Title	A Survey on Internet of Things for Smart Health Technologies [課題研究報告書]
Author(s)	Le, Duy Tan
Citation	
Issue Date	2018-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/15251
Rights	
Description	Supervisor:丹　康雄，先端科学技術研究科，修士（情報科学）

A Survey on Internet of Things for Smart Health Technologies

LE Duy Tan
Graduate School of Advanced Science and Technology
Japan Advanced Institute of Science and Technology
March, 2018

A Survey on Internet of Things for Smart Health Technologies

s1610217 LE Duy Tan

Supervisor : Yasuo Tan
Main Examiner : Yasuo Tan
Examiners : Yuto Lim
Yoichi Shinoda
Razvan Beuran

Graduate School of Advanced Science and Technology
Japan Advanced Institute of Science and Technology
[Information Science]
February, 2018

Abstract

Nowadays, accompanying with prolonged life expectancy, there is an increase of chronic diseases which include obesity, diabetes, cardiovascular diseases, cancer, osteoporosis and dental diseases at the global level. As a result, the task of supporting, enhancing and improving upon the existing healthcare services poses a wide range of challenges. Accordingly, the utilization of sophisticated technologies in the healthcare field, which is called "smart health", is not an optional value, but a requirement. It is accepted that the Internet of Things (IoT) is the core technology of smart health. The adoption of advanced technology, particularly Internet of Things (IoT), in the field of medicine and healthcare, benefits patients with better medical assistance, reduce a treatment time, lower medical costs and more satisfying healthcare services.

Many efforts have been made and a large body of research regarding the potential and implementation of IoT in healthcare already existed. The main problem, however, is the need for a comprehensive literature review which states the extensive overview of the field and advances in IoT-based healthcare technologies in the aspect of system architecture. This study aims to state the comprehensive overview of the field and advances in IoT-based healthcare technologies in the aspect of system architecture. The research survey investigates and summarizes the existing knowledge, state-of-the-art technologies of each aspect in the field. Further, this research provides an exhaustive understanding of the successful case studies on the IoT in the healthcare context which is expected to be useful for further research. Following the structure of standard IoT-based healthcare system architecture, the summary of our project paper is given as follows.

First of all, we identify physiological information, the environmental aspects, as well as indoor location information. They are the necessary data that needs to be collected in order to ensure the quality of life, safety, and well-being. Vital signs, which primary are temperature, pulse rate (heart rate), respiration rate (breathing rate) and blood pressure, are essentially physiological signs to indicate the status of the body. Besides, several extra medical signs for example muscle activity level, blood glucose level, gaseous carbon dioxide levels (CO_2), electrical activity in the brain are all essential to track down the healthcare status. Since an abnormal change of indoor environment may cause dangerous situations and result in undesired consequences of directly or indirectly harming occupants, environmental aspect is another relevant information. Moreover, based on the advantages of the information, the demand for indoor localization especially in healthcare area is increased recently. The specific normal range, effectiveness, and typical application of each type of data are sufficiently discussed.

Secondly, we discuss and investigate between recent advances in sensor technology, microelectronics, and telecommunication which applied to collect and monitor healthcare data. The environmental sensors can be used to capture environmental parameters. For instance, accelerometer and compass sensors are applied in location monitoring application while the biomedical sensors such as electrocardiogram (ECG) and photoplethysmography (PPG) sensors can be used to monitor the health figure of the human. The collected data can be analyzed and sent to the server through communication technologies. Wildly known technologies in smart health field are RFID, NFC, Wifi, 6LowPAN, Bluetooth, ZigBee. Advantages of in sensor technology and common sensor devices are defined. The flexibility, communication efficiency and cost-effective of using Body Area Networks Architecture in healthcare are also presented. Furthermore, we compare among IEEE 802.15.1 (Bluetooth), 802.15.4 (the basis for Zigbee) and 802.15 (standard for wireless personal area network - WPAN), these are the most commonly employed wireless communication standards in BANs.

In the healthcare field, especially in healthcare monitoring system, the reliability of input data is extremely important. Accurate healthcare decisions can only be obtained with accurate input data. However, it is possible to have noise and outliers in sensor data. It is caused by (1) low quality sensors or errors of the sensors and (2) occurrence of noise or motion artifacts in all sensor networks. Consequently, these might lead to faults in reading and giving alarm for patient and healthcare provider deliveries. In other to control data quality, the collected data need to go through preprocessing steps before being analyzed by algorithms. We employ several criteria including usage, referable, possibility, popularity, standardization and intelligence to evaluate the impact of algorithms. The most critical data preprocessing algorithms in the healthcare system can be classified into two categories: imperfect data and imbalanced data preprocessing methods. The chart is used to demonstrate the evolution of academic publications concerning the data preprocessing algorithms from 1990 to 2016 while constructing tables thoroughly reviews the detailed figure of them from 2010 to 2016. Publication statistics are acquired from Google Scholar; the search query is defined as the subfield name of algorithms and at least one term of medical or health appearing, for example, "'support vector machine' AND medical OR health".

IoT-based healthcare systems can be utilized to a diverse array of fields, including care for pediatric and elderly patients, the supervision of chronic diseases, and the management of private health and fitness, among others. Starting from the most successful support vector machine algorithm, we identify plausible architectures in the healthcare domain. The research trends in algorithms applied for healthcare decision support systems as well as the number of publication referring to several algorithms are analyzed. Among abundant IoT-based healthcare applications, we can classify them into two main categories: pervasive monitoring and medical informatics applications. In order to provide the overview, we summarize the feature IoT-based healthcare applications.

Sensor data issues, the accuracy of indoor location monitoring system and system design are emerging as the principal challenges and limitations in the IoT-based healthcare area. We briefly clarify these issues and through the final discussions, further areas that would benefit from the introduction of IoT technologies will be identified.

As a second study, based on the general system architecture we conduct the implementation of the smart healthcare system. The system applied modern technologies to solve the current problem in smart healthcare. Besides, the implemented system is expected to be useful for further research.

Regarding future works, we propose an Integrated Home-Based Healthcare Monitoring System ensuring wellness for each individual lives in the house. In this system, the data collected by health sensors, environment sensors, and location sensors will be pre-processed and sent to the Smart Health Platform through the IoT gateway. The Smart Health Platform not only stores the data but also controls the connectivity of devices and provides the APIs for the third-party application. The Smart Health Applications can provide several safety applications such as fall detection and emergency assistance.

Through the survey, researchers can get enough understanding, save their time in searching articles and reduce repetitious work for supporting the task of designing and developing the smart healthcare system. Besides, stakeholders, healthcare manufacturers, family members and especially patients will get benefit from our research.

Keywords. Internet of things, healthcare, technologies, smart health, applications, networks, sensors

Acknowledgements

I would like to express my very great appreciation to my supervisor, Professor Yasuo Tan. From commencement in finding an appropriate subject to the finalization of my research, Professor Yasuo Tan offers his unreserved help and guidance and lead me to finish my thesis step by step. His lectures are always inspiring me and bring me to a higher level of thinking. What I learn from him is not just only about the knowledge but also how to view and approach problems in a new perspective. Without his kind and patient instruction, it is impossible for me to finish this thesis.

I would like to thank Associate Professor Yuto Lim, my second supervisor, for his guidance and support throughout my study. Thanks to his pieces of advice, not only my presentation slides but also my thesis become better and better.

I am particularly grateful for the assistance given by Associate Professor Razvan Beuran and Professor Yoichi Shinoda. As committee members of my defenses, they provided instructive questions and helpful comments to clarify and improve my research.

I wish to acknowledge the help provided by Professor Nak Young Chong. Without his provided equipment, it is difficult for me to finish the implementation part of my thesis.

I would like to convey my special thank to Associate Professor Nguyen Le Minh and Dr. Tuan Nguyen for giving me a chance to study in Japan.

Completing this work would have been more difficult without the support and friendship provided by the other lecturers, friends at JAIST, V-Project 2016 members, especially Quyen Dang, Quynh Le, Duy Huynh, Viet Lai, Phi Hung, Pong Pi, Kris, and Quynh Dao, etc. I am indebted to them for their help.

Finally, I would like to express the most sincere gratitude to my family and parents for their support and provide a carefree environment for me, so that I can concentrate on my study.

