error that computes the difference between the actual and predicted value.

optimizer='adam': ADAM is an efficient optimization algorithm.

epoch: For this system, 100 passes have been made for the entire training dataset.

Upon completion of training and testing using the LSTM classifier, performance has been measured in terms of Root Mean Square Error(RMSE).

D. DEPLOYMENT OF THE MODEL INTO MOBILE APPLICATION

In Fig. 12, it can be seen that after training and testing the ML model, a pickle file can be exported from the model. This pickle file will be used to create a Flask Rest API that gives output in JSON format. API acts like a request-response carrier between frontend Java(Android Studio application) and backend Python(ML model and Pickle file). The need for Rest API is that the response needed for the application is in a structured and straightforward format rather than a complete webpage. Then Flask Rest API will be embedded into the "MedAi" android application to complete the system.

TABLE 8. Classification algorithm's performance on hypertension dataset.

	SVM-Linear	SVM-Poly	SVM-RBF	LR	RF	KNN	GB	XGB
	HTN Testing Performance using Classification Algorithm							
Accuracy	90.00%	90.00%	90.00%	90.00%	90.56%	90.65%	75.00%	75.00%
Precision	90.68%	90.68%	90.68%	90.68%	90.67%	90.68%	75.00%	75.00%
Recall	90.00%	90.00%	90.00%	90.00%	90.56%	90.65%	75.00%	75.00%
F1-Score	90.22%	90.22%	90.22%	90.23%	90.22%	90.22%	75.00%	75.00%

TABLE 9. Regression algorithm's performance on hypertension dataset.

	HTN Testing Performance using Regression Algorithm		
	LSTM		
RMSE	4.27%		
	SVR		
	SVR-RBF	SVR-Linear	Regression Linear
MSE	98.73%	34.26%	37.78%
R-Squared	36.49%	96.98%	96.88%

Fig. 13 states that after creating the "MedAi" application as mentioned above, when data is fetched to the android application from the smartwatch, it will send a request to API with the parameters (features value of the disease in this case). API will forward the request to the ML model with the extracted feature values to be tested for prediction. The ML model after performing test will send prediction response to API, which will further be forwarded to the application for display notification.

Fig. 15 consists of proposed application interfaces. (A) interface illustrates the opening page of the application. The "getting started" button will show the permissions page on the (B) interface. The "I agree" button will grant permission to establish a Bluetooth connection, collect data from the smartwatch, and allow app notifications over the lock screen. Upon agreeing, the (C) interface will include the sign-up page that will take necessary inputs like name, gender, age, and other pieces of information which the watch will not be able to provide. Once sign-in is complete, all the readings will be shown in the interface (D) and let anyone click on the predict button to know the disease he/she is at risk. Interface (F) shows predicted disease with some essential remedies. Finally, interface (G) shows how the notification panel will look when the "MedAi" application notification pops up.

Sensors naturally produce a large amount of data that might require a storage medium, but offline storage can be a costly and extra burden on the hardware, while online storage can create privacy issues. To this concern, "MedAi" has been enabled to clear out all unnecessary data and store only a minimal amount of data valid for evaluation.

IV. EXPERIMENTATION AND RESULT ANALYSIS
The major experimental component of this system is the creation of a reliable prediction model. On each of the disease datasets, eight algorithms were applied, including SVM, LR, RF, KNN, GB, XGB, SVR, and LSTM.

The datasets were then amalgamated, and 5-fold cross-validation was used to apply the same approaches to the entire dataset. The cross-validation enabled hyperparameter optimization and assess model overfitting [35]. The process is depicted in Fig. 16.

A. PERFORMANCE EVALUATION METRICS
There are a number of performance measures that can be used to assess the classification and regression analyses. Some popular classification performance metrics used in this study are as follows:

Accuracy: The ratio of the number of correct predictions to the total number of predictions is known as accuracy.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

Precision: The ratio of the number of correctly predicted positives to the total number of positive predictions. Precision is effective in lowering the number of False Positives.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

Recall: The ratio of the number of correctly predicted positives to the total number of actual positives. The recall is particularly useful in reducing the number of False Negatives.

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

F1-Score: It is equivalent to the harmonic mean of both precision and recall.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

Here,

TP = True Positive (Correctly predicted as positive).
TN = True Negative (Correctly predicted as negative).
FP = False Positive (Incorrectly predicted as positive).
FN = False Negative (Incorrectly predicted as negative).

