9-3-2019

Machine Learning Enabled Vital Sign Monitoring System

Varun Kumar Vats

University of Windsor

Machine Learning Enabled Vital Sign Monitoring System

By

Varun Kumar Vats

A Thesis
Submitted to the Faculty of Graduate Studies
through the Department of Electrical and Computer Engineering
in Partial Fulfillment of the Requirements for
the Degree of Master of Applied Science
at the University of Windsor

Windsor, Ontario, Canada

2019

Machine Learning Enabled Vital Sign Monitoring System

by

Varun Kumar Vats

APPROVED BY:

P. Zadeh
Department of Computer Science

E. Abdel-Raheem
Department of Electrical & Computer Engineering

K. Tepe, Advisor
Department of Electrical & Computer Engineering

September 3rd , 2019

DECLARATION OF ORIGINALITY

I hereby certify that I am the sole author of this thesis and that no part of this thesis has been published or submitted for publication.

I certify that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my thesis and have included copies of such copyright clearances to my appendix.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other University or Institution.

ABSTRACT

Internet of Things (IoT)- based remote health monitoring systems have an enormous potential of becoming an integral part of the future medical system. In particular, these systems can play life-saving roles for treating or monitoring patients with critical health issues. On the other hand, it can also reduce pressure on the health-care system by reducing unnecessary hospital visits of patients. Any health care monitoring system must be free from erroneous data, which may arise because of instrument failure or communication errors. In this thesis, machine-learning techniques are implemented to detect reliability and accuracy of data obtained by the IoT-based remote health monitoring. A system is a set-up where vital health signs, namely, blood pressure, respiratory rate, and pulse rate, are collected by using Spire Stone and iHealth Sense devices. This data is then sent to the intermediate device and then to the cloud. In this system, it is assumed that the channel for transmission of data (vital signs) from users to cloud server is error-free. Afterward, the information is extracted from the cloud, and two machine learning techniques, i.e., Support Vector Machines and K-Nearest Neighbor are applied to compare their accuracy in distinguishing correct and erroneous data. The thesis undertakes two different approaches of erroneous data detection. In the first approach, an unsupervised classifier called Auto Encoder (AE) is used for labeling data by using the latent features. Then the labeled data from AE is used as ground truth for comparing the accuracy of supervised learning models. In the second approach, the raw data is labeled based on the correlation between various features. The accuracy comparison is performed between strongly correlated features and weakly correlated features. Finally, the accuracy comparison between two approaches is performed to check which method is performing better for detecting erroneous data for the given dataset.

DEDICATION

To my Parents

Vinod Kumar Vats & Varsha Rani Sharma

&

To my Brother

Sarthak Kumar Vats

&

To my Friends & Family

&

To IEEE Windsor Section

ACKNOWLEDGEMENTS

I would like to express my gratitude to my supervisor Dr. Kemal Tepe for his invaluable support, guidance, and good sense of humour throughout the course of my time in WiCIP lab. I would like to extend a big thanks to Sabbir Ahmed, Danilo Coral-De Witt, Elvin Enziama, Lining Zhang and other lab members of WiCIP lab for their guidance and support. I am grateful to my family for their guidance and sacrifice. I would also like to thank Dr. Esam Abdel-Raheem and Dr. Pooya Moradian Zadeh for their useful feedback and assistance.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AE	Autoencoder
AES	Advanced Encryption Standard
AI	Artificial Intelligence
ANN	Artificial Neural Network
API	Application Program Interface
BG	Blood Glucose
BP	Blood Pressure
BLE	Bluetooth Low Energy
BSN	Body Sensor Network
CT	Computer Tomography
DIA	Diastolic
IEEE	Institute for Electrical and Electronics Engineers
IBM	International Business Machine Corporation
IoT	Internet of Things
KNN	K-Nearest Neighbor
LoRA	Long Range
RFID	Radio Frequency Identification
RHM	Remote Health Monitoring
RR	Respiration Rate
SL	Supervised Learning
SQL	Standardised Query Language
SVM	Support Vector Machine
SYS	Systolic
TCP/IP	Transmission Control Protocol/Internet Protocol
t-SNE	t-Stochastic Nearest Embedding

UID	Unique Identifier
USL	Unsupervised Learning
Wi-Fi	Wireless Fidelity
WMD’s	Wearable Mobile Devices

Chapter 1 - Introduction

1.1. Background

Internet of Things (IoT) has been revolutionizing the healthcare industry. The entire market is positioning itself to the ability to monitor remotely, use smart sensors, integration of medical equipment. Improved patient health and safety, enhanced physician care delivery, and better doctor-patient engagement are few benefits that come with IoT. In countries like the USA, UK, Germany, Canada and Australia, the smart healthcare is mostly associated with providing healthcare facilities to ambient assisted living. The smart healthcare also helps in educating people about the healthcare facilities available to them, and they can use those facilities for self-managing emergencies. There are various disadvantages associated with smart healthcare, such as security and privacy, integration challenges, and technology adoption.

Machine Learning (ML) has evolved from pattern recognition, where computers can learn without being programmed to specific tasks. In recent days, ML is used for various applications related to healthcare such as Disease Identification, Drug Discovery, Medical Image Diagnosis, Robotic Surgical Tools, etc.
However, "The Institute of Medicine at the National Academics of Science, Engineering and Medicine" reports that diagnostic errors contribute to around (10% -17%) of hospital complications and also accounts for approximately (10%) of patient deaths [1].

1.2.Motivation

In the modern day, there are various fitness and medical devices available in the market, which monitor and provide different vital and medical parameters such as calories burnt, steps count, oxygen saturation level (SPO_2), blood pressure, heart rate variability and respiration rate. These devices use Bluetooth Low Energy (BLE) technology to connect to other devices such as mobile phone. The major drawback of these devices is that the data flow occurs in one way although some companies such as Fitbit and Garmin provide the functionality of making their phone applications (apps) to receive data with limited features.

There is a need to build a system which can analyze collected data for accuracy, consistency and viability. With such a system, the collected data can be used by medical purposes. For example, if any of the parameter crosses the threshold value, an automatic message can be sent to the medical practitioners. The medical practitioners can, therefore, observe the data and provide medications to the concerned person on time. This system can play an essential role in providing remote health monitoring to patients and most notably be useful for elderly, women, and infants.

This work proposed an ML integration for smart healthcare. There are various techniques in which ML algorithms can be implemented, such as feature learning, sparse dictionary learning, anomaly detection, decision trees, and association rules. In ML perspective, the main focus of this work is towards anomaly and outlier detection in IoT based data collection. The primary purpose of outlier detection is to detect errors in the normal behavior of the system. This methodology is essential in remote monitoring systems for providing accurate measurements to medical practitioners.

1.3. Problem Statement

The following two significant challenges need to be addressed:

The first challenge comes while establishing the communication phase. There are a lot of smart devices which enables data flow from the sensor device attached a patient to the mobile phone. However, in some of these devices there is no provision by which data can flow from the device to a medical practitioner for monitoring or examination. In order to provide a proof of concept, sensor devices with having open APIs are selected to demonstrate this challenge can be overcome and data can be collected and processed in a cloud system.

The second challenge that needs to be addressed arises in the data processing and usage phase. In the data processing phase, it must be ensured that the collected data is reliable. To check if the data is reliable, ML will be used in anomaly detection algorithms.

1.4. Thesis Contribution

In this thesis, we have developed a prototype to collect healthcare-related data using IoT devices. The data being collected from devices includes different parameters, i.e., respiration rate, heart rate, and blood pressure (SYS and DIA). The data (Vital Signs) was collected considering different body movements such as sleeping, walking, running, rest and exercising. The respiration rate data is obtained from one device, while blood pressure and heart rate data are collected using another device. These two devices collect data simultaneously, and data is stored on a SQL database. This data was further extracted from the SQL database for analysis. The data analysis considered samples from different users but due to shortage of samples from multiple users the analysis was performed considering samples (1000) from single user. Two different approaches are considered for identifying anomalies in the data. In the first approach, the raw data is being sent as the input to the Autoencoder (AE). The primary function of the AE in our thesis is to determine internal relation among different parameters and classify them accordingly. The classification from AE results in labeled data as inliers and outliers. The inliers, in our case, are the values that are being considered as correct values while incorrect values are regarded as outliers. The labeling being obtained from the AE serves as a ground truth for training supervised learning models. The labelled data obtained from AE is further divided into training set and test set. The training set is used for training supervised models while test set is used for further validation. The models considered in this work are Support Vector Machine (SVM) and K-Nearest Neighbor (K-NN). Therefore, in this approach, these models are compared based on accuracy being computed using the ground truth obtained from AE.

In the second approach, the correlation is being found among different parameters — the correlation results in two different types, such as weak and strong. The weak correlation is found among respiration rate and pulse rate, while a correlation is observed between systolic and diastolic pressures. The association between systolic and pulse rate, systolic and respiration rate, diastolic and pulse rate, diastolic, and respiration rate are negatively correlated. Two different methods are being considered in labeling the data. In the first method, labeling is performed using strong correlation while in another technique, labeling is performed using weakly correlated features. The labeling among weak and robust relationship is shown to observe the difference in accuracy achieved using both the methods. In the last step, the accuracy comparison is being conducted among two different approaches.

The main objective of using two different approaches in our work provides a way of improving accuracy of the model as well as saving implementation time of the model. The first way (Automatic Labelling) of determining accuracy is based on ML models (AE, SVM, KNN) and provides an accuracy about 94%. However, the second method (Manual Labelling) provides an accuracy of about 87% with the help some predefined threshold value.

1.5. Thesis Organization

This thesis is organized as follows: Chapter 1 describes the concepts of the IoT for smart healthcare systems and ML for medical diagnosis on an introductory basis and provides the thesis's contributions. Chapter 2 provides the technical background of the work and literature survey in the field of IoT and ML for smart healthcare. Chapter 3 presents the hardware setup of the system and the specifications of the devices used in the prototyping. It also explains the concept of outlier detection using ML algorithms. Chapter 4 details about the results based on ML algorithms taken into account along with the comparison of those algorithms based on data taken into consideration. Chapter 5 provides a conclusion and recommendations for future work in the proposed system.

Chapter 2 - Literature Review

2.1 Background

Remote Health Monitoring (RHM) is a system in which patients are monitored outside the current clinical environment. This technique is helpful in saving time and reducing healthcare delivery cost. In the case of chronic diseases, RHM is useful in improving the quality of life of a patient. The application areas of RHM includes body temperature monitoring, diabetes monitoring, congestive heart failure detection, infertility, heart rate variability, etc. In modern days, RHM is implemented using wearable devices, smartphone apps, communication systems such as Radio Frequency Identification, Lora/Bluetooth, ZigBee, etc.

Present day state of RHM notwithstanding, IoT and ML have helped further revolutionize RHM where doctors can monitor multiple patients and provide medication simultaneously, regardless of the type of medication needed or place inhabited. Hence, the combination of ML with IoT is useful not only for monitoring but also for ensuring efficient diagnosing illnesses.

There is a significant growth in the RHM devices market and various surveys have been conducted in the past. The Grand View Research Company conducted a study showing increasing demand for RHM devices from 2012-2022. Their study showed that in 2012 the market of RHM was around 5.00 billion USD. In 2016 the market of RHM was 10.00 billion USD, it is expected to reach 40 billion USD in 2022. RHM devices are used in various industries such as m-health, remote patient monitoring, wearable sensor technology, wearable devices and telehealth [2].

Having mentioned some of the advantages of RHM, there are some disadvantages associated with RHM such as data integration, cost and lack of proper guidelines.

2.2 IEEE 1073 Standard

The IEEE 1073 Standard provides guidelines for medical practitioners for capturing data from patients' body using medical devices [3]. This standard was particularly designed to meet the following requirements.

- Being able to handle frequent network re-configuration
- Offering clinic - friendly plug and play operation
- Providing robust, reliable communications for the safe execution of critical applications

The complete, accurate capture of data from all devices connected to the patient provides following applications.

- Designing and improving treatment protocols
- Providing remote access to real time patient data
- Providing real-time notification of drug infusion protocols
- Facilitating real-time analysis of drug effectiveness
- Providing network management of connected devices
- Enabling real-time changes in patient care

2.3 Internet of Things

IoT is defined as a system of interconnected devices each having unique identifiers (UID). Sharing of information between interconnected devices over the network does not require human-to-human or machine-to-machine communication. Applications of IoT includes smart cities, smart environment, smart agriculture-health, etc. However, this thesis focuses on role of IoT for healthcare industry. Thus, IoT in healthcare industry is broadly classified into three main categories namely, clinical care, remote monitoring, and context awareness discussed widely in [4].

A. Clinical Care

Clinical care requires IoT devices for monitoring physiological condition of patients. Collection and analyzation of patient's data is performed using body connected sensors and transmission of sensor data to processing centers provides suitable actions. Recently, IBM took an initiative to design hand washing monitoring system using Radio Frequency Identification (RFID). Main aim of this design is to avoid infections in hospitals after patient's treatment. Thus, preventing 90,000 deaths as well as saving $30 million yearly.