Contents

List of Figures

List of Tables

Abbreviation

APIs	Application Package Interfaces
BANs	Body Area Networks
BLE	Bluetooth Low Energy
BP	Blood Pressure
bPCA	Bayesian Principal Component Analysis
bpm	Beats Per Minute
BSN	Body Sensor Network
BVP	Blood Volume Pulse
DBN	Deep Belief Network
DM	Decision-Making
DNN	Deep Neural Network
ECG	Electrocardiogram
EDA	Electrodermal Activity
EEG	Electroencephalogram
EF	Ensemble Filter
EIRP	Effective Isotropic Rediated Power
EM	Expectation-Maximization
EMG	Electromyography
ETSI	European Telecommunications Standards Institute
FFT	Fast Fourier Transforms
FKM	Fuzzy K-Means
FL	Fuzzy Logic
GPUs	Graphics Processing Units
GSR	Galvanic Skin Response
GSR	Galvanic Skin Response

HITL	Human In The Loop
HR	Heart Rate
IBI	Inter Beat Interval
IDE	Integrated Development Environment
IoT	Internet Of Things
IPF	Iterative-Partitioning Filter
IPSO	Internet Protocol For Smart Objects Alliance
ITIF	Information Technology and Innovation Foundation
JSON	Javascript Object Notation
KNN	K-Nearest Neighbors
kNNI	K-Nearest Neighbor Imputation
LSTM	Long Short-Term Memory
MEMS	Micro-Electro Mechanical Systems
MI	Multiple Imputation
MICE	Multiple Imputations By Chained Equations
MICS	Medical Implant Communication Service
nW	Nanowatt
PCBs	Printed Circuit Boards
PDA	Personal Digital Assistant
PEM	Personal ECG Monitoring
PPG	Photoplethysmography
PSD	Power Spectral Density
RFID	Radio Frequency Identification
RNNs	Recurrent Neural Networks
RSSI	Received Signal Strength Indication
SDKs	Software Development Kits

SIG	Bluetooth Special Interest Group
SMOTE	Synthetic Minority Over-Sampling Technique
SOA	Service-Oriented Approach
SVD	Singular Value Decomposition
SVM	Support Vector Machines
TRL	Technological Readiness Levels
URUS	Ubiquitous Networking Robotics in Urban Settings
UWB	Ultra-Wideband
WHO	World Health Organization
WIP	Work In Progress
WLAN	Wireless Local Area Networks
WPAN	Wireless Personal Area Network
WSNs	Wireless Sensor Networks

Chapter 1

Introduction

1.1 Research Motivation

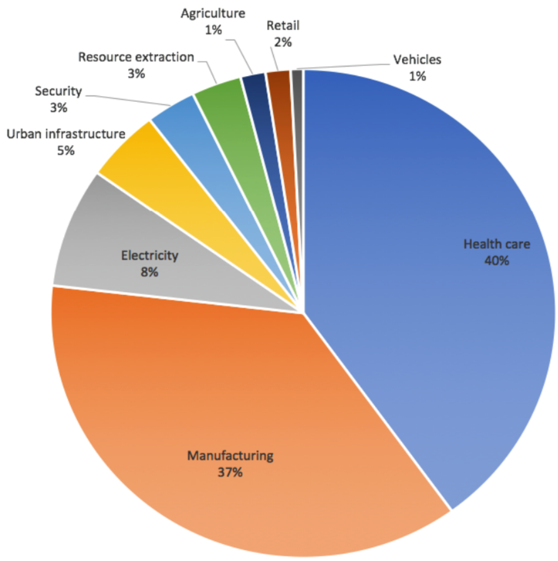


Figure 1.1: Applications of the Internet of Things by the year 2025

Nowadays, quality of life improves in advanced societies, life expectancy and the proportion of elderly citizens is projected to increase globally. In particularly, the proportion of Japanese population is above age 65 in 2015 is 26%. This proportion is estimated to reach one-third of the total population in 2050 according to United Nations Population Information Network [1]. Along with this prolonged life expectancy, there is an increasing of chronic diseases which include obesity, diabetes, cardiovascular diseases, cancer, osteoporosis and dental diseases at the global level; as reported by World Health Organization (WHO) [2]. As a result, the task of supporting, enhancing and improving upon the existing health care services poses a wide range of challenges. Therefore, the utilization of sophisticated technologies in healthcare field, which is called "smart health", is not an optional choice, but a requirement [3]. The concept of smart health is defined by the technology which must satisfy two main conditions. The first one is the achievement of better diagnostic tools and better treatment for patients. The second condition is identified through the improvement of life quality. The Internet of Things (IoT) is generally accepted not only as a connected set of anyone, anything, anytime, anyplace, any service and any network but also as the core technology of smart health. The use of modern technology, particularly Internet of Things (IoT) in the field of medicine and healthcare will benefit patients with better medical assistance, shorten treatment time, lower medical costs and more satisfying health care services [4]. Moreover, health care applications and related IoT-based services are expected to be about 40% of the whole annual economic impact caused by the IoT. While the IoT is estimated to be in range of $2.7 trillion to $6.2 trillion by the year 2025 [5], the IoT-based services can possibly contribute $1.08 trillion to 2.48 trillion. Fig 1.1 presents the projected market share of dominant IoT applications by the year 2025.

1.2 Research Background

Many efforts have been made as well as a large body of research regarding the potential of IoT and its applications in healthcare have already existed. Besides, there are several published survey papers that explore different aspects of the IoT technology in the field.

- López in [6] conducted a literature review of the IoT for specific clinical wireless devices using 6LoWPAN/IEEE 802.15.4, Bluetooth and NFC for mHealth and eHealth applications.
- The research survey at [7] focused on the development of pervasive healthcare from its origination for activity recognition using wearable sensors to the future of sensing implant deployment and data processing. The possibilities of the combination between pervasive health monitoring through data linkages and other health informatics systems including the mining of health records, clinical trial databases, multiomics data integration, and social media was also discussed in this survey.
- Regarding the smart health monitoring systems, over fifty different systems were selected, categorized, classified and compared in [8] to make an extensive review of smart health monitoring systems and an overview of their design and modeling. In addition, a critical analysis of the efficiency, clinical acceptability, strategies and recommendations on improving current health monitoring systems were presented.
- For the algorithms, the author of [9] reviewed models based on deep learning approach including Recurrent Neural Network, Deep Boltzmann Machine, Deep Belief Network, Convolutional Neural Network, Deep Autoencoder and Deep Neural Network in the broad context of health informatics applications. Ranging from genomic analysis to biomedical image analysis, the research was focused exclusively on deep learning techniques tailored to Electronic Health Record data.

However, the IoT still remains its infancy in the health care field.

1.3 Research Objective

There is a need for a comprehensive literature review which states the extensive overview of the field, advances in IoT-based health care technologies in the aspect of system architecture. Objectives of this research are clearly given by:

- **Investigating and summarizing the existing knowledge and state-of-the-art technologies in the field**. There are a lot of fundamental definitions and terms in healthcare research terms. Besides, multiple researchers have been studied for decays. They proposed a lot of problems and their own solutions with limitations. In addition, advanced techniques have developed through time and challenged new researchers. There may be a case that a problem was solved by multiple techniques; however, none of proposed manners gave a perfect solution. Researchers can waste years to study and understand identified problems and proposed solutions while they have to spend time on finding a new solution with updated technologies.

- **Provide an exhaustive understanding of the successful case studies on the IoT in the healthcare context which is expected to be useful for further research**. There are several of of the smart healthcare systems and applications which have been proposed with a lot of advantages of humans. However, each system and application have their own strength and weakness. It is necessarily to classify them into separated categories based on their purposes, applied manners and used techniques. It will help researchers to have deeper interpretation about systems in the same category to find a general solution dealing subproblems of each systems.

- **Identify the current problems of smart healthcare systems based on the view of general system architecture**. Besides, we will implement a smart health care system which applies methodologies to solve the current issues of the field. The implemented system is supposed to be useful for the smart healthcare research in the future.

Therefore, through the survey, researchers can get enough understanding, save their time in searching articles and reducing repetitious work for supporting the task of designing and developing the smart healthcare system.

In order to accomplish our objectives, we have conducted this literature review. In this research, we focus on the following aspects of the healthcare system.

- First of all, we identify all necessary data which is essential collected for providing health care services.

- Secondly, it is important to consider network connectivity technologies supported for collecting and monitoring the data. In oder to control the quality of data, the collected data needs to go through the data preprocessing step before being analyzed by algorithms. Data preprocessing techniques and smart health algorithms will be discovered. Moreover, we identify some specific applications in smart health area.