The regression analysis performance measures employed in this study include-

MSE: Mean Square Error defines how close the data points are to the fitted regression line.

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \quad (5)$$

TABLE 10. Classification algorithm's performance on entire dataset.

	Entire Dataset Testing Performance using Classification Algorithm (5-fold cross-validation							
	SVM-Linear	**SVM-Poly**	**SVM-RBF**	**LR**	**RF**	**KNN**	**GB**	**XGB**
Accuracy	89.86%	74.00%	78.00%	98.57%	99.4%	99.32%	98.40%	98.56%
Precision	90.87%	75.21%	79.13%	98.64%	99.6%	99.36%	98.94%	98.61%
Recall	89.86%	74.00%	78.00%	98.57%	99.4%	99.32%	98.90%	98.56%
F1-Score	89.81%	74.87%	78.22%	98.59%	99.1%	99.33%	98.88%	98.60%

FIGURE 15. Interfaces of android application.

TABLE 11. Regression algorithm's performance on entire dataset.

	Entire Dataset Testing Performance using Regression Algorithm (5-fold cross-validation)		
	LSTM		
RMSE	9.14%		
	SVR		
	SVR-RBF	**SVR-Linear**	**Regression Linear**
MSE	89.87%	96.89%	93.03%
R-Squared	58.75%	40.55%	49.41%

RMSE: Root Mean Square Error is simply the square root of Mean Square Error.

$$RMSE = \sqrt{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2} \quad (6)$$

R-Squared: R-squared indicates how much variance in the output / anticipated variable is explained by the input variables.

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y}_i)^2} \quad (7)$$

B. MODEL EVALUATION

Since the study intends to predict multiple diseases, the complete dataset has been partitioned into twelve subgroups for twelve different diseases for the convenience of use and better performance. Dataset is balanced, with almost equal number of records for each category. Each category of dataset was then subjected to all eight learning methods.

TABLE 12. Comparison with the other existing systems.

| Author & Study	Method	Predicted Diseases	Accuracy	Limitations
Narayanan et al. (2022) [36]	CNN, Random Forest and Logistic Regression	Malaria, Alzheimer's disease, TUBERCULOSIS, Pancreatic Cancer	0.9641, 0.88, 0.9836, 0.91 respectively	1.Needs user input which is hasslesome. 2. Lacks detection of fatal diseases.
Phasinam et al. (2022) [37]	Naive Bayes, ID3, C4.5, and SVM	Heart Disease	0.92, 0.85, 0.78, 0.7 respectively very commendable.	1. Lack of real-time implementation. 2. Less accurate.
Singh et al. (2022) [38]	KNN, Random Forest, XGBoost	Diabetes, Heart Disease, Liver Cancer	0.89, 0.85, 0.73 respectively 0.73 respectively	1. Require user input. 2. Less user-friendly UI.
Indukuri et al. (2021) [39]	LR, KNN SVM	Diabetes, Heart Disease, and Breast cancer	0.77, 0.83, 0.94 respectively	1. Limited to only three diseases. 2. Lack of optimum algorithm usage. 3. Lack of real-time implementation.
Atri et al. (2022) [40]	CNN	Parkinson Disease	0.9	1. Lack of assistive feature. 2. Lack of noble wearable design. 3. Not comfortable to wear. experimental watch for long.
MedAi	SVR, SVM, KNN, LR, RF, XGBoost, Gradient Boosting, LSTM	Ischemic heart disease, Hypertension Respiratory disease Hyperthyroidism, Hypothyroidism, Stroke, Myocardial infarction Kidney failure, Gallstones, Diabetes, and Dyslipidemia	0.99 (RF)	1. Requires Robust data.
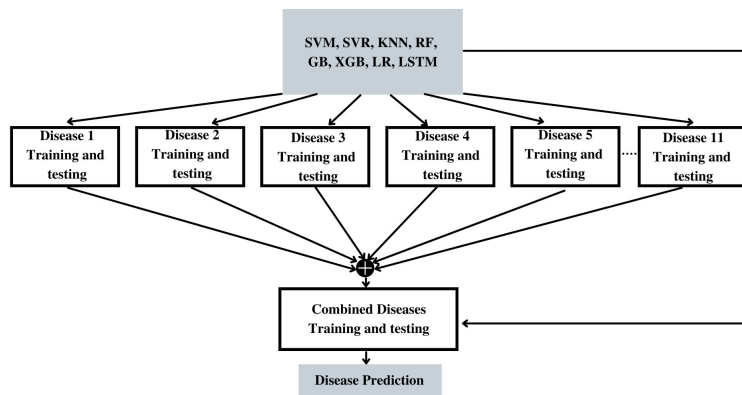
FIGURE 16. Learning steps from units to whole.