B. Remote monitoring

Remote monitoring is a method which uses IoT for providing medical facilities. In case of an emergency, remote monitoring plays an important role by providing medical facilities remotely; thereby, saving time by avoiding unnecessary hospital visits. In order to provide medical facilities to elderly people a connected e-health system is proposed. Sensors in this system are used for sending signals to doctors who can then continuously monitor the patient's data while also providing the latest information and data insights to emergency healthcare services in real time. By using a twin device (health-kits) arrangement wherein one of the devices (attached to the patient's body) synchronizes with the other (inside the ambulance) facilitating sound communication via Message Queuing Telemetry Transport (MQTT) sessions over the Transmission Control Protocol (TCP) protocol, the exact location of the ambulance can be determined in real time.

C. Context Awareness

Context awareness is an important method of determining a patient's condition where the patient's environment proves to be a vital consideration. In the event of a certain change in the patient's environment, healthcare professionals can now actively monitor these changes relative to the physical state of the patient. Studies have shown that the variations in physical being of a patient due to environmental changes can influence his/ her ability to endure adverse effects of various diseases that he/ she may have been inadvertently exposed to, in the past. Thus, using IoT in the form of sensors to monitor the physical state of the patient during various activities (walking, running, sleeping, etc.) which can then be directly relayed to healthcare professionals for real time monitoring proves to be an important aspect of making vital healthcare facilities available to the patient at the most appropriate point in time possible.

2.3.1. Challenges

Many researchers are working on designing and implementing healthcare-based systems. This section outlines some of those challenges, which are discussed widely in [5] .

A. Standardization

Standardization deals with maintaining the standard of devices utilized in the healthcare industry. In recent days, IoT devices are being manufactured in sizably voluminous numbers by sundry local manufacturers who are concentrating on incrementing quantity while the standard is being compromised.

B. Continuous Monitoring

Continuous monitoring deals with body connected sensors utilized for analyzing and accumulating the physiological condition of the patient for perpetually observing the health of the patient in case of an emergency. There are very few devices available in the market having precise results.

C. Data Protection

Data Protection deals with providing security to health data. This data is considered crucial since it contains the information about body vitals as well as personal information. Ergo, bulwarking this data is considered crucial since any hindrance to health data from the outside environment can have devastating results and lead to life threating issues.

2.4 Machine Learning

2.4.1 Overview

ML arises from Artificial intelligence (AI) where computer systems learn without being explicitly programmed [6]. This process of learning begins with observations or data such as training examples and direct experiences to look for patterns in those examples and make predictions in the feature based on those examples.

There are various applications of ML, such as email spam detection, credit card fraud detection, speech recognition, etc.

2.4.2 Types of Machine Learning

There are two broad categories of ML algorithms. These algorithms are discussed briefly in the following sections.

2.4.2.1 Supervised Learning

The supervised learning algorithms are based on labeled data. The SL algorithms focus on the training of these labeled data. Each labeled data consists of the input value and the desired output value. These SL algorithms analyze the training data and make an inferred function, which is used for mapping new values. The SL algorithms are preferred for data analyses.

2.4.2.2 Unsupervised Learning

The Unsupervised Learning algorithms contains data that has not been labeled, classified, or categorized. These algorithms focus on detecting commonalities in the data and react based on the presence or absence of such commonalities. The USL algorithms are also used for data analyses.

2.4.3. Applications of Machine Learning for Healthcare

Healthcare industry has various benefits such as providing value-based care. Moreover, this industry has helped various countries in earning revenue. Technology in healthcare is useful not only in providing patient care, billing, and medical records but also for medical practioners in developing staffing models, providing smart care, reducing administrative and supply costs. ML is useful in healthcare industry for various purposes such as analyzing data points and providing outcomes, precise resource allocation, etc. Thus, some of the applications of ML in healthcare industry are discussed widely.

A. Predictive Analytics

   Predictive analytics is defined as the branch of analytics which is used to make predictions about unknown future events. This method uses various techniques such as data mining, ML and AI to make predictions about the future. ML devices are used by clinicians of Beacon Health operations to extract actionable insights from structured and unstructured patient data. Determining the most appropriate treatment with minimum recovery time for people afflicted with second- or third-degree burns is recommended by using ML techniques [7].

B. Personalized Medicine

Rapidly advancing field of healthcare i.e., personalized medicine comprises of an individual's unique clinical, genetic and environmental information. Factors that enable personalized medication includes phenotype categories, population size and statistical analysis [8]. Phenotype categories considers knowledge about human disease, including their subtypes. Statistical Analysis considers mechanism to classify, predict, or diagnose on the basis of statistical method. Statistical analysis uses ML to develop models from dataset for obtaining quantitative results.

However, certain advancements are required for practical implementation of personalized healthcare. In case of population size, more enrollment of patients is required for improving visibility. Phenotype categorization is achieved by improved sharing of published/ unpublished data and efficient use of biotechnologies with improved collaboration. Advances in statistical data is currently being hindered by two factors; namely, computational genomics and voluntary negligence by clinicians towards statistical analysis.

2.4.4 Anomaly Detection

Anomaly detection in healthcare is defined as the deviation of healthcare data from the average value; a strange condition, or inconsistency. The anomaly is often described as something that determines whether the event is malicious or not. The best way to determine anomaly is to understand the various techniques for anomaly collection. In the medical domain, most of the techniques considers defining point anomalies. There are two approaches to Anomaly detection techniques, as shown in the Figure 2.2.

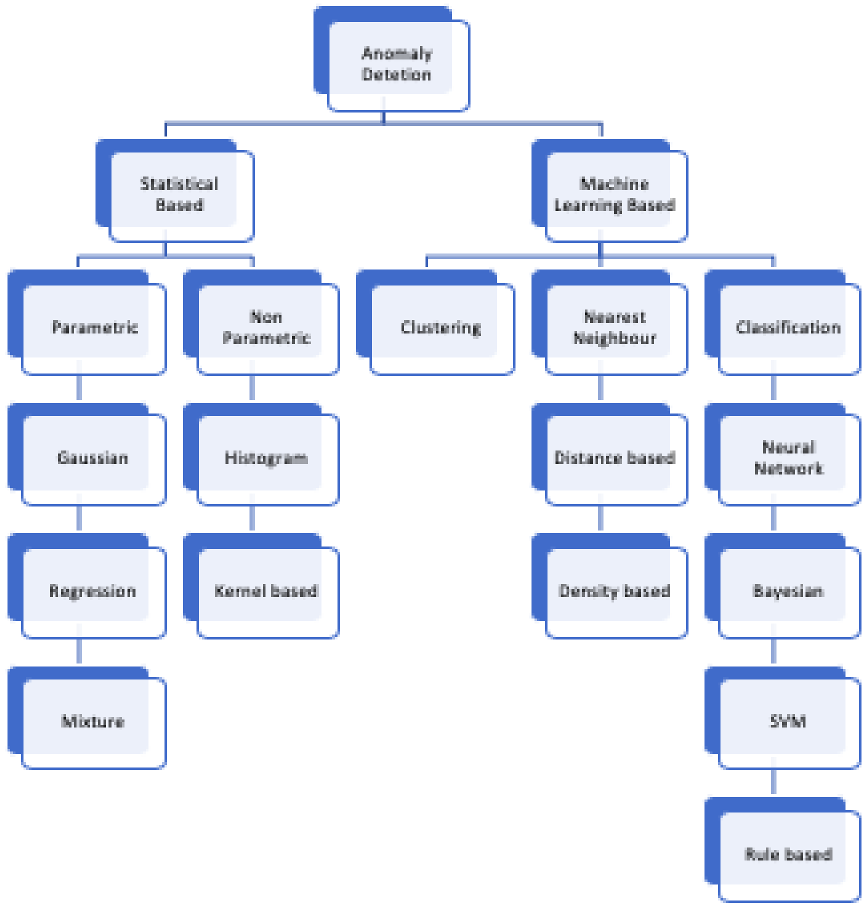

Figure 2.1: Anomaly Detection Techniques

In the healthcare field, there are certain challenges associated in the anomaly detection. One of most important challenge of anomaly detection in healthcare is reducing False Negative (FN) to make ensure that disease is detected and treated properly. In order to overcome this challenge False positive (FP) gets increased which is more fatal because it results in alarm fatigue which most of the medical practioners will ignore considering it FP. Therefore, as FN tends to 0, then FP should be maintained below alarm threshold.

2.4.5 Applications

Some of the application areas anomaly detection in healthcare are discussed briefly in this section [9].

A. Medical Image Analysis

A suggestion for making things consisted, “Existence of noise in Computer Tomography (CT) or Magnetic Resonance Imaging (MRI) images may severely impede their interpretation. ML-enabled smart analysis techniques like object detection, noise removal, feature extraction, anomaly detection and so forth can make life easier for healthcare practitioners by accelerating the data pre-processing time and fast-tracking results to the point of absolute decisiveness.”

B. Biomedical Signal Analysis

One of the applications of anomaly detection in healthcare industry includes biomedical signal analysis where temporal anomaly detection plays a major role. Temporal anomaly detection consists of two types such as point anomaly and sequential anomaly where sequential anomaly is considered most important for biomedical signals for example, arterial premature contraction using ECG traces for anomalous activity in brain, power spectral analysis in EEG for anomalous activity in neurons, arrhythmia condition from PPG, etc.,. Thus, biomedical signal analysis includes signal sources from various parts of body which contains important information about concerned body parts. Therefore, these signals require preprocessing for noise removal, filtering for compact frequency domain analysis, feature extraction and temporal/ frequency domain analysis and anomaly detection as shown in the Figure 2.3.

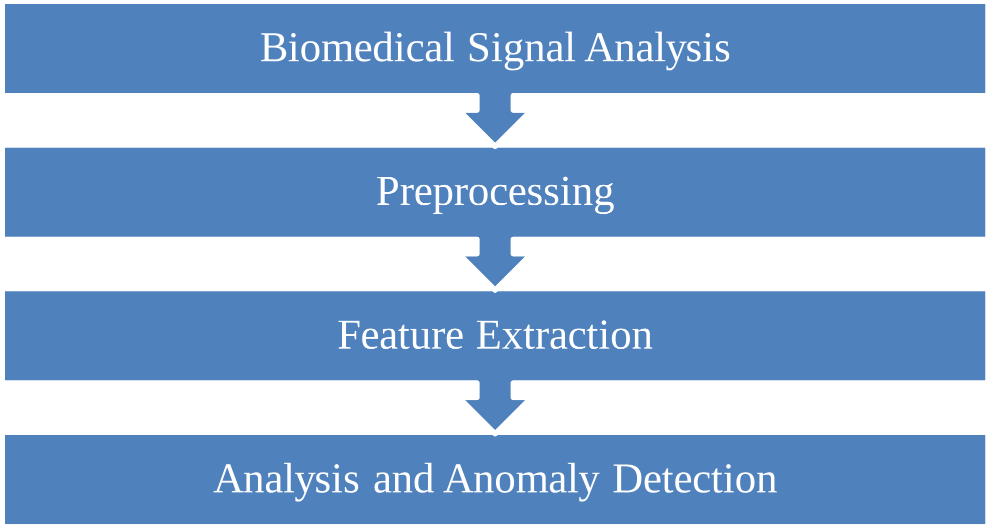

Figure 2.2: Biomedical Signal Analysis

2.4.6 Challenges

There are various challenges associated with implementing machine learning algorithms. This section briefly discusses these challenges [10].

A. Security Challenges

B. Data Provenance

C. Standards and platforms

D. Trust on Data

2.5 State of the Art: IOT and Machine Learning for Healthcare Monitoring

The main focus is providing efficient healthcare facilities to patients and saving time of not just doctors but also patients. Various remote health monitoring systems have been developed and researched.

IOT based healthcare monitoring using Wearable Body Sensors

Wearable devices like Fitbit, Mi-band, Apple watch, etc. are widely used in today's market for various purposes such as monitoring calorie count, step count, oxygen saturation level, heart rate, breathing rate, etc. and provides additional features such as clock display, reminder setup, call and text message notifications, etc. In [11]Yimeng Fang and et. al demonstrated the importance of using wearable devices for vital sign measurement. In this work, the vital signs such as ear temperature, pulse, oxygen saturation level and blood pressure were measured by both traditional methods as well as using intelligent wearable body devices. Bluetooth center and nurse digital assistant servers as integral part of this system where sensors such as infrared thermometer (IRTI 101 B), Pulse oximeter (AM-807-B-C), and arm type electric sphygmomanometer (ESM201B-01) are used for vital sign measurements. Wearable body devices are connected to the server with the help of built-in Bluetooth module and provides real-time data of patients. Android application in Nurse Dataset Administration (NDA) is used to bind patients with wearable body devices via NDA scan code and displays the record uploaded by wearable biomedical devices as well as transmits them to intelligent ward platform. Once this system is docked to hospital information system, then records can be sent to the desired location. This system utilizes Wearable Mobile Devices (WMDs) and traditional devices for same batch of patients and achieved consistent results for vital signs ($P > 0.05$). Therefore, WMDs worked much faster than traditional devices ($P< 0.01$) and achieved same level of accuracy in monitoring vital signs as traditional devices. As a result, WMDs are preferred over traditional devices for monitoring vital signs of patients.