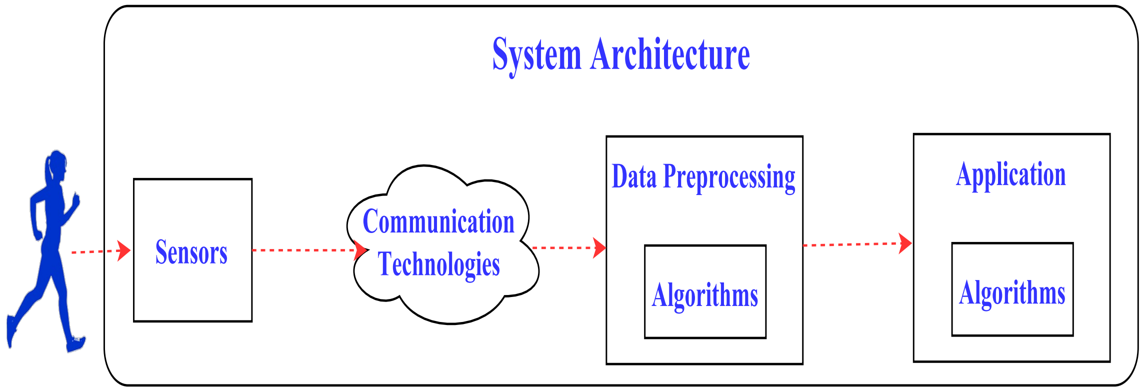


Figure 1.2: The process of literature review based on the general smart healthcare system architecture.

- Finally, we will study the system architecture design of smart health technologies. Through these discussions, further areas that would benefit from the introduction of IoT technologies will be identified. Moreover, based on this study, we will implement a smart healthcare system as a demo solution for the current problems of the field.

Fig 1.2 visualizes the process of literature review following the smart health care system architecture.

1.4 Research Outline

This report consists of eight chapters: Introduction, Data Collected by Sensors, Sensor and Communication Technology, Data Preprocessing, Health Care Application, Open Challenges and Issuses, Implementation, Discussion and Concluding Remarks. Besides, the Appendix which is a further detailed for the finding will be attached at the end of this report research. The remained sections of this research are organized as follows:

- The necessary data in smart health including physiological information, the environmental aspects, indoor location information are introduced in Section II.
- Sensors and their communication technologies in the smart health care field are represented in Section III.
- Data preprocessing and health care applications along with their algorithms and methods are reviewed in Section IV and V.
- Later, opportunities for research and development will be identified in Section VI.
- Chapter VII will be the implementation of smart healthcare system.
- Section VIII will be the discussion and concluding remarks of our research survey.

Chapter 2

Sensor Data

2.1 Introduction

Physiological information, the environmental aspects, indoor location information are certainly collected in order to ensure the quality of life, safety and well-being. Physiological information can reflex the health of humans while environmental aspects and indoor location information helps to identify the current position of an individual and warn the individual if the current state is dangerous because of unexpected objects or situations. Regarding details of mentioned information and aspects, following sections are presented.

2.2 Physiological information

2.2.1 Vital signs

Vital signs which primary are temperature, pulse rate (heart rate), respiration rate (breathing rate) and blood pressure, are the most important physiological signs that indicate the status of the body. They are often the fundamental information requested for during emergencies and required in any healthcare system. Vital signs can be used to measure the body's most basic functions for effectively identifying or monitoring health problems. Either rapid changes or gradual differences in vital signs have been proven to be a significant predictor of life-threatening health events. The normal ranges for a person's vital signs vary with patient demographics, age, gender, weight, overall physical conditions, and with environment variations, diet, uid status, stresses, and time of day [10]. Table 2.1 depicts the 4 basic vital signs.

Although individual body temperature fuctuation depends upon age, gender, exertion, time of day, food and fluid consumption, and the stage of the menstrual cycle of women, it has a range of 97.8 degrees F, or Fahrenheit to 99 degrees F (equivalent to 36.5 degrees C, or Celsius to 37.2 degrees C) for a healthy adult. The abnormal body temperature can be a sign of various health problems such as infection, heatstroke, faltering neurological heat regulation centers or sweat production mechanisms, malignant hyperthermia, side effects of certain medications or illicit drugs, vasodilatory malfunction, and some toxins.

The abnormal temperature can be classified into two types: fever (abnormally high body temperature) or hypothermia (abnormally low temperature).

- Fever is the stage that body temperature rises above the normal temperature of 98.6°F(37°C) with one degree or more. The body temperatures higher than 104°F (40°C) becomes a medical emergency requiring urgent treatment to avoid potentially disability or death.

- In contrast, hypothermia is a seriously low body temperature indicated when the body loses heat faster than it can produce. The body temperature drops below 95°F (35°C) when hypothermia occurs.

Pulse rate or heart rate is the speed of the heartbeat determined by the number of times a minute that human's heart contracts or beats. Although the typical fluctuation of pulse rate are provoked by physical exercise, smoking, caffeine, illness, injury, mood changes and ingestion of drugs, the accepted normal resting adult human pulse rate are from 60 to 100 beats per minute (bpm). An irregular heartbeat or an arrhythmia is a symptom of basic fitness level, cardiovascular health such as the risk of a heart attack, hypertension and atherosclerosis. Based on not only origination (atria or ventricles) but also the speed of heart rate, arrhythmias can be classified into two typical types: tachycardia (fast heartbeat) and bradycardia (low heartbeat). Both tachycardia and bradycardia are not abnormal and often self recovery unless they become symptomatic arrhythmias.

- Tachycardia: a resting heart rate greater than 100 bpm (beats a minute).

- Bradycardia: a resting heart rate less than 60 bpm (beats a minute).

Respiration rate (breathing rate) is the rate at which breathing occurs which is measured by the number of breaths a person takes per minute. The regular range of respiration rate for a healthy mature at rest is accepted as 12 to 18 breaths per minute. However, some research denote this range as high as 24 breaths per minute for matures. There is a wide range of variability in children commonly higher than matures from 17 to 40 breaths per minute [11]. Due to fever, illness, or other medical conditions, the normal respiration rates may change. There are also two types of abnormal respiration rate: tachypnea (rapid, abnormal breathing) and hypopnea (abnormally low respiratory rate). This abnormal respiration rate is an indicator of increasing nervousnessstar, exercise, infection, or stress and can be reversible by either calming measures or removing the infection.

Blood pressure is the force of circulating blood against the walls of blood vessels including both venous and arterial. When conducting blood pressure measurement, two numbers are captured: a systolic pressure and a diastolic pressure. A systolic pressure, the higher number, indicates the pressure strength (mm/Hg) inside the artery when the heart contracts and pumps blood through the body. A diastolic pressure, the lower number, indicates to the pressure inside the artery when the heart is at rest and is filling with blood. Blood pressure is affected by cardiac output, total peripheral resistance and blood vessel stiffness and varies depending on situation, stress, exercise, and related medical problems. 140 mm/Hg or higher systolic pressure or 90 mm/Hg and higher diastolic pressure is defined as high blood pressure. The normal range of blood pressure for healthy adult are 119 mm/Hg or less systolic and 79 mm/Hg or less diastolic.

Table 2.1: The basic vital signs.

Vital Sign	**Description**	**Normal Range**
Temperature	Depending on gender, recent activity, food and fluid consumption, time of day, and, in women, the stage of the menstrual cycle.	97.8°F to 99.1°F (36.5°C to 37.3°C) with an average of 98.6°F (37°C).
Pulse (Heart) Rate	A measurement of the heart rate, or the number of times the heart beats per minute. Tachycardia is the heart rate that higher than 100 bpm while the heart rate that lower than 60 bpm is called Bradycardia.	60 to 100 beats per minute (bpm).
Respiration Rate	The number of breaths a person takes per minute.	12 to 18 breaths per minute for an adult person, children's ranges as high as 24 breaths per minute.
Blood Pressure	The pressure of circulating blood on the walls of blood vessels. BP is composed of 3 pressures: systolic pressure, diastolic pressure.	119 mm/Hg or less systolic and 79 mm/Hg or less diastolic.

2.2.2 The Extra Medical Signs

Several other parameters such as muscle activity level, blood glucose level, gaseous carbon dioxide levels (CO_2), electrical activity in brain, etc. are all very important to track down the healthcare status [12], [13], [14], [15].

Blood glucose or blood sugar [16] is the amount of glucose circulating in the blood of humans and other animals. Although blood glucose level fluctuates throughout the day, the accepted normal range for non-diabetics is from 3.9 to 5.5 millimoles per litre (mmol/L) (or 70 to 100 milligrams per deciliter - mg/dL) while the mean normal blood glucose level in humans is about 5.5 mmol/L (100 mg/dL). There are two types of abnormality in blood sugar levels: hyperglycemia and hypoglycemia.

- Hyperglycemia or high blood sugar is a condition in which blood sugar increases to higher than normal levels due to an excessive amount of glucose circulates in the blood plasma. The short-term hyperglycemia can make the body suppress appetite while the long-term high blood sugar is the reason of many the medical problems including heart disease, eye, kidney, and nerve damage. Moreover, high blood sugar is an indicator of prediabetes and diabetes (both type 1 diabetes and type 2 diabetes). Therefore, blood glucose monitoring is an essential part of diabetes control. Patients with type 2 diabetes must test their blood sugar concentrations at least once a day. Those who need to take insulin that consists of all type 1 diabetes and some with type 2 ones, have to test their blood several times a day.