1) TRAINING PERFORMANCE

The training process suffered no loss, as the datasets had been preprocessed to remove missing values and outliers.

2) TESTING PERFORMANCE

The outcomes of some of the dataset's testing on the application of each method will be presented in a tabular format here.

For each dataset, all the classifiers can be seen to provide similar performance. However, LR is seen to outperform other models on datasets for Respiratory Disease, Stroke, Gallstones, and Dehydration, with a maximum accuracy of 96.67 percent.

3) PERFORMANCE EVALUATION OF THE ENTIRE DATASET

Following the successful training and testing of each of the dataset subsets, the subsets were combined into a single dataset and split for training and testing using the same machine learning methods.

Table 10 and Table 11 shows the outcome of the performance measures obtained by using five-fold cross-validation on the entire dataset. It is evident from the table that all the algorithms provided quite a satisfactory accuracy score. However, the RF algorithm can be seen to outperform all the other algorithms providing 99.4% accuracy on test data. The confusion matrix for test data on the application of the RF algorithm is depicted in Fig. 14.

As can be seen from the confusion matrix of test set in Fig: a total of 3 instances were misclassified, which are: 1 instance of Ischemic Heart Disease (1), 1 instance of Hypertension (6) and 1 instance of Myocardial Infarction. The numbers from 0-12 "Actual label" in the confusion matrix corresponds to the diseases provided in Table 3. The reason for such misclassification could be the ambiguous nature of those tests data. Some values of certain parameters can be similar for two different diseases e.g patients having IHD and HTP both have high blood pressure and some similar traits in other vitals, so getting a misclassification is unavoidable.

4) CREATION OF PICKLE FROM THE NOTEBOOK

Since it is now evident that RF is the best learning method for the dataset under consideration, the next step is to create another notebook that contains all the twelve diseases' datasets cross-validated, trained, and tested with the RF algorithm. The final model is then exported as a '.pkl' file for the development of the mobile application.

V. CONCLUSION AND FUTURE WORK

In this research, we successfully developed a framework for the system "MedAi" that included designing a smartwatch with many sensors, building a machine learning model by evaluating eight algorithms, and proposing a framework for an Android health application that predicts twelve kinds of disease. This system will help people of any age and gender care for their bodies. It is not a replacement for medical specialists but rather an aid in identifying symptoms that might lead to serious diseases or even death and motivating individuals to live a healthy lifestyle. The system is expected to greatly reduce the sudden death and undiagnosed terminal disease rate. This low-cost, energy saving and user-friendly system is a lifesaving package. Any user can monitor the condition of his or her body and be alerted to any significant changes since diseases manifest themselves through various changes in bodily statistics. The system provides a robust prediction strategy based on a highly accurate prediction model and a real-life observational health dataset of patients validated by doctors. The model was built with the RF algorithm, which provided an accuracy of 99.4% over other popular methods, such as KNN (99.3%) and XGB (98.56%). The system is reliable and protects user privacy, preventing the possibility of security breaches. We believe the system will be a healthy addition to one's life.

Future research on "MedAi" application framework includes increasing the dataset since a large amount of data will further improve the model. Production of the smartwatch is also a future task, as we already have a complete design of "Sense O'Clock". We will also add more seasonal diseases to the prediction list since the system already contains prominent diseases, and we will publish the mobile application on the Google Play Store.