In [12], Zia Uddin and et al. implemented an IOT- based health monitoring system for wearable biomedical device. This system has features such as displaying patient’s health status automatically and storing data on the cloud. In case of an emergency, the system provides alerts to doctors and model comprises of two sides namely, i) Web application and ii) Patient.

The main purpose of web application side is providing doctors access to real-time data belonging to multiple patients which can then be acted upon appropriately. For patients, it serves as a medium for uploading medical data in future instances where monitoring of health status would also be simple and easy. The patient side comprises of wearable body connected sensors such as temperature sensor (LM 35) and blood pressure sensor as well as an Arduino microcontroller which is used for transmission of data from sensors to a Global System for Mobile Communication (SIM 900D) every 10 minutes (min.). In case of an emergency, the microcontroller uses GSM module and sends text messages to the registered emergency contact while also generating an alarm signal.

This system provides various benefits such as minimizing patient's health risks and critical loss of time while maximizing cost-effectiveness and improving the quality of treatment offered by the doctor. Efficient collection and transmission of data notwithstanding, analyzing this data is still proving to be the need of the hour for providing accurate health-related information to doctors for appropriate action.

In [13], Mauricio and et. al proposed a system for monitoring people of health risk groups using IoT. This work mainly focuses on physiological data monitoring using wearable device and storing information in cloud. The main functionality of cloud is data processing and data storage.

This system comprises of three main parts where first part is used for collecting user's physiological data and transmitting data using Bluetooth Low Energy (BLE). The second part is responsible for capturing and storing data in cloud and the third part is web application part which provides statistical inferences of data to healthcare professionals.

This system used low cost wearable device for capturing data from users and implemented an application to validate patient's physiological data which interacts with an Android application and a Java Web application on a server in the cloud to enable data persistence in a relational database.

The main drawback of this system is lack of an API for capturing wearable device data. Most low-cost wearable devices do not have official APIs or libraries provided by manufacturers or readable documentation, causing difficulty in implementing the application.

IOT based healthcare monitoring using Body Sensor Network:

The body sensor networks (BSN) are also used for healthcare monitoring along with IoT is demonstrated by Snehal Sanjay Kale and DS Bhagwat who proposed a Secured IoT Based Web care Healthcare Controlling System using BSN [14] .The main aim of this work is transmitting patient's health parameters from remote device to cloud server. The sensors used in their project are heart rate sensor, body temperature sensor (DS18B20) and fall detection sensor (ADXL 345). Raspberry pi is used to configure the sensor and send data to the cloud. They have provided security using AES and used Error Detection to avoid data loss in cloud but still there is no provision for data analysis. The data analysis can provide a complete description and is useful for doctors or medical professionals to provide medications accordingly.

IOT based healthcare monitoring using Smartphone Interface:

RHM systems also uses smartphone interfaces to monitor health status of patient thereby solving certain problems such as unnecessary hospital visits, long waiting time, doctor's unavailability, etc. In [15], Akshat and et. al proposed a smart healthcare monitoring system using smartphone interface for remotely monitoring health status of patients as well as addressing the problem of personnel shortage. This work provides an appropriate framework for real-time monitoring and analyzing vital signs using IoT and Zigbee protocols. Vital signs such as heart rate, blood pressure, heart rate beat, blood glucose level and ECG data are monitored using sensors such as temperature sensor (LM 35), Glucose Level Detector, Heartbeat sensor and Galvanic Skin Response (GSR).

There are mainly three important phases of this system in which the first phase deals with connection of sensors to patient and data transmission using Zigbee to the server. Second phase involves data update in the server once all the parameters are properly received and third phase is associated with android application that fetches data from server and displays on smart phone application. Therefore, the proposed system successfully measured and monitored vital health parameters using low-cost effective solution. Thus, this system demonstrates the importance of smartphone interface in RHM systems.

IOT based healthcare monitoring using Machine Learning:

Machine learning has revolutionized the healthcare industry where ML in integration with IOT is useful in determining the diagnosis and treatment along with remotely monitoring health status of patients. In [16], "Rani J. Utekar and Jayant S. Umale" proposed an Automated IOT based healthcare monitoring for Remotely located patients where implementation of this system helps doctors and guides them along with providing E-mails in case of abnormal conditions. This implemented system also provides an interface among doctors, nurses and concerned relatives of patients.

This system's major purpose is to remotely monitor health status of a patient, simultaneously alert if heart rate goes in abnormal values and provides data analysis using supervised learning models such as SVM, KNN, Naive Bayes and J48 classifier. The system generates alert when result of decision support system shows the critical condition of patient.

In this system, remote patient healthcare monitoring is achieved with developing website and storing real time heart rate data on web server's database with IoT and provides prediction of heart disease with supervised machine learning classification techniques where J48 classifier performed better (92% accuracy) on the basis of predefined dataset. The probability of predicting heart problem will increase with generating rules from UCI heart database with decision tree algorithm. Thus, this system can be uses UCI heart rate dataset and provides an example of ML models in integration with IOT for healthcare monitoring.

In , [17]"A.Raji and et. al" proposed a system for Chronic disease monitoring using IoT for vital sign monitoring using wearable devices such as heart rate, blood pressure and temperature sensor and the patient can determine these vital signs in the form of text document without the need of nurse. This system is trained for these vital sign data using data mining techniques (Support Vector Classifier, J48, and Naïve Bayes algorithms) and enables high risk patients to be timely checked in order to improve their quality of life.

This system is primarily designed for elderly patients where raw data is collected from aforementioned sensors and model returns real-time data which can be healthy or unhealthy for particular patient.

Main concentration of this system is towards problems related to health parameters such as Electrocardiogram (ECG), Oxygen Saturation (SpO_2), Heart Rate (HR), Photoplethysmography (PPG), Blood Glucose (BG), Respiratory Rate (RR), and Blood Pressure (BP). For testing purpose, this system used instances of raw data from any patient.

For data processing, data mining approaches such as SVM, Naïve Bayes, J48 classifier are used where SVM was performing better having accuracy of 92%.

This system captured the vital sign data for patients using medical sensor and noise from the raw data is removed. Afterwards, this raw data is used as test data in preprocessed step. Model training are learned by supervised methods (SVM, Naïve Bayes and J48). The model returned the vital sign data as healthy or unhealthy. From this status, patient takes medicine to detect early disease.

The main drawback of system is classification of large amount of data. Therefore, in that case Big data techniques and Map reduce techniques will be used to find health status as well as dependency can be found for further analysis.

2.6 Conclusion

This chapter aims to provide literature review of our thesis and explains the need and importance of RHM devices. The main focus is towards technologies promoting RHM devices such as IoT, ML, Anomaly Detection. Therefore, brief insight about these technologies along with applications and challenges are discussed in relation to healthcare industry. In recent days, lots of researchers and companies are focused on RHM as it is growing market and makes it easier to provide medical facilities remotely. Thus, some of their contribution in accordance with our work is also discussed in this chapter. However, the next part of this work will provide insights about the practical design and implementation of our model.

Chapter 3 - Design and Implementation

3.1 Functional Block Diagram

This chapter provides an insight about the designing and implementation of the proposed model of this thesis. The designing involves prototype for data collection using IoT devices and implementation involves various ML algorithms used in our thesis. The functional block diagram of the work implemented in our work is being displayed as shown in the Figure 3.1. It begins with the user interface which is used for data collection. In our work, we are dealing with health-related data. This data is being collected using two devices. These two devices then transmit the data to the intermediate devices using Bluetooth as the transmission medium. This data is then stored in the intermediate device and further transmitted to the cloud via Wi-Fi as the transmission medium. This data from cloud is further extracted to the remote server for analysis. In our work, analyzing data involves erroneous data detection. The analysis is performed using ML techniques and the processed data is stored in SQL database located in remote server.

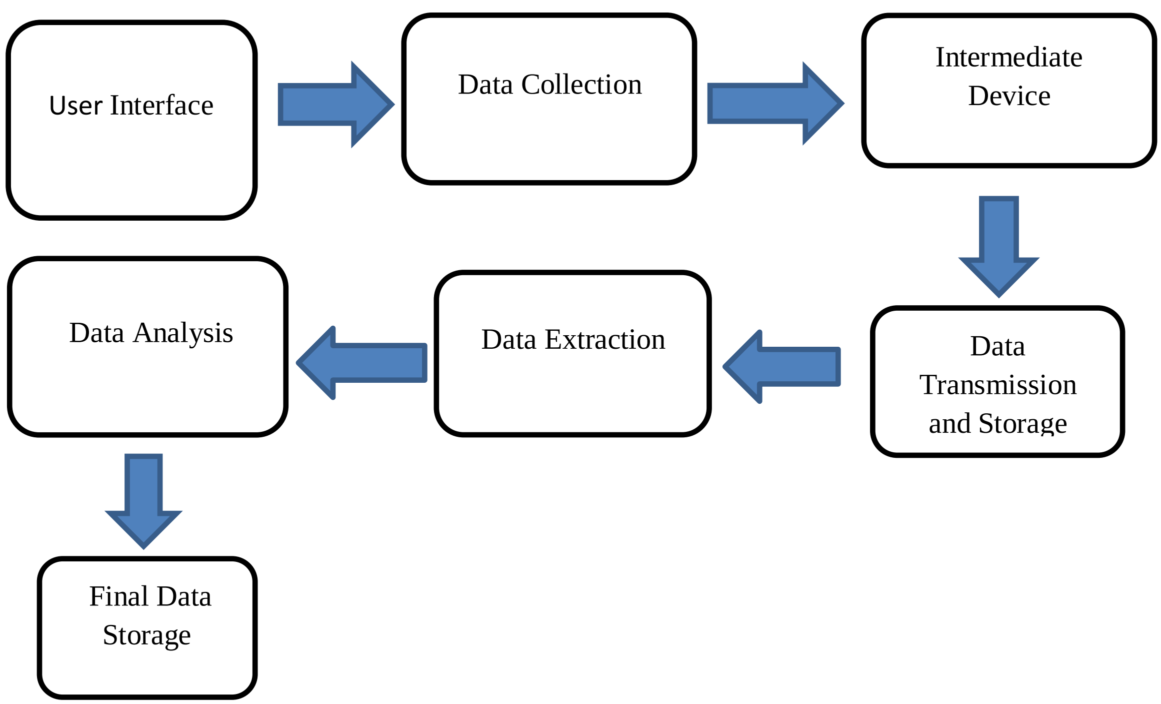

Figure 3.1:System Functional Block Diagram

3.2 Phases of The Design

This section explains about the proposed methodology of the thesis. The block diagram shown in Figure 3.2 depicts the three main phases of design prototype. These three central phases as shown in the figure have different functionalities.

Figure 3.2: Phases of the Design

The first phase of the work is the sensing period in which the data is being collected. The task of data collection is performed in two steps. The first step in data collection involves a device to be used for data collection. In our work, we are making use of IOT devices, namely spire stone and iHealth Sense.

The second phase includes the communication phase. This phase also contains three essential steps. The first step comprises establishing a communication between the devices and the medium for controlling the device, i.e., smartphone. In our system, we are using Bluetooth Low Energy (BLE) technology. The second step includes establishing a communication to transmit the data from the interface, i.e., mobile phone to the cloud, and in our work, we are using Wi-Fi interface. The last step includes sending the data from the cloud to the server using either the concept of open API's or using the file transfer protocol.

The third phase includes data processing phase, which deals with the data analysis. The primary purpose of data analysis for our work is to detect anomalies in the collected data. Outlier detection depends on the type of information being collected. There are three main types of outliers, such as global outliers (point outlier), collective outliers, and contextual outliers. These outlier detection methodologies will be done by using ML. The last step of the proposed method includes the processed data storage in the SQL database.

3.2.1. Sensing Phase

The sensing phase deals with the physical interaction of the devices with the patient whose measurements are being taken. The main aim of this work is to provide Remote Monitoring (RM) of the patient. This phase can be further divided into two parts. The first part deals with the appropriate selection of the medical device for measuring the patient. In the modern era, a lot of accessories are available in the market, which can be used for monitoring the health of a human being. These devices can measure heart rate, calorie count, step count, oxygen saturation level (SPO_2). Most of the methods available in the market are not accurate as these devices are of low cost and can be used as a reference to check the health and well-being of the person. The comparison of some of those devices is shown in Table 3.1.

Table 3.1: Device Comparison

S.NO	Devices	Advantages	Limitations
1	Fitbit –Charge 3	Easy to set goals, constant reminder of commitment to goals,	Only for HR measurements
2	Fitness Tracker	Personal accountability, individual-tailored goals, cost effective, up-to-date weight loss tool	Connectivity Issues
3	Apple Watch	Accurate fitness tracker, telephone calls, mapping, streaming music	Compatible with iPhone
4	Garmin Watch	GPS tracker, Heart rate monitoring, long lasting battery power	Expensive as compared to Fitbit, Fitness tracker, and Apple Watch

Therefore, the devices used for this work are Spire Stone for measurement of respiration rate and iHealth Sense for measurement of BP (systolic and diastolic) & HR. The second step involved in this phase deals with the actual collection of data and transmitting the collected data to the mobile phone, which acts as an intermediate system between the patients and the cloud. The collection of data from the respiration devices considers the compressions and expansions of the torso. The respiratory equipment is a hardware device which must be placed at chest on the patient for collection of data. These novel sensors in this device are patented sensors and uses ML algorithms. The process involved in collection of data from iHealth Sense involves a wireless wrist BP monitor. This BP device is monitored wirelessly from the mobile phone for the collection of the data. This device, while collecting data contracts the hand and after the readings are being taken, the device automatically loses its grips. In the current generation, it is recommended to take at least 2-3 BP and HR readings in a single day. These two devices work simultaneously, and the data is further processed to the next stage. The next step involves establishing a connection between the devices and the mobile phone and between the mobile phone and cloud.