- Hypoglycemia or low blood sugar is a condition in which blood sugar drops to lower than normal levels. Hypoglycemia tends to occur on quickly and can vary from person to person with a feeling of hunger, sweating, shakiness, and weakness. Hypoglycemia is a potentially fatal condition which causes many health problems including uncoordinated, trouble talking, confusion, unconsciousness, convulsions, or even death. Glucose plays an important part of metabolism and nutrition, which is the suitable functioning of the body's organs. Consequently, hypoglycemia is more dangerous than hyperglycemia.

The muscle activity level at rest and during contraction is another important information which can be analyzed to identify medical abnormalities, neuromuscular diseases, assessing low-back pain or activation level. The muscle activity level can aslo be used as a control signal for prosthetic devices such as prosthetic hands, arms, and lower limbs.

Gaseous carbon dioxide levels [17] is the amount of carbon dioxide (CO_2) in the blood serum. There are two major forms of CO_2 in the human body with more than 90% in the form of bicarbonate (HCO_3) and nearly 10% in the form of carbon dioxide (PCO_2). The normal range for CO_2 in the human blood is around 23 to 29 mEq/L (milliequivalent units per liter of blood). The imbalance between the oxygen and carbon dioxide in human blood is the indications of a kidney, respiratory, or metabolic disorder. Shortness of breath, breathing difficulties, nausea or vomiting are the common signs of an imbalance of oxygen and carbon dioxide in blood. The abnormality blood carbon dioxide levels condition includes low bicarbonate and high bicarbonate combined with high blood pH and low blood pH.

- Metabolic acidosis is a condition of low bicarbonate and low blood pH (less than 7.35) because the body produces excessive quantities of acid or when the kidneys are not able to remove enough acid from the body. Kidney failure, severe diarrhea, lactic acidosis, seizures, cancer, prolonged lack of oxygen from severe anemia, heart, failure, or shock, diabetic ketoacidosis are the common reasons of metabolic acidosis, which can lead to serious consequences including coma and death.

- Respiratory alkalosis is a condition of low bicarbonate and high blood pH (more than 7.45) because the levels of carbon dioxide and oxygen in the blood are not balanced. Hyperventilation, fever, pain and anxiety are the common reasons of respiratory alkalosis
- Respiratory acidosis is a condition of abnormally increasing carbon dioxide levels in the blood, the blood pH is low (less than 7.35). It occurs when the lungs can not remove enough of the carbon dioxide produced by the body. There are various reasons for respiratory acidosis ranging from pneumonia, chronic obstructive pulmonary disease, asthma, pulmonary fibrosis, exposure to toxic chemicals, drugs that suppress breathing, tuberculosis, lung cancer, pulmonary hypertension and severe obesity.
- Metabolic alkalosis is a condition of both bicarbonate and pH level in blood are higher than normal. The pH figure is more than 7.45. The common reasons of metabolic alkalosis are chronic vomiting, low potassium levels, hypoventilation, and decreased CO_2 elimination.

Electrical activity in brain is caused by the communication activity of neurons across short and long ranges in the human brain. The abnormal change in electrical activity can lead to seizures, especially epilepsy. Moreover, the information is applied to identify mental health problems and physical problems such as problems in the brain, spinal cord, or nervous system of the patients.

Electrodermal Activity (EDA) or Galvanic Skin Response (GSR) refers to electrical properties of the human body captured at the surface of the skin that arise when the skin receives innervating signals from the brain. When individuals encounters nervous, emotional activation, increase cognitive workload or physical exertion, the brain sends signals to the skin to increase sweat gland activity. The result of these actions is the increasing of skin conductance. Consequently, skin conductance can be applied to measure emotional and sympathetic responses such as relaxation or distress biofeedback, emotional mapping, the polygraph test of an individual.

2.3 The Environmental Aspects and Indoor Location Information

Since people spend about 80-90% of their time indoor, the research at [18] indicated that indoor environmental quality can impact on the comfort, health and productivity of the individual who lives inside the house. Abnormal change of indoor environment may cause dangerous situations and result in undesired consequences of directly or indirectly harming occupants [19]. A range of indoor factors such as acoustic, thermal, lighting aspects, moisture, mould, noise and vibration, visual, radiation, chemical compounds and particulates are the examples of many indoor stressors. Sleep and the life quality can be affected and reduced by high or low thermal condition. High temperature and humidity can not only cause heat diseases but also result in slow to react or inappropriate judgement that indirectly cause harm with respect to nursing or taking care of young children. These indoor stressors can exert their effects additively or through complex interactions (synergistic or antagonistic) [20]. These relationships can have both short-term and long-term impact on the wellbeing of the occupants. The recent researches have specified that there is a link between indoor building conditions and mental health; cardiovascular diseases, asthma-related issues and obesity which are not easily detectable in the short term, however can be considerable problems in the long term [21]. Therefore, along with human's health-related information, the environmental aspects as gas, smoke, light, temperature, sound, oxygen and humidity also have their roles to determine the environmental health and safety of an indoor space .

Indoor location is another neccesary information. In recent years, there is an increasing demand for indoor localisation, especially in healthcare area.

The indoor location information is useful for positioning people within buildings such as hospitals and nursing homes to track patients, either for their safety or their care. For specific mental disease like Alzheimer disease, there is the problem of lossing or forgetting patients around the hospital. For emergency response such as when a patient has suddenly fallen down or has not moved for a period of time, he or she may require a prompt assistance from doctors or other healthcare providers to avoid additional injuries, the location information can help to determine where in the facility a patient is located. Furthermore, a health facility also need location information for tracking expensive equipment, preventing theft, and precise locating for robotic assistants during surgeries [23].Therefore, observation of the precise position of a person inside a medical facility at any time is a significant problem inside indoor localization research area.

For senior citizens living alone, indoor location information can be used to monitor their daily activities and ensure the safety. Besides, daily movement patterns are effective for detection of early indications of new or deteriorating health issues [24].

Chapter 3

Supporting Infrastructure and Technology

3.1 Introduction

Recent advances in sensor technology, microelectronics, and telecommunication, both physiological, environmental and location information can be easy collected and monitored. The environmental sensors can be used to capture environmental parameters. Accelerometer and compass sensors are applied in location monitoring application while the biomedical sensors such as Electrocardiogram (ECG), Electrodermal Activity (EDA) and Photoplethysmography (PPG) sensors can be used to monitor the health figure of the human [25]. The collected data can be analyzed and sent to the server by applying communication technologies. There is a wide range of communication technologies in smart health field which is well known such as RFID, NFC, WiFi, 6LowPAN, Bluetooth, ZigBee and 2G/3G/4G cellular.

3.2 Sensor Technology

3.2.1 The Advance in Sensor Technology

Sensors which is the bridge between the physical world and electronic system, play a critical role in IoT healthcare fields. The analog signals corresponding to body's physiological actions, human activities, location and environmental aspect are collected and then forwarded to the radio receivers. Commonly, there are three main building blocks in a wireless sensing node including sensors, processing, and wireless electronics. These blocks are embedded on printed circuit boards (PCBs) made of the glass-refined epoxy laminate (FR4) while the flexible material such as polyimide has been applied in the field recently [7]. Because these sensors directly communicate with human or are even implanted, size of sensors and front-end electronics has become one of the significant difficulties to the adoption of sensing technology for many years. Since Moore's law [26] is also applicable for sensors which become smaller and smaller, sensors will be embedded in more devices than the imagine of a human.

Traditionally, textitinvasive sensors [27], in which the probe must enter the human body through the natural cavities (nostrils, throat, ears, skin), are widely used in healthcare applications. Due to the enormous potential of continuously monitoring physical and chemical parameters of the patient, optical becomes the key technology of invasive sensors. Besides, the possibility of accurate measurements improves thanks to this direct connection. Moreover, invasive sensors become small enough to fit in the tissue with minimal damage, either as a part of a catheter or some other probe. However, the limitation of invasive sensors is in where and when they can be applied and the comfort of a human.