REFERENCES

[1] G. Kessler. (2016). *Are There Really 10,000 Diseases and Just 500 'Cures'?*. Accessed: Apr. 1, 2022. [Online]. Available: https://www.washingtonpost.com/news/fact-checker/wp/2016/11/17/are-there-really-10000-diseases-and-500-cures/

[2] (2019). *Global Burden of Disease (GBD 2019), Institute for Health Metrics and Evaluation*. Accessed: Apr. 1, 2022. [Online]. Available: https://www.healthdata.org/gbd/2019

[3] (2022). *Health Resources—Health Spending—OECD Data*. Accessed: Apr. 1, 2022. [Online]. Available: https://data.oecd.org/healthres/health-spending.htm

[4] (2022). *Diseases Potentially Acquired by Travel to Southeast Asia*. Accessed: Apr. 1, 2022. [Online]. Available: https://www.uptodate.com/contents/diseases-potentially-acquired-by-travel-to-southeast-asia

[5] (2022). *Global Health and Fitness app Downloads 2020 | Statista*. Accessed: Apr. 1, 2022. [Online]. Available: https://www.statista.com/statistics/1127248/health-fitness-apps-downloads-worldwide/

[6] (2022). *Digital Around the World—DataReportal—Global Digital Insights*. DataReportal-Global Digital Insights. Accessed: Apr. 1, 2022. [Online]. Available: https://datareportal.com/global-digital-overview

[7] M. Whaiduzzaman, M. Mahi, A. Barros, M. Khalil, C. Fidge, and R. Buyya, "BFIM: Performance measurement of a blockchain based hierarchical tree layered fog-IoT microservice architecture," *IEEE Access*, vol. 9, pp. 106655–106674, 2021, doi: 10.1109/access.2021.3100072.

[8] (2022). *What Is Sudden Death?*. Accessed: Apr. 1, 2022. [Online]. Available: https://sudden.org/about-sudden-death/

[9] (2022). *How Often Should You Get Routine Checkups at the Doctor?*. Accessed: Apr. 1, 2022. [Online]. Available: https://www.healthline.com/health/how-often-should-you-get-routine-checkups-at-the-doctor#access-to-care

[10] (2022). *Global Connected Wearable Devices 2016–2022 | Statista*. Accessed: Apr. 1, 2022. [Online]. Available: https://www.statista.com/statistics/487291/global-connected-wearable-devices

[11] T. Ramya, "Disease prediction system using fuzzy C-means algorithm," *Int. J. Eng. Res. Technol. (IJERT)*, vol. 6, no. 3, pp. 1–4, 2022, Accessed: Apr. 1, 2022. [Online]. Available: https://www.ijert.org/research/disease-prediction-system-using-fuzzy-c-means-algorithm-IJERTCONV6IS03003.pdf

[12] W. Wang, J. Lee, F. Harrou, and Y. Sun, "Early detection of Parkinson's disease using deep learning and machine learning," *IEEE Access*, vol. 8, pp. 147635–147646, 2020. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9165732

[13] V. Jackins, S. Vimal, M. Kaliappan, and M. Y. Lee, "AI-based smart prediction of clinical disease using random forest classifier and naive Bayes," *J. Supercomput.*, vol. 77, no. 5, pp. 5198–5219, May 2021, Accessed: Apr. 1, 2022. [Online]. Available: https://link.springer.com/article/10.1007/s11227-020-03481-x#citeas

[14] (2022). *The Top 10 Causes of Death*. Accessed: Apr. 1, 2022. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death

[15] T. Fernández-Caramés and P. Fraga-Lamas, "Towards the internet-of-smart-clothing: A review on IoT wearables and garments for creating intelligent connected E-textiles," *Electronics*, vol. 7, no. 12, p. 405, Dec. 2018, Accessed: Apr. 1, 2022. [Online]. Available: https://www.mdpi.com/2079-9292/7/12/405

[16] K. Sivaraj, K. Jeyabalasuntharam, H. Ganeshan, K. Nagendran, J. Alosious, and J. Tharmaseelan, "Medibot: End to end voice based AI medical chatbot with a smart watch," *Int. J. Creative Res. Thoughts*, vol. 9, no. 1, pp. 201–206, 2021. Accessed: Apr. 1, 2022. [Online]. Available: https://www.researchgate.net/profile/Janani-Tharmaseelan/publication/348834221

[17] R. Lutze and K. Waldhor, "Personal health assistance for elderly people via smartwatch based motion analysis," in *Proc. IEEE Int. Conf. Healthcare Informat. (ICHI)*, Aug. 2017, Accessed: Apr. 1, 2022, doi: 10.1109/ichi.2017.79.