3.2.2 Communication Phase

3.2.2.1 Overview

The communication phase deals with establishing a communication between two systems taken into consideration. This work concentrates on two parts of the communication. The first part includes making a secure communication between the devices in the sensing phase and one intermediate device which is used for monitoring the data of the instruments taken into consideration. The communication protocol used for establishing this connection is known as " BLE." The second part of this communication establishing a proper communication channel between the mobile phone and cloud which is used as an intermediate device for controlling BP readings and HR readings, and while measuring RR, it acts a medium of verification of the values being received at the server. The functional working of the communication that is involved in this work is being illustrated in Figure 3.3 and Figure 3.4.

Part A: Respiration System Setup

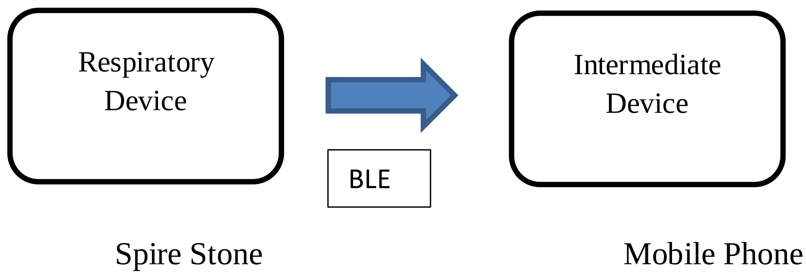

Figure 3.3: Respiration System Setup

Part B: Blood Pressure and Heart Rate System Setup

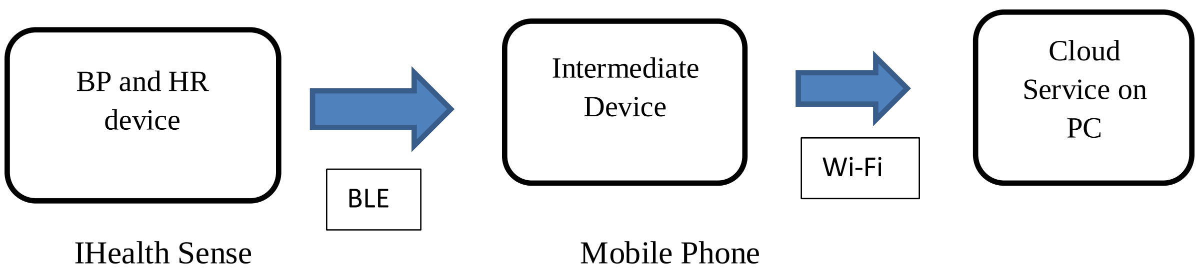

Figure 3.4: Blood Pressure and Heart Rate System

3.2.3. Data Processing Phase

3.2.3.1 Overview

The data processing phase is the final stage of this work. The basic working of this phase is to deal with analysis of the data being obtained from communication phase. The data that is being obtained from communication phase contains some values that deviate from rest of the values. These values are being referred to "outliers" in this context. This work primarily focuses on data related to body vitals (i.e., BP {sys, dia}, HR and RR). The outliers are considered very important in the medical field. They can be used by doctors to see the medical condition of the patient. In this section, different ML algorithms taken into consideration are discussed. The ML algorithms are used in the work as it allows to reduce human effort and with more accurate results. The data processing phase is further divided into two main parts. The first part includes arranging the raw data for processing. This is also referred to as data preprocessing. The data preprocessing is required because data is being obtained from different sources. The RR data is coming from respiratory device using the concept of Oauth2.0 and the device used for respiration rate is Spire Stone. The HR and BP data is coming through the cloud service. The device used for HR and BP data is IHealth Sense. The working diagram is shown in the Figure 3.5.

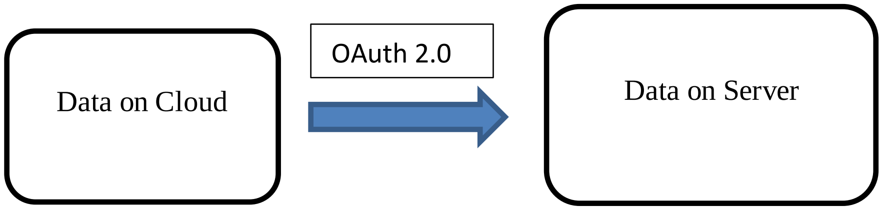

Figure 3.5: Data Processing

3.2.3.2. Methods Implemented

A. Autoencoder

The Autoencoder (AE) is a type of neural network. It plays an important role dimensionality reduction. In the case of anomaly detection, it tries to find an optimal subspace [18] . We can assume the normal training set as $\{x_1, x_2, x_3, \dots \dots \dots \dots . x_n\}$ where each of them represents a d dimensional vector. The training phase consists of constructing a model to project these data into low dimensional subspace and reproduce the data to obtain the output $\{x_1, x_2, x_3, \dots \dots \dots \dots . x_n\}$. The reconstruction error is defined by the formula given below:

$$\varepsilon(\mathrm{x}_{\mathrm{i},}, \mathrm{x}_{\mathrm{i}}') = \sum_{\mathrm{j}=1}^{\mathrm{d}} (\mathrm{x}_{\mathrm{i}} - \mathrm{x}_{\mathrm{i}}')^2 \qquad (3.1)$$

The basic architecture of AE consists of an encoder and a decoder. Figure 3.7 shows a simple AE having an encoder, decoder and a hidden layer. In the encoder section, as depicted by Figure 3.6, the input vectors ($x_i \in R^d$) are compressed to develop hidden layer. The activation of neurons is given by:

$$\mathrm{h}_{\mathrm{i}} = \mathrm{f}_\theta(\mathrm{x}) = \mathrm{s}\left(\sum_{\mathrm{j}=1}^{\mathrm{n}} \mathrm{W}_{\mathrm{ij}}^{\mathrm{input}} \mathrm{x}_{\mathrm{j}} + \mathrm{b}_{\mathrm{i}}^{\mathrm{input}}\right) \qquad (3.2)$$

Thus, input vector is encoded to low-dimensional vector. In the decoder section, the hidden representation h_i is decoded back to R^d . The mapping function is given by:

$$x_i' = g_{\theta'}(h) = s\left(\sum_{j=1}^{n} W_{ij}^{hidden} h_j + b_i^{hidden}\right) \quad (3.3)$$

The AE is optimized to minimize the average reconstruction error with respect to θ' and θ given by:

$$\theta^*, \theta'^* = \text{argmin}_{\theta,\theta'} \frac{1}{n} \sum_{i=1}^{n} \epsilon(x_i, x_i') \quad (3.4)$$

$$= \text{argmin}_{\theta,\theta'} \frac{1}{n} \sum_{i=1}^{n} \epsilon\left(x_i, g_\theta'\left(f_\theta(x_i)\right)\right)$$

Where ε is defined as the reconstruction error.

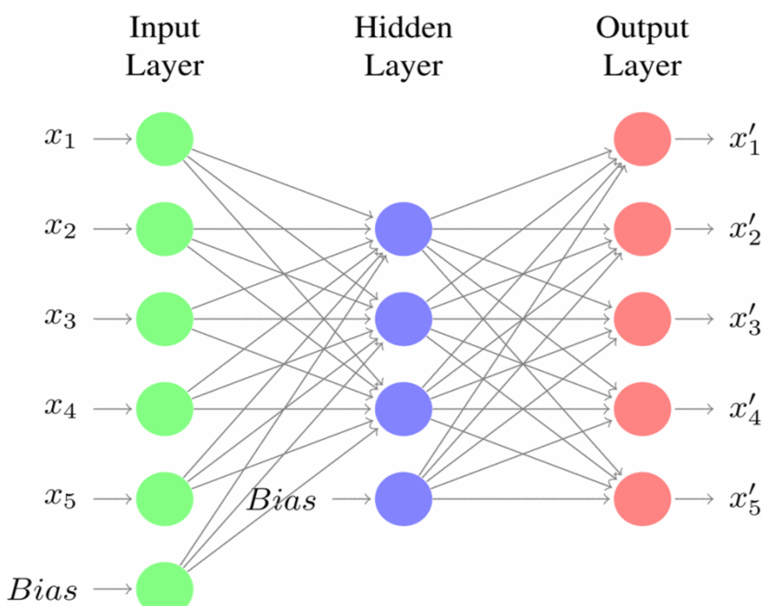

Figure 3.6: Autoencoders [18]

B. KNN

The KNN is often called as lazy learner [19]. This algorithm stores and classify new cases on the basis of similarity measure. It is a non-parametric classification method which is classified into two types

1) Structureless NN techniques
2) Structure-based NN techniques

The structureless NN classifies whole data into training and test sample data. The structure-based NN are based on the structure of data such as orthogonal structure tree (OST), ball tree, k-d tree, axis tree, and various others. KNN stores the training examples and waits for the test examples. In this method, all training tuples are stored in n-dimensional space. It starts working once the tuples from test data are provided and search for the k training tuples closest to the unknown tuple. Therefore, in order to classify unknown record, it computes the distance between other training records. Based on the distance, the class labels and k nearest neighbors are identified. Figure 3.7 helps us explain how KN works.

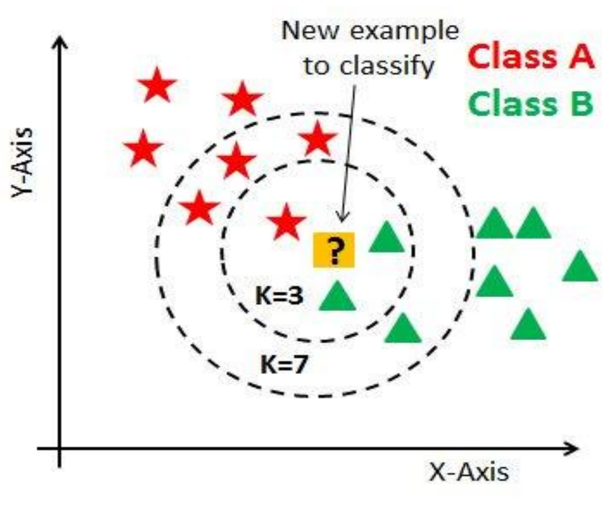

Figure 3.7: KNN Algorithm [19]

The KNN computes distances using three different methods but the most commonly used method is Euclidian distance. The distance metric using all three methods are explained further.

a. Euclidean distance

$$d(x_i, y_i) = \sqrt{\sum (x_i - y_i)^2} \tag{3.5}$$

b. Manhattan distance

$$\sum_{i=1}^{k} |x_i - y_i| \tag{3.6}$$

c. Minkowski distance

$$\left(\sum_{i=1}^{k} (|x_i - y_i|)^q\right)^{\frac{1}{q}} \tag{3.7}$$

Figure 3.7 depicts basic working KNN algorithm. In this figure, suppose we want to classify if a new data point belongs to class A or class B then, initially we will randomly choose the value of k as three. This chosen value of k will look for the nearest 3 data points and tries to identify their respective classes. Thereafter, it will calculate the distance of that new data point from all other data points. The distance can be calculated either by using Euclidean, Manhattan, or Minkowski formulas. The most commonly used is Euclidian distance. The Euclidean distance is calculated by taking two data points and then using Pythagorean formulae to find the distance between those two data points. In this algorithm, the decision boundaries are not explicitly computed whereas Voronoi diagram visualization is used for it [20]. Voronoi diagram visualization is basically partitioning of plane into specific regions based on distance between them. One of the examples of decision boundary is given by Figure 3.8.

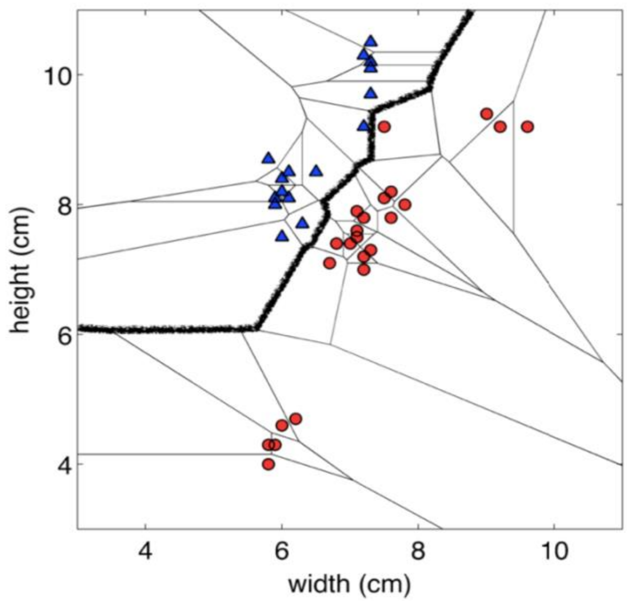

Figure 3.8: KNN Decision Boundary[20]

In KNN the best way to choose the value of $[k]$depends on following factors:

a. The value of k should be set larger i.e., the k in KNN should be greater.

b. Using cross-validation to find the value of $[k]$

c. The rule of thumb is $k < (sqrt(n))$, where n is the number of training examples.