The miniaturization of electronic circuits based on the use of microelectronics has taken an essential part in the advancement of sensor technology [24]. Moreover, the development of signal processing and especially Micro-Electro Mechanical Systems (MEMS), sensory data can be monitored in *a non-invasive fashion.* MEMS is an advanced technology for sensors design based on miniaturized mechanical and electro-mechanical components (i.e., devices and structures) that are made utilizing the techniques of micro-fabrication. Accordingly, sensors progressively tinier in the scope of 1 to 100 micrometers. As a result, a new chance of ubiquitous healthcare applications has been created with the low medical cost, independence, more comfortable and high quality of healthcare services. MEMS technology has been applied to the design of different kinds of sensors such as an accelerometer, blood glucose, blood pressure, carbon dioxide (CO_2) gas sensor, ECG, EEG, EMG, gyroscope, pulse oximetry, as well as some sensors typically used in WSNs recently. However, there is an increasing failure of electrical contacts along with skin irritation problems due to the long-term utilization of these types of electrodes. Fortunately, this issue can be alleviated by applying the textile structured electrodes of MEMS technology. These textile-structure electrodes or smart textiles contain pressure, chemical, humidity and temperature sensors in clothes fabrics which will not cause any skin irritation, adaptable shape to human activity. Therefore, smart textiles are comfortable, flexible, and advisable for long-term monitoring.

3.2.2 Some Common Sensor Devices

In this section, we introduce some commonly available sensor devices for the smart healthcare system. These sensors can be classified into two main categories: (1) contextual sensors and (2) physiological sensors. Contextual sensors are the embedded sensors in the environment around patients such as temperature sensors, weather sensors, motion detection sensors, audio, video sensors, location sensors to measure various contextual properties. Physiological sensors capture patient vital signs or physiological statistics such as temperature, pulse rate (heart rate), respiration rate (breathing rate) and blood pressure or Electrocardiography (ECG), blood glucose sensor, etc. These two main categories are summarized in table 3.1 The feature applications in the summarization are based on the technological readiness levels (TRL) with TRL = 4 indicating in-lab component validation through to TRL = 9 where technology is in its final form, being used under operational conditions.

Table 3.1: Some common sensor devices in healthcare applications.

Sensor	*Measurement*	*TRL*	*Clinical Focus*
	Environmental Sensors and Location Sensors		
Pressure Sensor	Pressure on mat, chair, etc.	6	Human detection and tracking [31]
Humidity Sensor Temperature Sensor	Water in air Temperature	9	Indoor abnormality detection (Sensorcon®) [37]
Accelerometer Gyroscope	Direction Orientation	6	Indoor location monitoring [113]
	Physiological Sensors		
Electrocardiography (ECG) Photoplethysmogram (PPG)	Pulse rate and variability Pulse rate and blood variability	6	Cardiac arrhythmia [30]
Electroencephalogram (EEG)	Brain activity	7	Epilepsy (NeuroPro®)[32]
Electromyography (EMG)	Muscle activity	6	Neonatal intensive care unit [33]
Thermal	Body temperature	6	Infection [34]
Glucometer	Blood glucose	9	Diabetes (Dexcom®) [36]
Blood Pressure (BP)	Oscillometric	9	Hypertension (iHealth®) [35]

Humidity and temperature sensor: measuring the temperature of the human body and the humidity of the immediate environment around a person.

Electrocardiography (ECG) sensor: ECG is a process of amplifying and recording the heart's electrical activity that controls the expansion and contraction of heart chambers. ECG provides graphic information about the part of the heart that triggers each heartbeat, the nerve conduction pathways, the rate and rhythm of the heart. This information is used to help healthcare provider analyze a cardiovascular disease.

Blood glucose sensor: Measuring the amount of glucose circulating in the blood. Commonly, blood glucose-level measurement requires taking a blood sample by pricking a finger which causes pain or trauma. A blood sample is placed and reacted on a test strip that changes color when reduced. An optical meter is used to analyze the blood sample and gives a numerical glucose reading. Recently, through infrared technology and optical sensing, the techniques that do not require drawing blood or noninvasive glucose monitoring has become available [28].

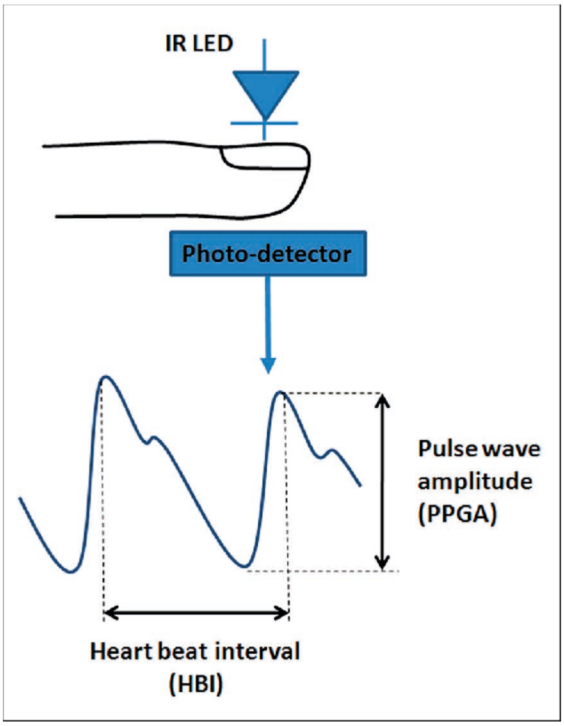


Figure 3.1: The standard process for capturing PPG signal and example of PPG signal [29].

Electroencephalography (EEG) sensor: measure the electrical activity of the brain by placing small electrodes on the scalp. EEG measures voltage fluctuations resulting from ionic current within the neurons of the brain. The brain's electrical activities data sensed by the electrodes is sent to an amplifier for generating a pattern of tracings. Synchronous electrical activities in different brain regions are generally assumed to imply functional relationships between these regions.

Electromyography (EMG) sensor: measure the electrical activity of muscles during contractions or at rest. The more body fat an individual has, the weaker the EMG signal. Since nerves control the muscles in the body by electrical signals, nerve conduction studies are often done together while measuring the electrical activity in muscles. The ideal location for placing the EMG sensor is the longitudinal midline.

Photoplethysmogram (PPG) or scientifically called Blood Volume Pulse (BVP) use a light-based technology to detect the rate of blood flow as controlled by the heart's pumping activity. PPG makes uses of low-intensity infrared (IR) light. When light travels through biological tissues, it is absorbed by bones, skin pigments and both venous and arterial blood. Since light is more strongly absorbed by blood than the surrounding tissues, the changes in blood flow can be detected by PPG sensors as changes in the intensity of light. The voltage signal from PPG is proportional to the quantity of blood flowing through the blood vessels. Even small changes in blood volume can be detected using this method, though it cannot be used to quantify the amount of blood. Figure 3.1

illustrates the standard process for capturing PPG signal proposed based on the research of Bonhomme [29] and example of PPG signal. The derived unit of measurement for PPG is $nanoWatt(nW)$. From PPG data, researcher can apply several algorithms to extract the interbeat interval (IBI) or well-known as heart beat interval (HBI) data, which is the time interval between individual beats of human. The derived unit of measurement for IBI is $second(s)$. Heart rate (HR) can be simply calculated by $60/IBI(bpm-beatperminute)$.

Electrodermal Activity (EDA): There are various of methodologies to measure EDA for example admittance, conductance, impedance, resistance and skin potential. Measuring electrical conductance across the skin is one of the popular methods to capture EDA figure. By passing a minuscule amount of current between two electrodes in contact with the skin, the information can be achieved. MicroSiemens (μS) is the derived unit of measurement for conductance. The example of EDA signal is shown in Figure 3.2

Thanks to technological advances as mentioned in Section 3.2.1, invasive and non-invasive sensors can be placed in any part of the body to collect not only physiological data but also contextual data. For example, sensors placed on the brain can capture glucose level, ECC, PPG signal while wearable sensors on the wrist can measure skin conductance level (EDA), the activity level of an individual. Figure 3.3 depicts a graphical demonstration of physiological sensor placement on the human body.