[18] H. Qiu, X. Wang, and F. Xie, "A survey on smart wearables in the application of fitness," in *Proc. IEEE 15th Int. Conf Dependable, Autonomic Secure Comput., 15th Int. Conf Pervasive Intell. Comput., 3rd Int. Conf Big Data Intell. Comput. Cyber Sci. Technol. Congress (DASC/PiCom/DataCom/CyberSciTech)*, Nov. 2017, pp. 303–307, Accessed: Apr. 1, 2022, doi: 10.1109/dasc-picom-datacom-cyberscitec.2017.64.

[19] K. Arumugam, M. Naved, P. P. Shinde, O. Leiva-Chauca, A. Huaman-Osorio, and T. Gonzales-Yanac, "Multiple disease prediction using machine learning algorithms," *Mater. Today, Proc.*, Aug. 2021, pp. 1–4, Accessed: Apr. 1, 2022, doi: 10.1016/j.matpr.2021.07.361.

[20] G. Alfian, M. Syafrudin, M. Ijaz, M. Syaekhoni, N. Fitriyani, and J. Rhee, "A personalized healthcare monitoring system for diabetic patients by utilizing BLE-based sensors and real-time data processing," *Sensors*, vol. 18, no. 7, p. 2183, Jul. 2018, Accessed: Apr. 1, 2022, doi: 10.3390/s18072183.

[21] M.-J. Kim, "Building a cardiovascular disease prediction model for smart-watch users using machine learning: Based on the Korea national health and nutrition examination survey," *Biosensors*, vol. 11, no. 7, p. 228, Jul. 2021, Accessed: Apr. 1, 2022, doi: 10.3390/bios11070228.

[22] S. Ayon, M. Islam, and M. Hossain, "Coronary artery heart disease prediction: A comparative study of computational intelligence techniques," *IETE J. Res.*, vol. 68, pp. 1–20, Jan. 2020, Accessed: Apr. 1, 2022, doi: 10.1080/03772063.2020.1713916.

[23] O. Terrada, S. Hamida, B. Cherradi, A. Raihani, and O. Bouattane, "Supervised machine learning based medical diagnosis support system for prediction of patients with heart disease," *Adv. Sci., Technol. Eng. Syst. J.*, vol. 5, no. 5, pp. 269–277, 2020, Accessed: Apr. 1, 2022. [Online]. Available: https://astesj.com/v05/i05/p33/

[24] A. Tyagi, R. Mehra, and A. Saxena, "Interactive thyroid disease prediction system using machine learning technique," in *Proc. 5th Int. Conf. Parallel, Distrib. Grid Comput. (PDGC)*, Dec. 2018, pp. 689–693, Accessed: Apr. 1, 2022. [Online]. Available: https://ieeexplore.ieee.org/document/8745910

[25] Z. Wang, J. W. Chung, X. Jiang, Y. Cui, M. Wang, and A. Zheng, "Machine learning-based prediction system for chronic kidney disease using associative classification technique," *Int. J. Eng. Technol.*, vol. 7, no. 4.36, p. 1161, Dec. 2018, Accessed: Apr. 1, 2022, doi: 10.14419/ijet.v7i4.36.25377.

[26] M. Ganesan and N. Sivakumar, "IoT based heart disease prediction and diagnosis model for healthcare using machine learning models," in *Proc. IEEE Int. Conf. Syst., Comput., Autom. Netw. (ICSCAN)*, Mar. 2019, pp. 1–5, doi: 10.1109/ICSCAN.2019.8878850.

[27] M. Rahman, B. Morshed, B. Harmon, and M. Rahman, "A pilot study towards a smart-health framework to collect and analyze biomarkers with low-cost and flexible wearables," *Smart Health*, vol. 23, Mar. 2022, Art. no. 100249, Accessed: Apr. 1, 2022, doi: 10.1016/j.smhl.2021.100249.

[28] N. A. A. M. Johari, N. Mohamad, and N. Isa, "Smart self-checkup for early disease prediction," in *Proc. IEEE Int. Conf. Autom. Control Intell. Syst. (ICACIS)*, Jun. 2020, pp. 33–38, Accessed: Apr. 1, 2022. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9140205

[29] M. Rathi and V. Pareek, "Mobile based healthcare tool an integrated disease prediction & recommendation system," *Int. J. Knowl. Syst. Sci.*, vol. 10, no. 1, pp. 38–62, Jan. 2019, Accessed: Apr. 1, 2022. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9140205

[30] R. S. Sethi et al., "Disease prediction application based on symptoms," *Int. J. Sci. Res. Comput. Sci., Eng. Inf. Technol.*, pp. 641–646, 2019, doi: 10.32628/cseit1952189.