There are some advantages and disadvantages associated with using KNN algorithm.

Advantages

a. Effective for large training data

b. Robust to noise

Disadvantages

a. High Computational Cost

b. Distance metric for calculation is not specified

c. Sensitive to labelling.

d. Sometimes results lead into overfitting the data.

C. Support Vector Machine

The SVM algorithms can be used for studying linearly separable sets. However, for non-linearly separable sets, a non-linearity mapping is required to change inseparable sample of low-dimensional space to high-dimensional space [21]. It is also used for structural risk management for finding optimal separating hyperplane. The principle of SVM will be discussed considering three different cases.

a. Linearly Separable Case:

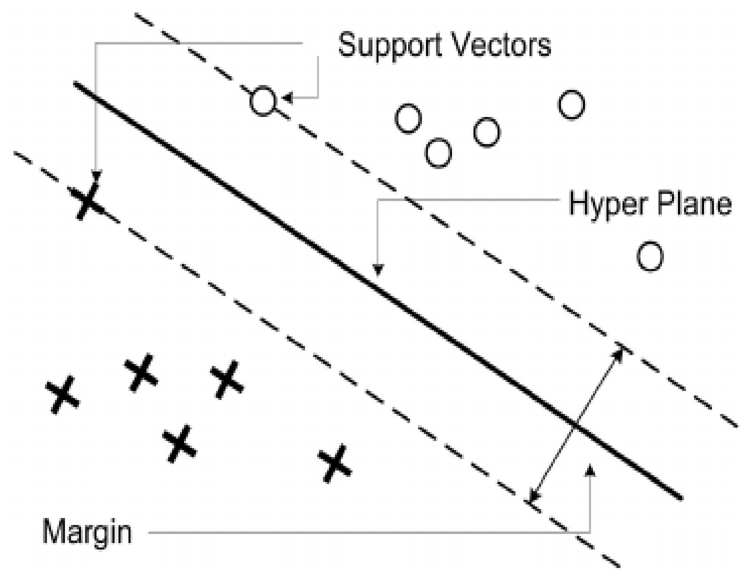

Figure 3.9: Linearly Separable SVM [21]

Figure 3.9 shows the case of linear classification. It depicts two lines separating the two different types of samples. The distance between these two straight lines is called margin. The line separating the two straight lines is often called hyperplane. The two different samples are called support vectors. The classification line is expressed by using the equation $w * x + b = 0$ The linear sample set is expressed as: (x_i, y_i)

$$\mathrm{i} = 1,2, \ldots\ldots \mathrm{n}, \mathrm{x} \in \mathrm{R}^{\mathrm{d}}, \mathrm{y_i} \in \{1, -1\} \tag{3.8}$$

The classification line is normalized to satisfy the condition:

$$\mathrm{y_i}[(\omega, \mathrm{x_i}) + \mathrm{b}] - 1 \geq 0, \mathrm{i} = 1,2, \ldots\ldots \mathrm{n} \tag{3.9}$$

The class interval is given equal to $\frac{2}{||w||}$. For maximum class interval it is equivalent to $\frac{1}{2}$ $\|w||^2$.

b. Linearly Inseparable Case:

There are some classification problems which are linearly inseparable. Therefore, their hyperplane is limited. For linearly separable problems, a slack variable ε is introduced and constrain equation becomes:

$$y_i[(w.x_i + b)] \geq 1 - \epsilon; i = 1,2, \ldots \ldots n \tag{3.10}$$

The objective is given as:

$$L(\omega, \varepsilon) = \frac{1}{2}||\omega||^2 + C\,|\textstyle\sum_{i=1}^{n} \varepsilon_i|, i = 1,2, \ldots . n \tag{3.11}$$

c. Nonlinear Case:

There are two different types of nonlinear case. One of them is Nonlinear SVM while the other one is based on kernel function. In Nonlinear SVM, a nonlinear mapping of ϕ is performed to high-dimensional or infinite dimensional feature space in order to apply linear SVM to solve non-linear problems. This method is given by Figure 3.10.

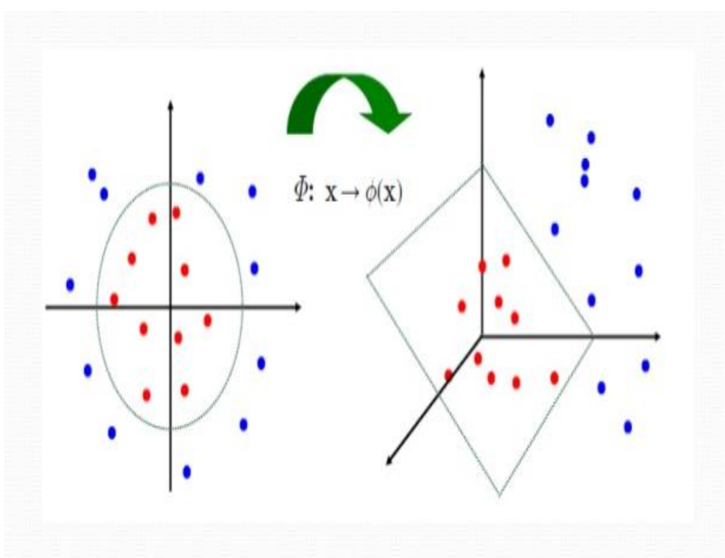

Figure 3.10: Non-Linear SVM [21]

The kernel function is defined as $K(x_i, x_j) = \emptyset(x_i).\emptyset(x_j)$. The functional theory tells that if kernel function satisfies Mercer condition, it corresponds to inner product in transformation space. There are three types of kernel functions in SVM, namely linear, polynomial and Radial Basis Function. They are explained below.

a. Polynomial kernel:

$$k(x, x_i) = (\gamma x. x_i + r)^d \tag{3.12}$$

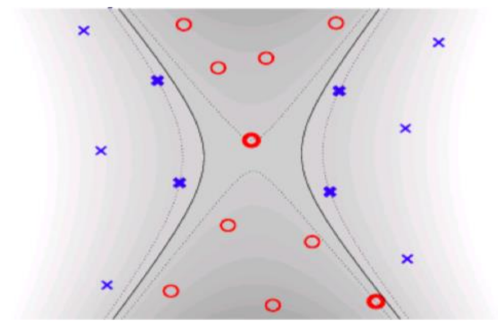

Figure 3.11:Polynomial Kernel [22]

b. RBF kernel:

$$k(x, x_i) = \exp\left(-\frac{1}{2\delta^2}||x - x_i||^2\right) \tag{3.13}$$

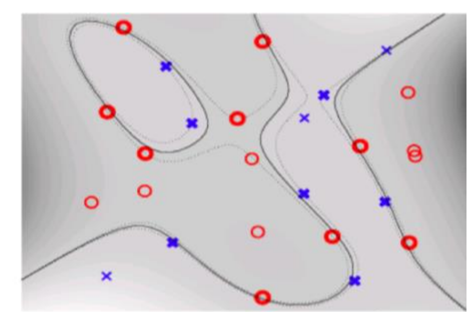

Figure 3.12: RBF Kernel [22]

c. Sigmoid kernel:

$$k(x, x_i) = \tanh[\gamma(x.x_i) + r] \tag{3.14}$$

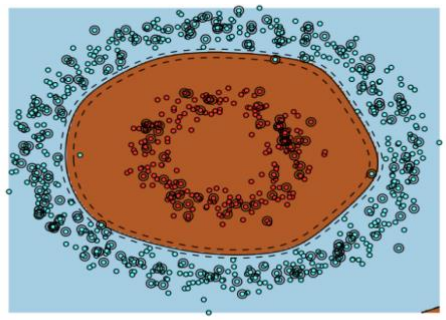

Figure 3.13:Sigmoid Kernel [23]

In this section, the methods implemented in our work were discussed in detail. It dealt with briefly discussing about the working of Supervised models (SVM, KNN) as well as Unsupervised model (AE) and also provides mathematical formulations of the models taken into consideration. The mathematical formulation of Unsupervised model (AE) discusses the role of hidden layer in formulation of relationship among different parameters taken into consideration whereas mathematical formulation of Supervised models (SVM, KNN) discusses their working. The design implementation of this thesis will be discussed further in the next section.

SVM can be used to avoid the difficulties of using linear functions in the high-dimensional feature space, and the optimization problem is transformed into dual convex quadratic programs [24]. KNN is more robust towards noisy training data and more effective towards large training data as compared to other models [25].

3.3. Design Prototype

3.3.1. General Design

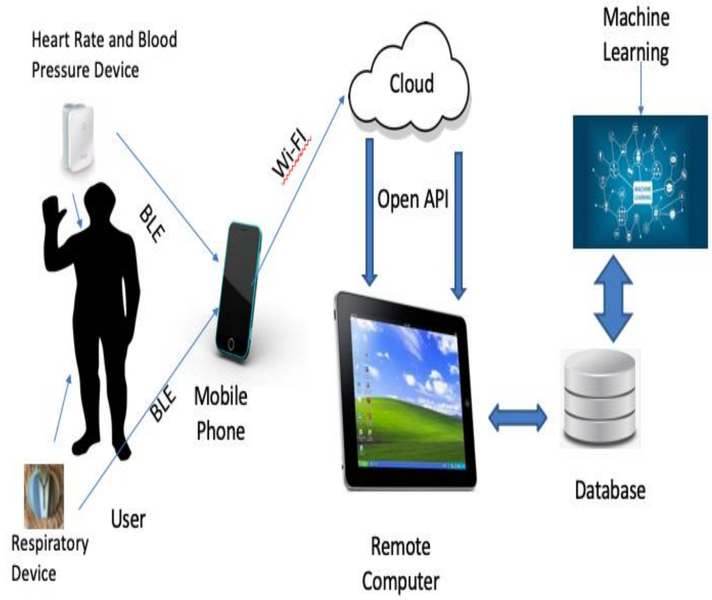

Figure 3.14:General Architecture

The general architecture of our design is depicted as shown in Figure 3.14 and begins with the data collection from the user. This design is using two different devices (Spire Stone and iHealth Sense) for data collection. The spire stone is attached to user body on torso for collecting respiration rate while IHealth sense is attached to human wrist for collecting blood pressure readings as well as heart rate readings. These two IOT devices transmit data using BLE to the intermediate device and data is stored in it. This stored data is transmitted to cloud using Wi-Fi and then extracted for performing data analysis. The data extraction is done using the concept of Open API where python script is created and using access token the data is transmitted over to the remote computer where a SQL database is created for storing unprocessed data. This unprocessed data is then further analysed using two approaches. These two approaches consider combination of unsupervised and supervised learning algorithms for detecting erroneous data. At the final stage, the processed data is again stored on SQL database on the remote computer.

3.3.2 Implementation Design

This section deals with the implementation design of the work. The design prototype consists of main two devices which are used for measuring body vitals and intermediate device used is mobile phone. The output is generated using ML algorithms after data preprocessing. The detailed explanation is provided further in this section.

Part A: Respiration System

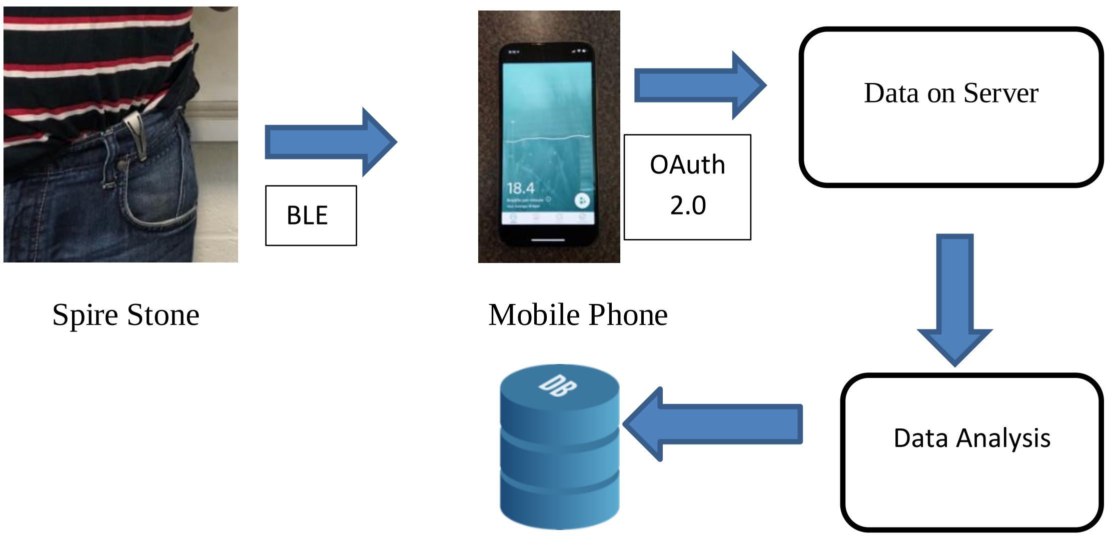

Figure 3.15:Respiratory System Design Implementation

The Respiratory System Design begins with measuring the data using Spire Stone. The spire stone has inbuilt patented sensors for measuring the RR. This device further transmits the data to the intermediate device (mobile phone) and the data from the device can be extracted on the server using the concept of APIs and in our work, we are making use of Spire API. The concept behind extracting data from APIs involves OAuth 2.0. Once the data is being extracted on the server it is further sent for preprocessing. The preprocessing in our work involves arranging the raw data. The raw data being received is in Json format which is further converted into csv format using standard python functions. Once the data is arranged properly it is being processed for detecting outliers using AE, KNN and SVM. The outliers are considered very important in the medical field. They can be used to detect the condition of the patient taken into consideration. The general framework of OAuth 2.0 being used for extracting data on the server is explained in the Figure 3.16 [26].

```
+--------+                               +---------------+
|        |--(A)- Authorization Request ->|   Resource    |
|        |                               |     Owner     |
|        |<-(B)-- Authorization Grant ---|               |
|        |                               +---------------+
|        |
|        |                               +---------------+
|        |--(C)-- Authorization Grant -->| Authorization |
| Client |                               |     Server    |
|        |<-(D)----- Access Token -------|               |
|        |                               +---------------+
|        |
|        |                               +---------------+
|        |--(E)----- Access Token ------>|    Resource   |
|        |                               |     Server    |
|        |<-(F)--- Protected Resource ---|               |
+--------+                               +---------------+
```

Figure 3.16:OAuth 2.0 framework [26]

The framework shown above consists of Client and the Server. The Client in our work is the spire stone and server are the computer for storing the data. There are 3 steps involved. The first step involves resource owner in which authorization request is being sent by client and the resource owner provides the authorization grant on the basis of authorization request. The second step involves authorization server where authorization grant is being sent by the client to the authorization server and authorization server provides access tokens. The final step involves resource server which provides protected resource for using these access tokens. The design prototype of the second part is discussed further in detail. The second part of the design deals with measuring HR and BP.