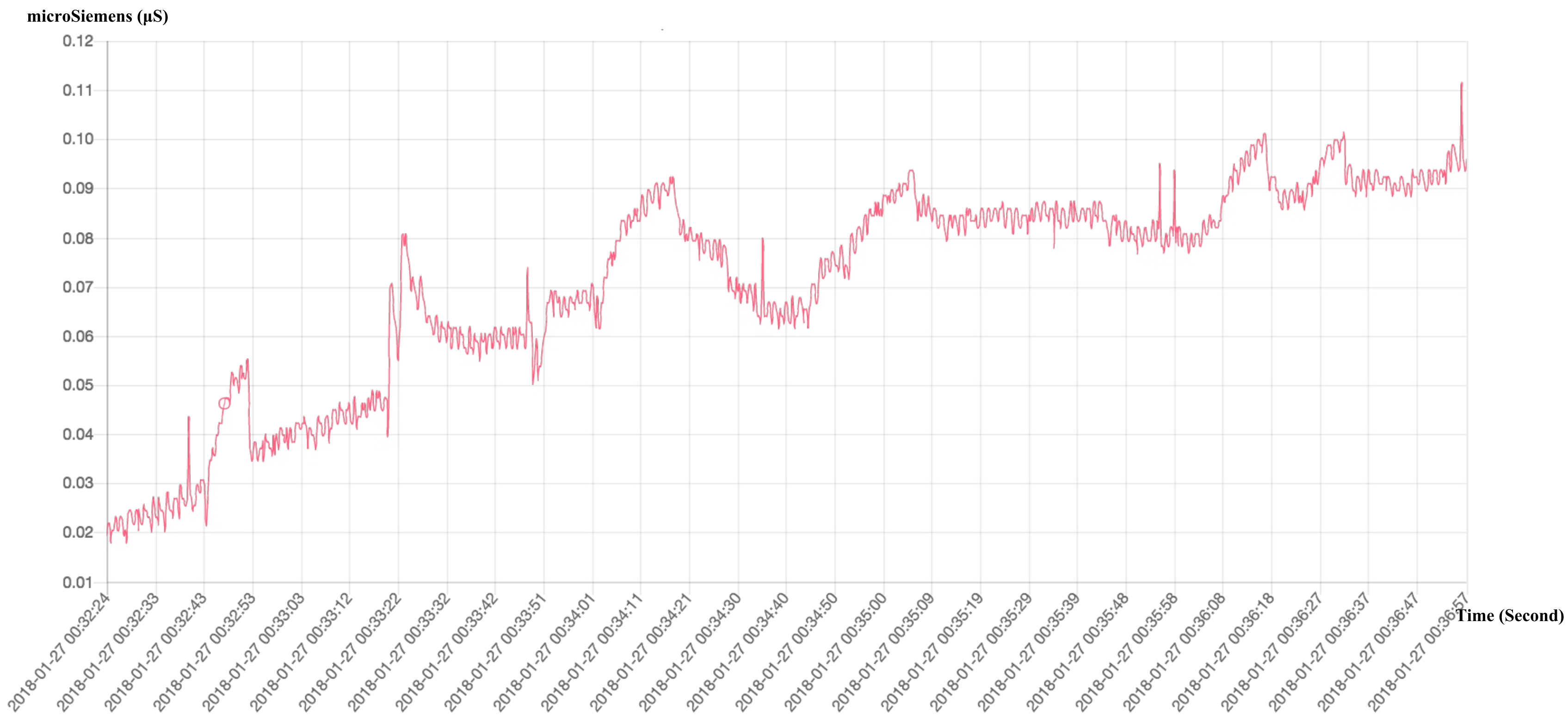


Figure 3.2: The example of EDA signal.

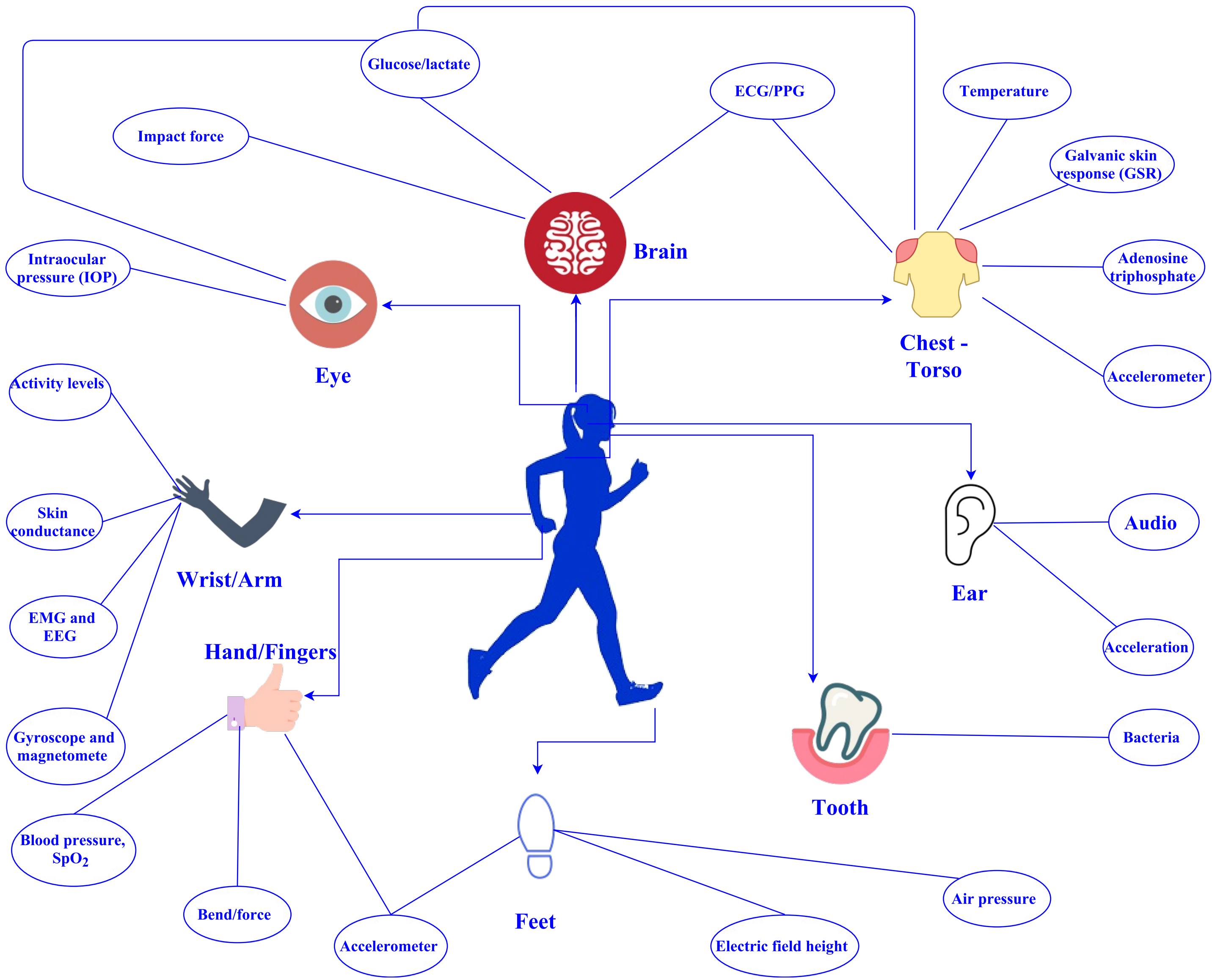


Figure 3.3: Graphical demonstration of physiological sensor placement.

3.3 Communication Technology

The sensors and equipment are connected through a communication network to share and exchange data. There are three purposes for the transmission of measured data in the overall context of smart healthcare needs. Firstly, for retrieving sensory data from human body and environment. Secondly, for transmitting the collected physiological signals from that biosensors to the system's central node. Finally, for sending the aggregated measurements from the local system to remote medical stations.

The communication between sensors and system's central node can be handled either by wires or by multiple wireless networks. In the past, the utilizing of wires not only severely hindered the user's mobility and comfort but also increased the risk of system failure [38]. Many advanced technologies have been applied to overcome this problem. For example, conductive yarns have been used to transfer the collected data from sensors integrated into some flexible smart-textile clothing [39].

Currently, autonomous sensor nodes can construct a body area networks (BANs) or body sensor network (BSN).

3.3.1 Body Area Networks Architecture

The development of BANs has been empowered by the extensive use of wireless networks and the constant miniaturization of electrical devices. There are several benefits introduced by using wireless BANs in healthcare application, mainly, flexibility, communication efficiency and cost-effective. Positively, non-invasive sensors can be used to flexibly monitor and transmit physiological data to the central node of BAN, then forward to nearby devices based on the application needs. Moreover, the signals that body sensors provide can be efficiently processed to obtain reliable and accurate physiological estimations. Besides, the ultra-low power consumption of such sensors makes their batteries long lasting. Furthermore, more sensors, especially for healthcare purposes, will be mass-produced at a relatively low-cost thanks to the increasing demand of body sensors in the consumer electronics market. BANs may interface with other wireless technologies, such as WSNs, radio frequency identification (RFID) technology, Bluetooth, Zigbee, Bluetooth Low Energy, video surveillance systems, wireless personal area network (WPAN), wireless local area networks (WLAN), internet, and cellular networks.

Figure 3.4 depicts the research by M. Chen [40] which clearly separated the BAN communications system into three-tier architecture. The three different layers are tier 1 or intra-BSN communication, tier 2 or inter-BSN communication, and tier 3 or beyond-BSN communication. These three architectural layers cover various aspects of communication, from low-level to high-level design problems, assist the progress of a component-based, efficient BAN system for multiple healthcare service provisionings.