[31] T. S. Jahnavi, N. Thumma, K. A. Reddy, U. V. Teja, M. Gunasree, and M. Kiran, "Health monitoring smart glove using esp32 microcontroller," *J. Phys., Conf.*, vol. 2325, no. 1, Aug. 2022, Art. no. 012007, Accessed: Oct. 5, 2022, doi: 10.1088/1742-6596/2325/1/012007.

[32] S. M. R. Al Masud, A. A. Bakar, and S. Yussof, "A modified WBANs MAC superframe using priority-criticality index table for managing pilgrims emergency traffic in Hajj," *Indonesian J. Elect. Eng. Comput. Sci.*, vol. 14, pp. 1499–1507, Feb. 2019, doi: 10.11591/ijeecs.v14i3.pp1499-1507.

[33] H. Murtaza, M. Iqbal, Q. Abbasi, S. Hussain, H. Xing, and M. Imran, "Correlation analysis of vital signs to monitor disease risks in ubiquitous healthcare system," *EAI Endorsed Trans. Ind. Netw. Intell. Syst.*, vol. 7, no. 24, Oct. 2020, Art. no. 165676, Accessed: Aug. 2, 2022, doi: 10.4108/eai.18-5-2020.165676.

[34] N. Faruqui, M. A. Yousuf, M. Whaiduzzaman, A. K. M. Azad, A. Barros, and M. A. Moni, "LungNet: A hybrid deep-CNN model for lung cancer diagnosis using CT and wearable sensor-based medical IoT data," *Comput. Biol. Med.*, vol. 139, Dec. 2021, Art. no. 104961.

[35] K. T. Chui, B. B. Gupta, W. Alhalabi, and F. S. Alzahrani, "An MRI scans-based Alzheimer's disease detection via convolutional neural network and transfer learning," *Diagnostics*, vol. 12, no. 7, p. 1531, Jun. 2022, Accessed: Sep. 26, 2022, doi: 10.3390/diagnostics12071531.

[36] S. Narayanan, N. M. Balamurugan, and P. B. Palas, "Leveraging machine learning methods for multiple disease prediction using Python ML libraries and flask API," in *Proc. Int. Conf. Appl. Artif. Intell. Comput. (ICAAIC)*, May 2022, pp. 694–701, Accessed: Oct. 5, 2022, doi: 10.1109/icaaic53929.2022.9792807.

[37] K. Phasinam, T. Mondal, D. Novaliendry, C.-H. Yang, C. Dutta, and M. Shabaz, "Analyzing the performance of machine learning techniques in disease prediction," *J. Food Qual.*, vol. 2022, pp. 1–9, Mar. 2022, Accessed: Oct. 5, 2022, doi: 10.1155/2022/7529472.

[38] A. Singh and A. Yadav. (2022). *International Research Journal of Engineering and Technology (IRJET)*. Accessed: Oct. 5, 2022. [Online]. Available: https://www.irjet.net/archives/V9/i3/IRJET-V9I3312.pdf

[39] I. Mohit, K. S. Kumar, U. A. K. Reddy, and B. S. Kumar, "An approach to detect multiple diseases using machine learning algorithm," *J. Phys., Conf.*, vol. 2089, no. 1, Nov. 2021, Art. no. 012009, Accessed: Oct. 5, 2022, doi: 10.1088/1742-6596/2089/1/012009.

[40] R. Atri, K. Urban, B. Marebwa, T. Simuni, C. Tanner, A. Siderowf, M. Frasier, M. Haas, and L. Lancashire, "Deep learning for daily monitoring of Parkinson's disease outside the clinic using wearable sensors," *Sensors*, vol. 22, no. 18, p. 6831, Sep. 2022, Accessed: Oct. 5, 2022, doi: 10.3390/s22186831.

SHINTHI TASNIM HIMI was born in Dhamrai, Dhaka, Bangladesh, in 1998. She received the B.Sc. degree in computer science and engineering from Jahangirnagar University, Savar, Dhaka, in 2020, where she is currently pursuing the master's degree.