Part B: Blood Pressure and Heart Rate System

This part deals with measuring BP and HR readings of a person. These two parameters play an integral part in determining the health of a person. This work considers IHealth sense for measuring the HR and BP. This device is a wireless wrist monitor. The main focus of this work is to detect outliers in the readings obtained from the devices which are useful for doctors in monitoring the condition of the patient. The system design of BP and HR is discussed below.

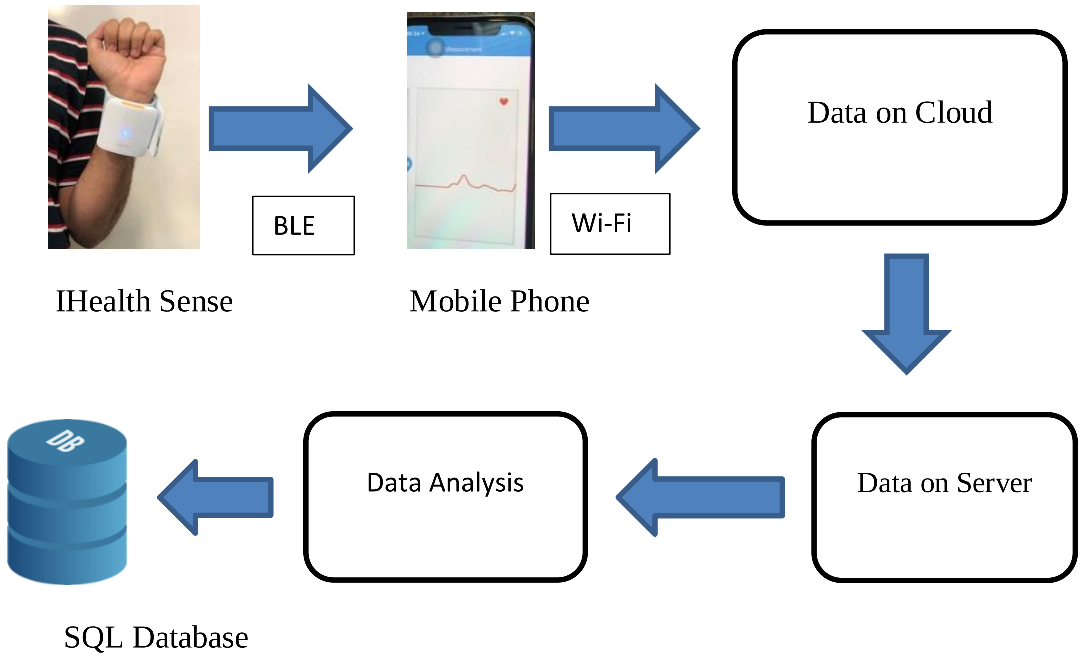

Figure 3.17:Blood Pressure and Heart Rate Design Implementation

3.4. Device Specification

3.4.1 Blood Pressure Device

The specifications [27] of BP Device- iHealth Sense BP7 are provided in the table.

Table 3.2: Blood Pressure Specifications

Parameters	Specifications
Connectivity	Bluetooth 3.0 + EDR
Dimensions	72mm x 74mm x 17.6mm
Cuff Circumference	13.5 – 22 cm
Weight	105g
Method of Measurement	Oscillometric, Automatic Inflation and deflation
Power	DC: 5V, 1A, 1x 3.7 V, Li-ion: 400mAh
Battery Life	80 (full battery)
Measuring Range	SYS: 60-260 mmHg, DIA: 40- 199 mmHg, Pulse: 40-180 pulse/min
Measurement Accuracy	Pressure: +/- 3mmHg, Heart Rate: +/- 5%
Conditions for use	Temperature: 5°C- 40°C, Humidity: < 90% RH
Storage and Transportation	Temperature: -20°C-55°C, Humidity: < 90% RH

3.4.2. Respiration Device

The specifications [28] of the respiratory device are provided by the table.

Table 3.3: Respiration Rate Specifications

Parameters	Specifications
Battery	Technology: Li-ion
	Standby Time: up to 7 days
	Recharge Time: 180 min
Connectivity	Bluetooth
General	Product type: Activity tracker
	Supported Host Device OS: iOS, Android
	Tracking Data: Calories burnt, activity, steps taken, breath, state of mind
	Set Goals: Yes
	Color: Stone
	Features: Inactivity alert, motion sensor
Dimensions	Length: 1.2 x 0.6 x 1.7 inches
	Weight: 0.8 ounces
Sensors	Patented Respiratory sensor
	3-axis accelerometer
	Vibration monitor

3.5 Device Working:

3.5.1. Blood Pressure and Heart Rate Device

The BP and HR device used in the work works on the basis of Non-invasive BP (NIBP) Measurement by Oscillometric Principle. This device is used to obtain sys and dia BP along with HR. The Oscillometric method came into existence in 1876 and involves observation of oscillations [29]. The basic working of NIBP devices is being presented by Jaafar e.t al [30]. The authors demonstrated the concept by means of a figure 3.20.

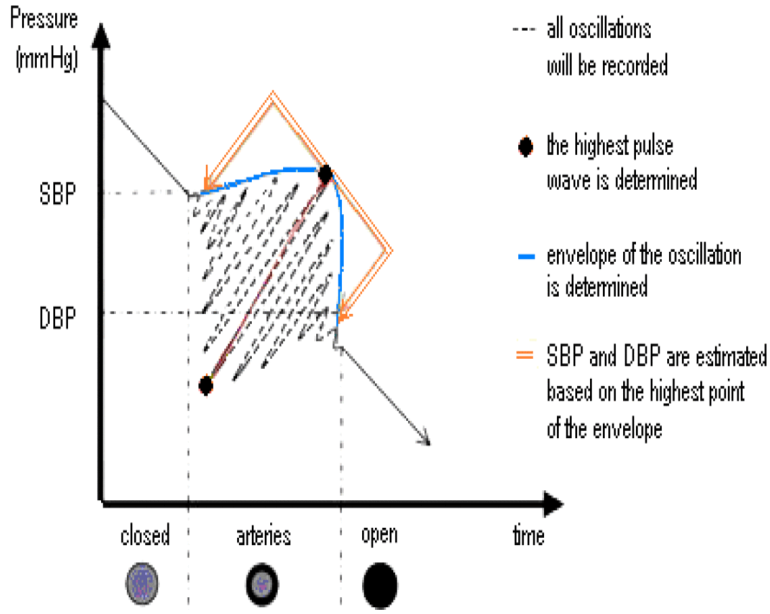

Figure 3.18:State of arteries[28]

The figure above illustrates the state of arteries and the pressure oscillations. The pressure oscillations are calculated during deflation of cuff pressure and recorded which is used for estimating SBP and DBP. These oscillations have a typical curve when blood flow stops temporarily interrupted and then starts flowing again. During the normal flow of the blood, these oscillations becomes stronger and diminish. The frequency of heart beat are represented by these oscillations. The peak to peak pressure oscillations determines pulse rate. The values of SBP, DBP, and pulse rate are displayed on the display panel. Finally, methods for NIBP are shown in the Table 3.4.

Table 3.4: NIBP Principles

Method	NIBP Principles
Palpatory	Cuff pressure equals SBP
Auscultatory	Based on Sound Waves
Ultrasonic	Based on frequency difference between transmitted and reflected ultrasounds
Tonometry	Surrounding Pressure equals artery pressure during partial collapse of blood vessel
Oscillometric	SBP and DBP using empirical algorithm

3.5.2. Respiration Rate Device

The Schematic diagram of the RR System is shown in the form of figure 3.20 as shown below.

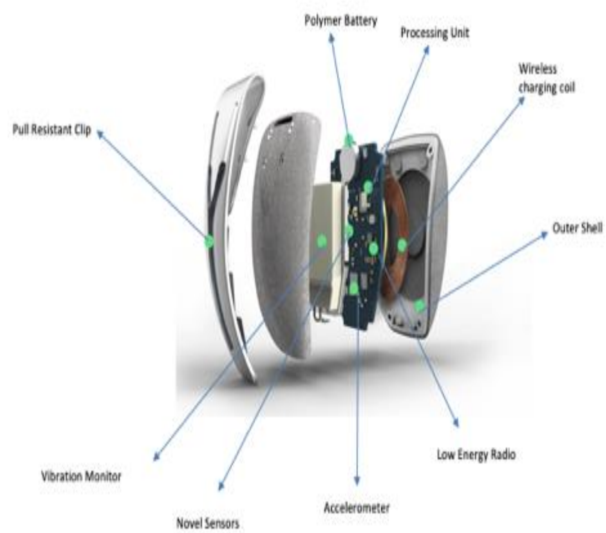

Figure 3.19:Respiratory Device

The various functional parts of the device are as follows [31].

1. Pull Resistant Clip: It helps to keep the stone close to the body.
2. Vibration Monitor: It helps in providing information about adjustment of breathing.
3. Novel Sensors: These are patented sensor which measures contraction and expansion of torso to measure breathe.
4. Accelerometer: It is useful for tracking activity, steps, and calorie count.

5. Low Energy Radio: It uses Bluetooth technology to pair the device with the phone.

6. Outer Shell: It is durable during wear and tear.

7. Wireless Charging Coil: It is useful for charging spire stone wirelessly.

8. Processing Unit: It uses spire algorithms for monitoring breathing patterns

9. Polymer Battery: It is the battery used in the device which can lasts up to 7 days.

3.6. Summary

In this chapter of our work, designing and implementation of the prototype for remote health monitoring is discussed in detail. Designing part not only discusses IoT devices used in our work for data collection but also various other devices with their limitation and advantages. Implementation section discusses about the ML algorithms along with their mathematical formulation. Finally, the working algorithm of IoT devices is provided with their specifications. Implementation results of ML models (Supervised and Unsupervised) considering accuracy as the prime factor for classification will be discussed in next chapter of the thesis.

Chapter 4 - Results and Discussion

4.1 Preprocessed Data

The preprocessed data is often called as raw data. This raw data is being obtained from the server for further analysis. In this work, the information is first collected from the IoT devices and then it is sent to the intermediate device, i.e., mobile phone. Afterward, it is being stored in the cloud using Wi-fi and gets transferred to the server on a remote computer. The preprocessed data in our work considers different parameters such as systolic blood pressure, diastolic blood pressure, heart Rate, and respiration Rate. This data examines 1000 samples. Some of the examples of preprocessed data is shown in the Figure 4.1.