The terminology "intra-BAN communications" known as tier-1 refers to the wireless communication that directly connects to a human body with the coverage of about 2 meters. The construction of intra-BAN is essential because of this direct relationship with body sensors. The paradigms of the layer can be further subcategorized as: (1) the communication between body sensors, which are strategically attached or implanted on

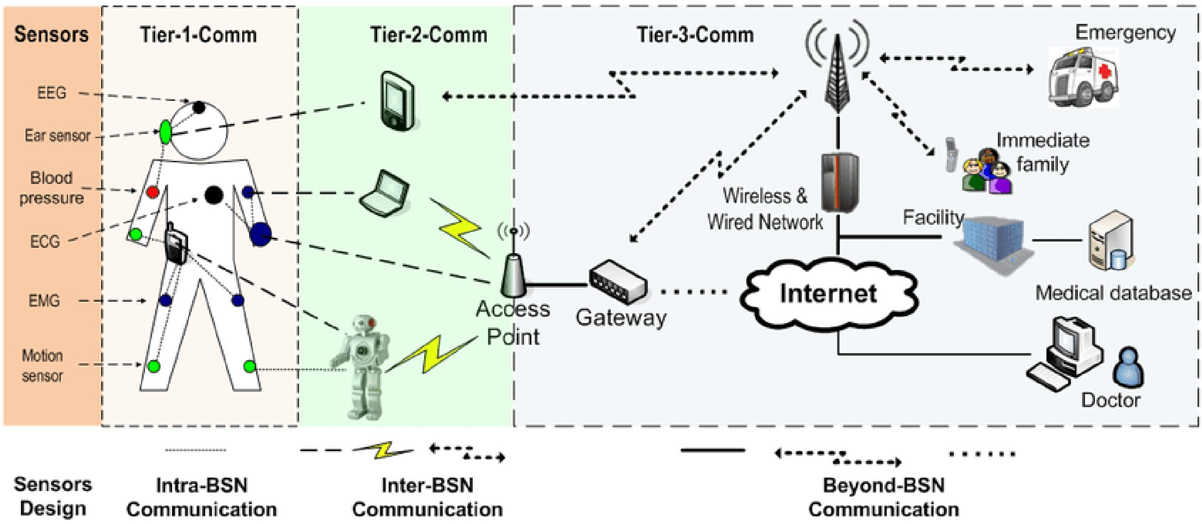


Figure 3.4: A three-tier architecture based on a BAN communication system [40]

the patient's body, as well as deployed within the human's clothing; (2) the communication between the body sensors and the portable Personal Server (PS).

The "inter-BSN communication" known as tier-2 refers to the communication between the PS and one or multiple access points (APs). These APs can be deployed as a part of the system's component or strategically placed in dynamic environment to handle emergencies. According to the terminology of "inter-BSN", the tier-2 has capability to interconnect BSNs with various networks ranging from internet to cellular networks. The paradigms of inter-BAN communications are devided into two categories, infrastructure-based architecture, and ad hoc-based architecture. The benefit of infrastructure-based architecture is the ability to provide larger bandwidth with concentrated control and flexibility while the ad hoc-based architecture can be deployed faster in a dynamic situation, for example in medical emergency response, or at a disaster area [41]

Tier-3 or the "beyond-BSN communication" is designed for streaming collected data to the remote healthcare applications and systems utilizing cellular network or the Internet. A gateway device, such as a personal digital assistant (PDA), is usually employed to create a wireless link between inter-BAN and beyond-BSN communications. Depending on the specificity of services and the requirement of user-specific applications, there is the proper design of beyond-BSN communication.nding on the specific characteristics of services and the requirement of user-specific applications; there is the appropriate design of beyond-BSN communication.

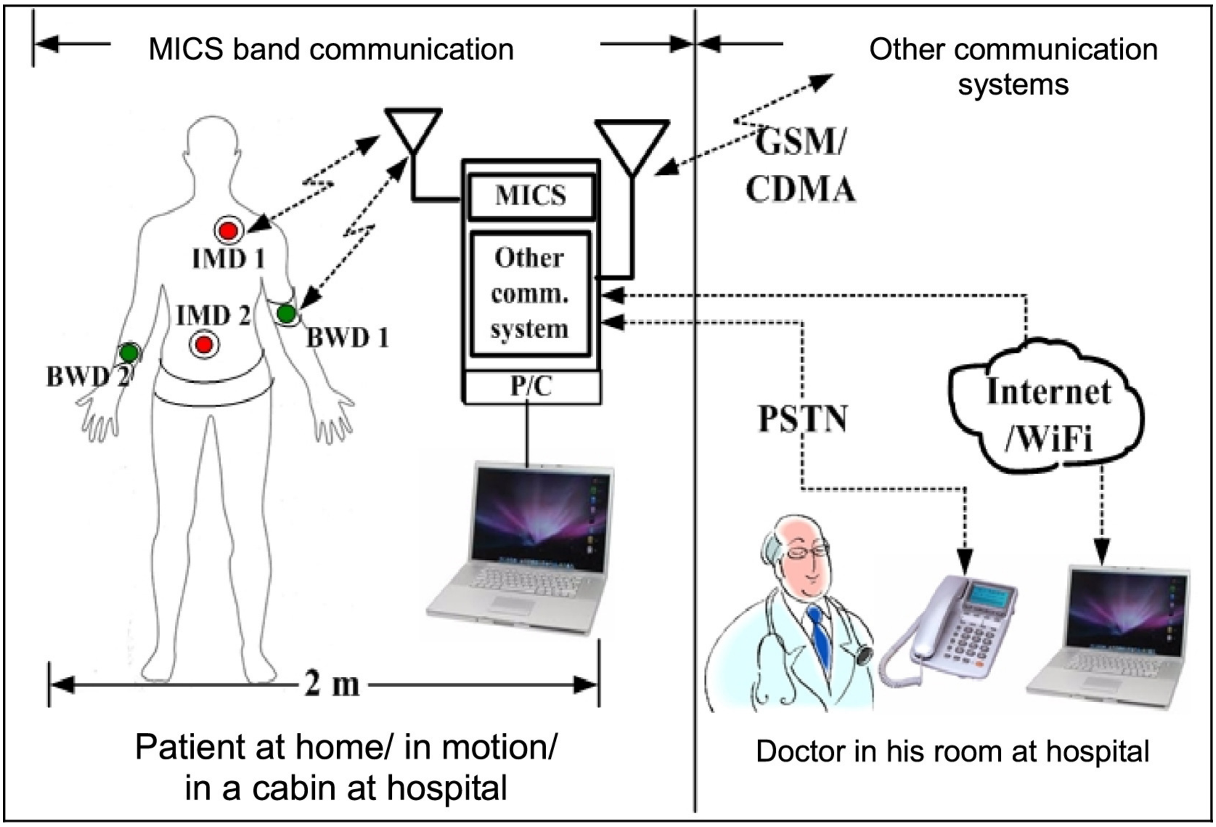


Figure 3.5: Complete MICS communication network. [48]

3.3.2 Wireless Communication Technology

IEEE 802.15.1 (Bluetooth), 802.15.4 (the basis for Zigbee) and 802.15 (standard for wireless personal area network - WPAN), presented in table 3.2, are the most commonly employed wireless communication standards in BANs.

The Zigbee standard [42] is refers to a low-cost, low data-rate, low complexity however long battery life solutions along with a set of globally accepted specifications for wireless sensor networks. The radio design utilized by Zigbee has been strategically optimized in order to archive low cost in large scale production. This standard specifies operation in 16 channels in the 2.4 GHz industrial, scientific, and medical (ISM) band (250 kb/s, OQPSK modulation), in 10 channels in the 915 MHz band (40 kbps, BPSK modulation) and in one channel in the 868 MHz band (20 kb/s, BPSK modulation). Zigbee supports various network architectures including star, tree cluster, and mesh topologies with the maximum transmission range is about 75 meters. There is an extensive potential of ZigBee standards in healthcare area. These standards have been applied in a wide range of applications including healthy and independent living support for the disabled or elderly, foster safe, wellness and fitness [43].

Bluetooth [44] is a low power and linexpensive wireless technology standard for short-range radio frequency connectivity not only between fixed but also mobile devices. This

Table 3.2: Wireless communication protocols in BANs.

Technique/Parameters	*Range (typical)*	*Data Rate (max)*	*Frequency*
Zigbee	10–75 m	20 Kbps 40 Kbps 250 Kbps	868 MHz 915 MHz 2.4 GHz
Bluetooth	10–100 m	1–3 Mbps	2.4 GHz
Ultra wideband	2 m	500 kbps	400 MHz
Infrared	1 m	16 Mbps	-
MICS	2 m	500 kbps	402-405 MHz

standard operates in the unlicensed 2.4 GHz spectrum. After breaking transmitted data into packets, bluetooth transmits each of these packet on one of 79 designated Bluetooth channels with the bandwidth of each channel is 1 MHz. Consequently, Bluetooth is able to transfer moderate amounts of data over a versatile, robust and secure wireless connection. Although the common transmission distance is around 10 meters, the maximum figure can reach to 100 meters. Piconet is the basic configuration of Bluetooth. It is an ad hoc network with one master interconnect with up to seven slaves, whereby the master provides the synchronization reference (common clock and frequency hopping pattern).