She is also working on two additional research projects and aspires to continue working on more such research in the future. She has been actively involved in research projects, since 2018. She has been keen to research from the very early days of her undergraduate life and attended a number of educational conferences. Her first journal publication was in *IJIM*, in 2020, and her first conference paper was published in IEEE Xplore, in 2020. Her research interests include big data analysis, machine learning, artificial intelligence, and human–computer interaction.

NATASHA TANZILA MONALISA was born in Dhaka, Bangladesh, in 1999. She received the B.Sc. degree in computer science and engineering from Jahangirnagar University, Savar, Dhaka, in 2020, where she is currently pursuing the M.Sc. degree.

Since 2018, she has been actively engaged in research works. She is also working on two new search initiatives that should be released soon. She attended a number of educational conferences and seminars while an undergraduate. Her first journal publication was in *IJIM*, in 2020, and her first conference paper was published in IEEE Xplore, in 2020. Her research interests include deep learning, artificial intelligence, natural language processing, and cyber security.

MD WHAIDUZZAMAN (Senior Member, IEEE) received the B.Sc. degree in electronics and computer science and the M.Sc. degree in telecommunication and computer network engineering from London, U.K., and the Ph.D. degree from the University of Malaya, Malaysia. He is currently a Professor with the Institute of Information Technology (IIT), Jahangirnagar University. He is also working as a Research Fellow on ARC-funded projects with the Queensland University of Technology, Australia. His research interests include cloud computing, edge computing, the IoT, and cyber security. Recently, he received the Best Paper Award for the *JNCA* (Elsevier), Paris, France.

ALISTAIR BARROS is currently a Professor of information systems and the Head of the Services Computing Program, Information Systems School, Queensland University of Technology. He has 32 years of ICT experience across industry and industrial research and development and academic roles, including a Global Research Leader and the Chief Development Architect at SAP AG. His research interests include cloud, enterprise systems and microservices engineering, and evolution and provisioning using model-based techniques.

MOHAMMAD SHORIF UDDIN (Senior Member, IEEE) received the B.Sc. degree in electrical and electronic engineering from the Bangladesh University of Engineering and Technology (BUET), Bangladesh, in 1991, the M.Ed. degree in technology education from Shiga University, Japan, in 1999, the Ph.D. degree in engineering from the Kyoto Institute of Technology, Japan, in 2002, and the M.B.A. degree from Jahangirnagar University, in 2013. In 1991, he started his teaching career at the Chittagong University of Engineering and Technology (CUET) as a Lecturer. Later, he moved as a Lecturer to the Department of Computer Science and Engineering, Jahangirnagar University, in 1992, where he is currently a Professor. In addition, he oversees the ICT Cell of Jahangirnagar University as a Teacher-in-Charge. He worked as the Chairperson of the Department of Computer Science and Engineering, Jahangirnagar University, from June 2014 to June 2017. He worked as an Advisor of ULAB, from September 2009 to October 2020, and Hamdard University Bangladesh, from November 2020 to November 2021. He undertook a postdoctoral research at the Bioinformatics Institute, Singapore; the Toyota Technological Institute, Japan; the Kyoto Institute of Technology, Japan; Chiba University, Japan; Bonn University, Germany; and the Institute of Automation, Chinese Academy of Sciences, China. He was the Coach of Janhangirnagar University ACM ICPC World Finals Teams, in 2015 and 2017. He supervised a good number of doctoral and master's theses. He holds two patents for his scientific inventions and has published more than 170 research papers in international journals and conference proceedings. In addition, he edited a good number of books and wrote many book chapters. He had delivered a remarkable number of keynotes and invited talks. His research interests include artificial intelligence, imaging informatics, and computer vision. He is a fellow of the IEB and the BCS. He received the Best Paper Award in the International Conference on Informatics, Electronics and Vision (ICIEV 2013), Dhaka, Bangladesh; and the Best Presenter Award from the International Conference on Computer Vision and Graphics (ICCVG 2004), Warsaw, Poland. He acted as a general chair or TPC chair or co-chair of many international conferences. He is an Associate Editor of IEEE Access. He is also serves as the President for the Bangladesh Computer Society (BCS).

• • •