Respiration Rate	SYS	DIA	PULSE
18.0	142.0	76.0	86.0
15.0	125.0	76.0	83.0
17.0	125.0	78.0	85.0
17.0	124.0	80.0	85.0
14.0	123.0	89.0	81.0
19.0	120.0	92.0	87.0
15.0	123.0	75.0	83.0
15.0	132.0	99.0	83.0
15.0	127.0	86.0	83.0
14.0	127.0	94.0	81.0
16.0	139.0	99.0	85.0
15.0	128.0	94.0	82.0
15.0	124.0	86.0	85.0
14.0	122.0	81.0	80.0
14.0	116.0	89.0	81.0
12.0	116.0	86.0	81.0
14.0	119.0	90.0	80.0
16.0	117.0	84.0	82.0
14.0	117.0	84.0	82.0
13.0	117.0	84.0	82.0
22.0	145.0	95.0	105.0
20.0	145.0	95.0	105.0

Figure 4.1 : Preprocessed Data

4.2 Performance Metrics

In our model, we are using two different supervised learning techniques for the purpose of accuracy measurement after obtaining the labelled data using unsupervised ML algorithm i.e., AE. These two different SL techniques are KNN and SVM. These two supervised learning techniques are being compared on the basis of accuracy in the two different approaches that are being considered for detecting erroneous data. Additionally, precision, recall, and f1- score are also provided for the second approach. These performance metrics are calculated as:

a. Accuracy:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

b. Precision:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

c. Recall:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

d. f1- score

$$\text{f1} = 2 * \left(\frac{\text{Precision} + \text{Recall}}{\text{Precision} * \text{Recall}}\right)$$

Where, True positives are considered as those values where actual value and predicted value is 1. In our model, True positives are the outliers. True negatives are those values where actual value and predicted value is 0. In our model, True negatives are the inliers. The other two parameters are False Positive and False Negative. The False Positive are the values where actual value is 1 but the predicted value is 0 while False Negative are those values where actual value is 0 but the predicted value is 1.

4.3 Approaches Used

The raw data obtained from the server is further analyzed. The analysis is done to determine the erroneous data and correct data from the raw data. This work concentrates on two different types of analysis. These analysis techniques are discussed further in this section.

4.3.1. Approach-1

In this approach, two different methods are considered. The first method feels raw data collected from the server and labeling using an unsupervised classifier "AE" where all features are taken into account for labeling. Afterward, in this method, this labeling is used as ground truth for computing accuracy in supervised learning models. The second method considers labeling using "AE" but dropping some features for increasing efficiency of the supervised learning models. These two methods are explained further in the following section.

4.3.1.1. Method-A

In this method, at first, the raw data is collected from the server. This raw data is an unlabeled dataset having some features. The AE is then used to provide labeling to the dataset. This work considers the simple AE having an input node, one hidden layer, and an output node. The input node of the AE is the same as the shape of our dataset. The output node is a replica of the input node, but it differs from the input node; otherwise, it would fail the purpose of finding patterns in the data. The hidden layer is used to enforce the model to prioritize the most important properties for reconstructing the output. In our work, the activation function used in the hidden segment is a ‘relu' activation function. The activation functions are required to introduce non-linearities in our data. The output from the AE, i.e., the detailed data is shown in the Figure 4.2.

Respiration Rate	SYS	DIA	PULSE	Class
18.0	142.0	76.0	86.0	1.0
15.0	125.0	76.0	83.0	0.0
17.0	125.0	78.0	85.0	0.0
17.0	124.0	80.0	85.0	0.0
14.0	123.0	89.0	81.0	0.0
19.0	120.0	92.0	87.0	1.0
15.0	123.0	75.0	83.0	0.0
15.0	132.0	99.0	83.0	1.0
15.0	127.0	86.0	83.0	1.0
14.0	127.0	94.0	81.0	1.0
16.0	139.0	99.0	85.0	1.0
15.0	128.0	94.0	82.0	1.0
15.0	124.0	86.0	85.0	0.0
14.0	122.0	81.0	80.0	0.0
14.0	116.0	89.0	81.0	0.0
12.0	116.0	86.0	81.0	0.0
14.0	119.0	90.0	80.0	0.0
16.0	117.0	84.0	82.0	0.0
14.0	117.0	84.0	82.0	0.0
13.0	117.0	84.0	82.0	0.0

Figure 4.2: Labelled Data from Autoencoder

The visualization of the labeling obtained from the autoencoder is performed using t-SNE display [32]. (t-SNE) is defined as the non-linear dimensionality reduction algorithm used for exploring high dimensional data, and it also maps multi-dimensional data to two or more dimensions. The main advantage of using t-SNE is that it focuses on preserving the distances between widely separated data points rather than on safeguarding distances between the nearby locations. The t-SNE visualization results are explained by Figure 4.3. This figure clearly shows inliers and outliers in two different colors. The inliers are represented by green color while outliers are represented by red color.

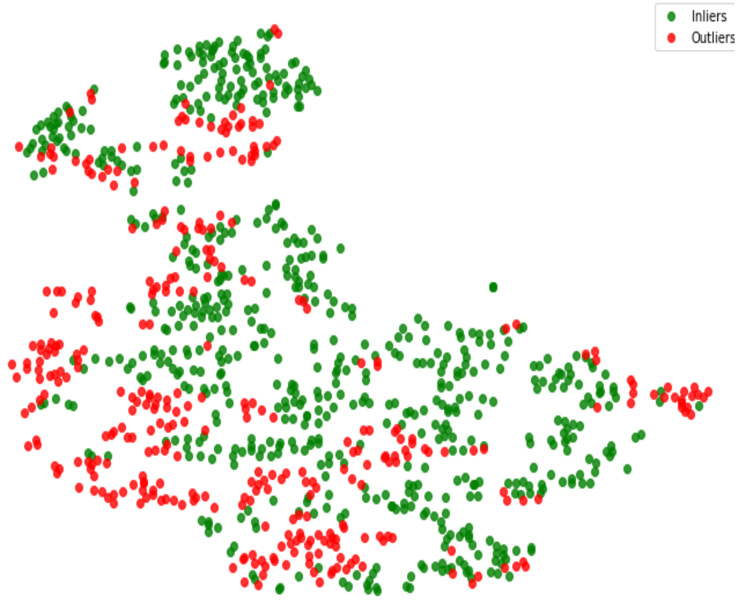

Figure 4.3: Autoencoder Visual Output

The ground truth obtained from the AE is used by the SL models for computing accuracy. Once the labelling is provided by AE then the data is further divided into two parts (training set and test set). The training set consists of 70% of the samples while rest 30% samples were kept for test set. The training set is used to train the models for the classification provided which was further validated by applying it on test set. The 70/30 rule is applied for dividing the samples into test set and training set [33]. Afterward, these SL models are compared based on accuracy achieved.

a. KNN

The KNN is one of the supervised learning models which uses ground truth obtained from the autoencoder and then computes the accuracy. The figure 4.4 depicts the confusion matrix of KNN.

N=300	Predicted: No	Predicted: Yes	
Actual: No	190	2	192
Actual: Yes	84	24	68
	274	26	

Figure 4.4: KNN Confusion Matrix

In this case, let us calculate the accuracy of the KNN classifier. The accuracy is being determined by using the formulae:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

$$= \frac{24+190}{84+2+190+24}$$

$$= 87\%$$

As shown in the Figure 4.4, we can clearly identify that TP in this case is 55 while TN are 190. Also, FP and FN are given as 54 and 1 respectively. It is to be noted that this accuracy is performed on test set having around 300 samples. This is done because the best possible results should be obtained on test sample.

b. SVM

SVM is one of the SL models in our work. It also uses the ground truth provided by AE. This ground truth is further implemented in SVM, and accuracy is being computed. The accuracy is being calculated using the confusion matrix, as shown in the Figure 4.5. The accuracy achieved by SVM is obtained as 81%.

N=300	Predicted: No	Predicted: Yes	
Actual: No	190	1	191
Actual: Yes	54	55	109
	244	56	

Figure 4.5: SVM Confusion Matrix

4.3.1.2. Method-B

In this method, at first unlabeled raw data is collected from the server. Afterward, AE are applied to provide labeling. This labeling serves as ground truth, and then some features are dropped to improve the accuracy of the supervised learning models. The features are dropped in which weak correlation existed i.e., Respiration Rate and Heart Rate. The dropping of features is done one by one with the help of method discussed further in this section.

The algorithm used for dropping features i.e., feature selection algorithm [33] is given in the Figure 4.6.

Tune/train the model on the training set using all predictors

Calculate model performance

Calculate variable importance or rankings

for *Each subset size* S_i, $i = 1 \ldots S$ **do**

- Keep the S_i most important variables
- [Optional] Pre–process the data
- Tune/train the model on the training set using S_i predictors
- Calculate model performance
- [Optional] Recalculate the rankings for each predictor

end

Calculate the performance profile over the S_i

Determine the appropriate number of predictors

Use the model corresponding to the optimal S_i

Figure 4.6: Feature Selection Algorithm

Once the feature selection algorithm is implemented, then accuracy of the SL models is discussed with the help of confusion matrices as shown in Figures 4.7 and 4.8.

a. KNN

N=300	Predicted: No	Predicted: Yes	
Actual: No	200	6	206
Actual: Yes	38	56	94
	238	62	

Figure 4.7: KNN Confusion Matrix

In the case of KNN, the accuracy achieved is 94%. It can be observed from the previous method that efficiency in this method increases as we dropped some features.

b. SVM

N=300	Predicted: No	Predicted: Yes	
Actual: No	190	7	197
Actual: Yes	39	64	109
	229	71	

Figure 4.8: SVM Confusion Matrix

In the case of SVM, the accuracy achieved is 83%. It can be observed from the previous method that efficiency in this method increases as we dropped some features.

4.3.1.3. Accuracy Comparison of Methods Used

The aforementioned methods are further compared on the basis of the accuracy achieved by them. The accuracy comparison of the methods is discussed further in the Table 4.1.

Table 4.1: Approach 1 Comparison Table

Algorithm Used	Method-1	Method-2
Support Vector Classifier	81%	83%
KNN	87%	94%

4.3.2. Approach-2

In this approach, at first, the raw data is collected from the server. The new information being collected contains features such as systolic blood pressure, diastolic blood pressure, respiration rate, and heart rate. This raw data contains erroneous data and correct data. The information is being distinguished by first finding the correlation between all the features and then providing labeling according to the relationship. At first, the labeling is provided based on strongly correlated features. Secondly, the labeling is provided based on weakly correlated features. In this work, strongly correlated features are systolic and diastolic blood pressure while weakly correlated features are heart rate and respiration rate. The figure 4.9 shows the correlation between all the features. It can be observed that negative correlation exists between respiration rate, systolic and diastolic pressure whereas positive correlation exists between respiration rate and heart rate. Also, positive correlation exists between systolic and diastolic blood pressure.

Correlation

	RR	SYS	DIA	HR
RR	1	-0.11	-0.06	0.16
SYS	-0.11	1	0.62	-0.048
DIA	-0.06	0.62	1	-0.048
HR	0.16	-0.048	-0.16	1

Figure 4.9: Correlation among features

a. KNN

In this method, the KNN classifier is applied on the dataset being labelled on the basis of strongly correlated features. The accuracy achieved by KNN is 92% and is being displayed using confusion matrix in the figure 4.10. Also, the classification report signifies other parameters such as precision, recall, and f1-score which is given by Figure 4.11. These parameters are calculated after evaluating the model for 7 times and signifies average value obtained after evaluation. In order to provide easier interpretation, the classification report as given by Figure 4.11 integrates numerical scores with color coded heatmap and all heatmaps are in the range of (0.0,1.0) for easier interpretation. These parameters differ for outliers as well as inliers. In this case, precision for inlier is more than outlier.

N=300	Predicted: No	Predicted: Yes	
Actual: No	190	7	197
Actual: Yes	12	91	103
	202	98	

Figure 4.10: Approach-2 KNN Confusion Matrix

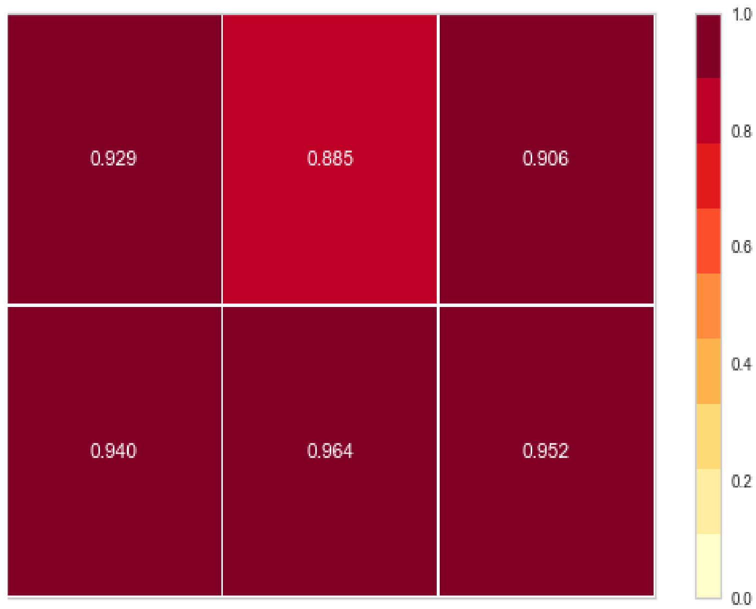

Figure 4.11: Approach-2 KNN classification Report

b. SVM

In this method, the SVM is applied on the dataset being labelled on the basis of strongly correlated features. The accuracy achieved by SVM is 90% and is being displayed using confusion matrix in the figure 4.12. Also, the classification report signifies other parameters such as precision, recall, and f1-score which is given by figure 4.13. These parameters are calculated after evaluating the model for 7 times and signifies average value obtained after evaluation. In order to provide easier interpretation, the classification report as given by figure 4.13 integrates numerical scores with color coded heatmap and all heatmaps are in the range of (0.0,1.0) for easier interpretation. These parameters differ for outliers as well as inliers. In this case, precision for inlier is more than outlier.

N=300	Predicted: No	Predicted: Yes	
Actual: No	190	4	194
Actual: Yes	25	81	106
	215	85	

Figure 4.12: Approach-2 SVM Confusion Matrix

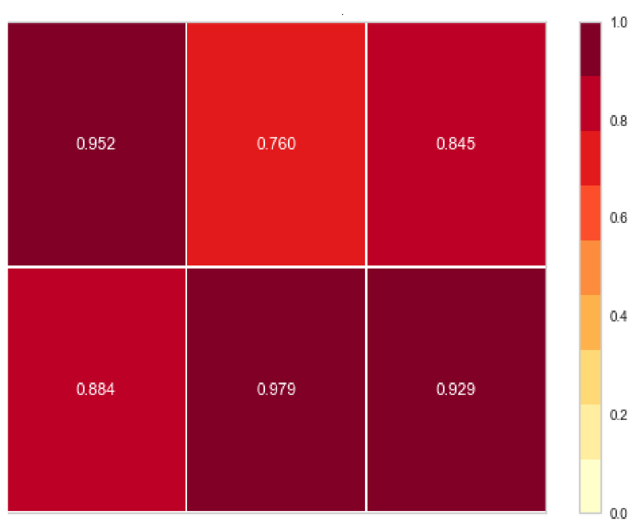

Figure 4.13: Approach 2 SVM Classification Report

4.3.2.1. Accuracy Comparison of Methods Used

The aforementioned methods are further compared on the basis of the accuracy achieved by them. The accuracy comparison of the methods is discussed further in the Table 4.2.

Table 4.2: Approach 2 Accuracy Comparison Table

Algorithm Used	Method-1	Method-2
Support Vector Classifier	85%	90%
KNN	88%	92%

The aforementioned approaches are further compared on the basis of the accuracy achieved by them. The accuracy comparison of the approaches is discussed further in the Table 4.3.

Table 4.3: Accuracy Comparison Table of two approaches

Algorithm Used	Approach-1	Approach-2
Support Vector Classifier	83%	90%
KNN	94%	92%

Chapter 5- Conclusions

5.1. Summary

IoT- based remote healthcare monitoring systems play an integral part in providing medical facilities to patients. In some cases, these systems play a pivotal role in preventing life-threatening issues. This thesis focuses on building the hardware prototype and detecting the erroneous data coming through the IoT devices on the server.

In the first phase of this thesis, it considered building a prototype system using IoT devices. In this phase, the two methods are used for collecting body vitals. The observations performed for this thesis considered data from well-being person to identify erroneous data coming due to either communication error or device error. Once the readings are performed from these devices then using BLE (Bluetooth Low Energy) the data is transferred to the central system. In our design, the standard method used was the mobile phone. Once the data was observed in the original order afterward, it was sent to the cloud. The raw data was extracted from the cloud using OAuth2.0 framework. This marks the end of the first phase.

In the second phase, initially, the raw data from the server was extracted for detailed analysis. In this phase, two approaches were used for detecting erroneous data.

In the first approach, unsupervised classifier "Autoencoder" is used for providing labeling to the raw data. The labeling from Autoencoder is further used as ground truth for comparing the accuracy of supervised learning algorithms such as Support Vector Machines, Logistic Regression, and K-Nearest Neighbor. KNN achieved the best accuracy. To check if the efficiency can be increased, some features were dropped iteratively. This process is often called feature selection. Once some features were lost when it was observed that the accuracy of the methods as mentioned earlier got increased. Therefore, K-NN was found to be performing better according to this approach.

In the second approach, the raw data was labeled by finding the correlation among the different features. Therefore, two methods are also used in this approach. In the first method, the labeling was performed using only strongly correlated features while in the second method, the labeling was performed using weakly correlated features. These two different methods were compared, and accuracy was being calculated. It was observed that the technique considering strongly correlated features was performing better. In the end, the comparison between the best methods from two different approaches was performed. It was observed from the comparison, that approach two was performing better for all of the supervised learning methods taken into consideration.

Therefore, we strongly recommend approach-2 if the sample size is small, whereas approach one if the sample size is large. It should be taken into account that strategy 1 is a time-consuming process, while method 2 is a time-saving process.

5.2. Future Work

From the growing interest of IoT-based remote health monitoring, it is evident that there is a massive scope of future research and study. As a path of future work, we can point out three different perspectives of future work from this thesis.

Hardware Implementation:

The first perspective is building a hardware implementation of the prototype being developed in this work. This work concentrates only blood pressure (SYS and DIA), Respiration Rate, and Heart Rate but other factors need to be taken into accounts such as Blood Sugar Level, Oxygen Saturation (SPO_2), Body Temperature, etc. This work takes into account the use of two devices but it possible to integrate different sensors in a single device which can measure other parameters also.

Security:

Another, perspective is concerned with providing security to the data. In this thesis, the data taken into account is health-related data. Therefore, it is required to make ensure that this data flows through a highly secured pathway. The various security related methods can be implemented, and there is also a possibility of providing protection using standard machine learning algorithms.

Data Processing:

Finally, considering data processing phase in this thesis. This work only concentrates a few machine learning methods such as Autoencoders, KNN, and SVM. This work focuses on basically four parameters for analysis but using online source dataset can be obtained with more features. Therefore, if features in the dataset are more than more advanced machine learning and deep learning techniques can be implemented to deal with a massive amount of dataset.

BIBLIOGRAPHY

[1] F.Nicolas, "Life Science Reader," *AI in Life Sciences- Seeing Past the Hype,* 1st March 2019.

[2] "wearable medical devices," [Online]. Available: https://www.grandviewresearch.com/industry-analysis/earable-medical-devices-market. [Accessed 22 May 2019].

[3] R. Kennely, "The IEEE 1073 Standard for Medical Device Communication," in *IEEE AUTOTESCON Proceedings*, Meredith, NH, 1998.

[4] S. R. Islam, D. Kwak, M. H. Kabir, M. Hossain and K.-S. Kwak, "The Internet of Things for Health Care: A Comprehensive Survey," *IEEE Access,* vol. 3, pp. 678-708, 2015.

[5] K. Rose, L. Chaplin and E. S., "THE INTERNET OF THINGS: AN OVERVIEW Understanding the Issues and Challenges of a More Connected World," [Online]. Available: https://www.internetsociety.org/wp-content/uploads/2017/08/ISOC-IoT-Overview-20151221-en.pdf. [Accessed 25 May 2019].

[6] A. l. Samuel, ""Some Studies in Machine Learning Using the Game of Checkers," IBM J. Res. Dev., vol. 3, pp. 210--229, 1959," *IBM J. Res. Dev,* vol. 3, pp. 210-229, 1959.

[7] A. Charleonnan, T. Fufaung, T. Niyomwang, W. Chokchueypattanakit, S. Suwannawach and N. Ninchawee, "Predictive analytics for chronic kidney disease using machine learning techniques," in *Management and Innovation Technology International Conference (MITicon).*, 2016.

[8] F. Ahamed and F. Farid, "Applying Internet of Things and Machine-Learning for Personalized Healthcare: Issues and Challenges," in *International Conference on Machine Learning and Data Engineering (iCMLDE).*, 2018.

[9] A. Ukil, S. Bandyoapdhyay, C. Puri and A. Pal, "IoT Healthcare Analytics: The Importance of Anomaly Detection," in *IEEE 30th International Conference on Advanced Information Networking and Applications (AINA)*, 2016.

[10] R. R. Reddy, C. Mamatha and R. G. Reddy, "A Review on Machine Learning Trends, Application and Challenges in Internet of Things," in *International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2018.

[11] Y. Fan, P. Xu, H. Jin, J. Ma and L. Qin, "Vital Sign Measurement in Telemedicine Rehabilitation Based on Intelligent Wearable Medical Devices," *IEEE Access,* vol. 7, pp. 54819-54823, 2019.

[12] Z. U. Ahmed, M. G. Mortuza, M. J. Uddin, M. H. Kabir and M. Hoque, "nternet of Things Based Patient Health Monitoring System Using Wearable Biomedical Device.," in *International Conference on Innovation in Engineering and Technology (ICIET)*, 2018.

[13] M. M. Neto, E. F. Coutinho, L. O. Moreira, J. N. d. Souza and N. Agoulmine, "A Proposal for Monitoring People of Health Risk Group Using IoT Technologies, 2018," in *20th International Conference on e-Health Networking, Applications and Services (Healthcom)*, 2018.

[14] K. S. Sanjay and D. Bhagwat, "A Secured IoT Based Webcare Healthcare Controlling System using BSN," in *Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*, 2018.

[15] Akshat, Gaurav, Zahid, Bhupendra, Aditi, S. Kumar, Maneesha and P. Pandey, "A Smart Healthcare Monitoring System Using Smartphone Interface," in *4th International Conference on Devices, Circuits and Systems (ICDCS)*, 2018.

[16] R. G. Utekar and J. S. Umale, "Automated IoT Based Healthcare System for Monitoring of Remotely Located Patients," in *Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, 2018.

[17] A. Raji, P. G. Jeyasheeli and T. Jenitha, "IoT based classification of vital signs data for chronic disease monitoring," in *10th International Conference on Intelligent Systems and Control (ISCO)*, 2016.

[18] Z. Chen, C. K. Yeo, B. S. Lee and C. T. Lau, "Autoencoder Based Network Anomaly Detection," in *Wireless Telecommunication Symposium*, 2018.

[19] K. I. Zhang and Z. f. He-Yan, "KNN Text Categorization Based on Semanic Center," in *International Conference on Information Technology and Computer Science*, 2009.

[20] R. Zamel, R. Urtasun and S. Fidler, "Nearest Neighbors," [Online]. Available: https://www.cs.toronto.edu/~urtasun/courses/CSC411- Fall 16/05_nn.pdf. [Accessed 24 May 2019].

[21] Z. Liu, X. Lv, K. Liu and S. Shi, "Study on SVM Compared with other Text Classification Methods," in *Second International Workshop on Education Technology and Computer Science*, 2010.

[22] "Support Vector Machine, Simply Explained, Medium," Towards Data Science, [Online]. Available: towardsdatascience.com/support-vector-machine-simplyexplained-fee28eba5496. [Accessed 23 June 2019].

[23] "The Kernel Trick," [Online]. Available: www.eric-kim.net/eric-kim-net/posts/1/kernel_trick.html. [Accessed 25 June 2019].

[24] " Support Vector Machines - an Overview | ScienceDirect Topics," ScienceDirect, [Online]. Available: www.sciencedirect.com/topics/neuroscience/support-vector-machines. [Accessed 13 August 2019].

[25] T. Kardi, "Strength and Weakness of K-Nearest Neighbor Algorithm." K Nearest Neighbors Tutorial: Strength and Weakness," [Online]. Available: people.revoledu.com/kardi/tutorial/KNN/Strength and Weakness.htm..

[26] E. D. Hardit, "The OAuth 2.0 authorization framework," 2012.

[27] "iHealthSense Specifications," iHealthLab, [Online]. Available: https://helpcenter.ihealthlabs.eu/hc/en-gb/articles/115003895485-Technical-specifications-iHealth-Sense-BP7. [Accessed 26 May 2019].

[28] Spire Health, [Online]. Available: https://www.ctnet.com/products/spire-mindefulness-and-activity-tracker-stone. [Accessed 26 May 2019].

[29] T. Pickering, J. Hall, L. Appel, B. Falkner, J. Graves, M. Hill, D. Jones, T. Kurtz, S. Sheps and E. Roccella, "Recommendations for Blood Pressure Measurement in Human and experimental Animals Part 1: Blood Pressure Measurement in Humans a Statement for Professionals From The Subcommittee of Professional and Public Education of American Heart Association Council, H," vol. 45, 2005.

[30] R. Jaffar, Z. Mahoodin, R. M. Abdullah, H. M. Desa and Z. Zaharudin, "Noninvasive Blood Pressure(NIBP) Measurement by Oscillometric Principle," in *2nd International Conference on Instrumentation, Communications, Information Technology, and Biomedical Engineering*, Bandung, Indonesia, 2011.

[31] "Spire Stone," Spire Stone, [Online]. Available: https://spirehealth.com/pages/stone. [Accessed 28 May 2019].

[32] Saurabh, " Comprehensive Guide on t-SNE algorithm with implementation in R & Python," Analytics Vidhya, 22 March 2017. [Online]. Available: https://www.analyticsvidhya.com/blog/2017/01/t-sne-implementation-r-python. [Accessed 27 May 2019].

[33] Data and Beyond, 24 August 2017. [Online]. Available: https://dataandbeyond.wordpress.com/2017/08/24/split-of-train-and-test-data/. [Accessed 4 September 2019].

[34] M. Kuhn, ""20 Recursive Feature Elimination," [Online]. Available: https://topepo.github.io/caret/recursive-feature-elimination.html. [Accessed 20 June 2019].

VITA AUCTORIS

NAME:	Varun Kumar Vats
PLACE OF BIRTH:	Delhi, India
YEAR OF BIRTH:	1994
EDUCATION:	APJAKTU University, B.Tech., Electronics and Communication Engineering, India, 2016
	University of Windsor, M. App.Sc, Electrical Engineering, Windsor, ON, 2019