Introduced under the Bluetooth 4.0 specification by Bluetooth Special Interest Group (SIG) since 2004, Bluetooth Low Energy (BLE) or also called Bluetooth Smart is a wireless personal area network technology. BLE has shown great promise in the development of applications not only in the health monitoring and fitness but also in smart home industries. In spite of the similarity in some regards, BLE is not backward compatible with Classic Bluetooth protocols as a consequence of applying a different controller, for example, physical and link layer. The key difference between BLE and previous Bluetooth protocols is low power consumption. Thanks to the applying of BLE, with just a small battery can let applications run on for years. It is ideal choice for healthcare applications, which only need to exchange small amounts of data periodically. In 2018, SIG believe that BLE will be supported by 90 percent of Bluetooth-enabled smartphones in the market [45]. The comparison of classic Bluetooth and BLE is shown in Table 3.3.

Along with Zigbee and Bluetooth, the medical implant communication service (MICS) and ultra wideband (UWB) are emerging technologies applied in short-range intra-BAN communication which have many potential applications to be researched.

Due to their size, power consumption and strong interference from other devices, Zigbee or Bluetooth do not comply with the medical standard. Therefore, a new band Medical Implant Communication Service band (MICS) has recently allocated to deliver high level of satisfaction, flexibility and superior healthcare services by the USA (Federal Communications Commission) and other countries communication authorities. It is a low power, unlicensed, short-range (2 meters) mobile radio service which has been accepted world-

Table 3.3: Comparison of Classic Bluetooth and Bluetooth Low Energy (BLE) [46]

Specifications	*Classic Bluetooth*	*Bluetooth Low Energy*
Range	100 m	Greater than 100 m
Data rate	1–3 Mbps	125 kbitps – 1 Mbps – 2 Mbps
Application throughput	0.7–2.1 Mbps	0.27 Mbps
Active slaves	7	Not defined
Frequency	2.4 GHz	2.4 GHz
Security	56/128-bit	128-bit AES with Counter Mode CBC-MAC
Robustness	Adaptive fast frequency hopping, FEC, fast ACK	24-bit CRC, 32-bit Message Integrity Check
Latency	100 ms	6 ms
Time Lag	100 ms	3 ms
Voice capable	Yes	No
Network topology	Star	Star
Power consumption	1 W	0.01 - 0.50 W
Peak current consumption	less than 30mA	less than 15mA

wide for transmitting high data rate to support of diagnostic or therapeutic functions associated with medical devices [47]. The universal radio frequency band of 402–405 MHz with 300 kHz channels is proposed in MICS. Effective isotropic rediated power (EIRP) is limited to 25 μW and targets mostly devices such as cardiac pacemakers and defibrillators, without causing interference to other users of the electromagnetic radio spectrum. Figure 3.5 present a high-level summary of the MICS band. Despite its beneficial element, because of the lack of commercially available solutions, MICS has not been utilized generally by scientists.

Ultra-wideband (UWB) radio is a technology that can use a very low energy level for short-range, high-bandwidth communications over a large portion of the radio spectrum [49]. UWB operates in the frequency range from 3.1GHz to 10.6GHz in America. However, the frequencies have been devided into two parts from 3.4 GHz to 4.8 GHz and 6 GHz to 8.5 GHz in Europe.

Chapter 4

Data Preprocessing

4.1 Introduction

In the healthcare field, especially in healthcare monitoring system, the reliability of input data is of extremely importance. Accurate health care decisions can only be made with accurate input data. However, sensor data is possible to contain noise and outliers due to the low quality or errors of the sensors, occurrence of noise, motion artifacts in any sensor networks. Consequently, these might lead to false readings, false alarms to be delivered to patients or healthcare providers. Therefore, a preprocessing of the raw sensor data is necessary.

Preprocessing in the healthcare domain involves filter imperfect data to remove artifacts or high-frequency noise; fill in missing values and imbalanced data preprocessing [50]. Many researchers have proposed various cleaning algorithms in order to increase the reliability of sensor data [51] - [52]. As an illustration, for filtering artifacts, designed modules have usually applied threshold-based methods to filter sensor data [53] or used statistical tools to interpolate the missing data points [54]. To remove frequency noise, the other methods in the frequency domain such as power spectral density (PSD) fast Fourier transforms (FFT), and low-pass/high-pass filtering tools are standard to remove the fluctuations in sensor signals [55]. For the missing data imputation, there are several methods such as Mean, K-nearest neighbors (KNN), fuzzy K-means (FKM), singular value decomposition (SVD), Bayesian principal component analysis (bPCA) and multiple imputations by chained equations (MICE) [56].

Among abundant technologies for data preprocessing, several algorithms emerge in popularity. In this section, we tackle the most important data preprocessing algorithms in health care system, which is presented in figure 4.1, based on the criteria proposed in [57]. They are:

- Usage: the algorithm is frequently used in a previous step of a DM process, or it is included in DM software packages.
- Referable: it must be described in a publication in the specialized literature.

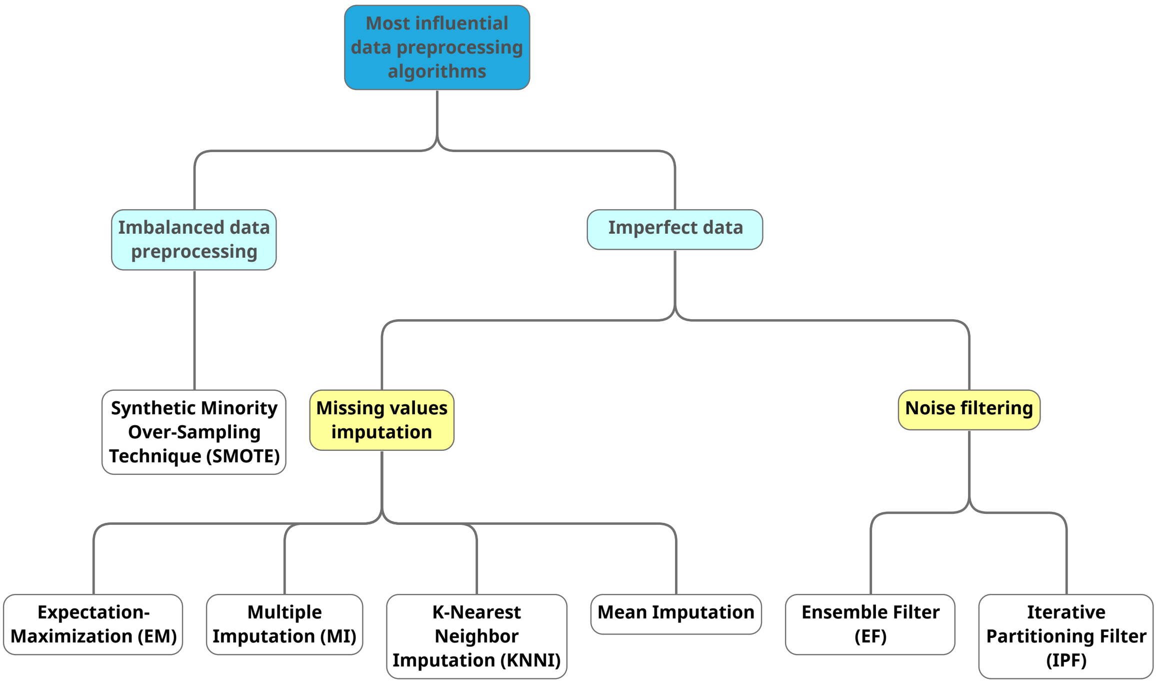


Figure 4.1: The classification of most influential data preprocessing algorithms in healthcare.

- Popularity: the associated publication is considered as a highly cited one in well-known databases: Web of Knowledge, Google Scholar, Scopus, etc.
- Standardization: the algorithm has been the baseline of inspiration of several modern and hybrid extensions.
- Smart: it must somehow incorporate a smart procedure in its definition, for the sake of not including direct and basic mechanisms as algorithms.
- Variability: there have been a minimum number of representatives belonging to each data preprocessing family.

Figure 4.2 illustrates the evolution of academic publications concerning the data preprocessing algorithms. Publication statistics are acquired from Google Scholar; the search query is defined as the subfield name of algorithms and at least one of medical or health appearing, e.g., “kNN Imputation” AND medical OR health.

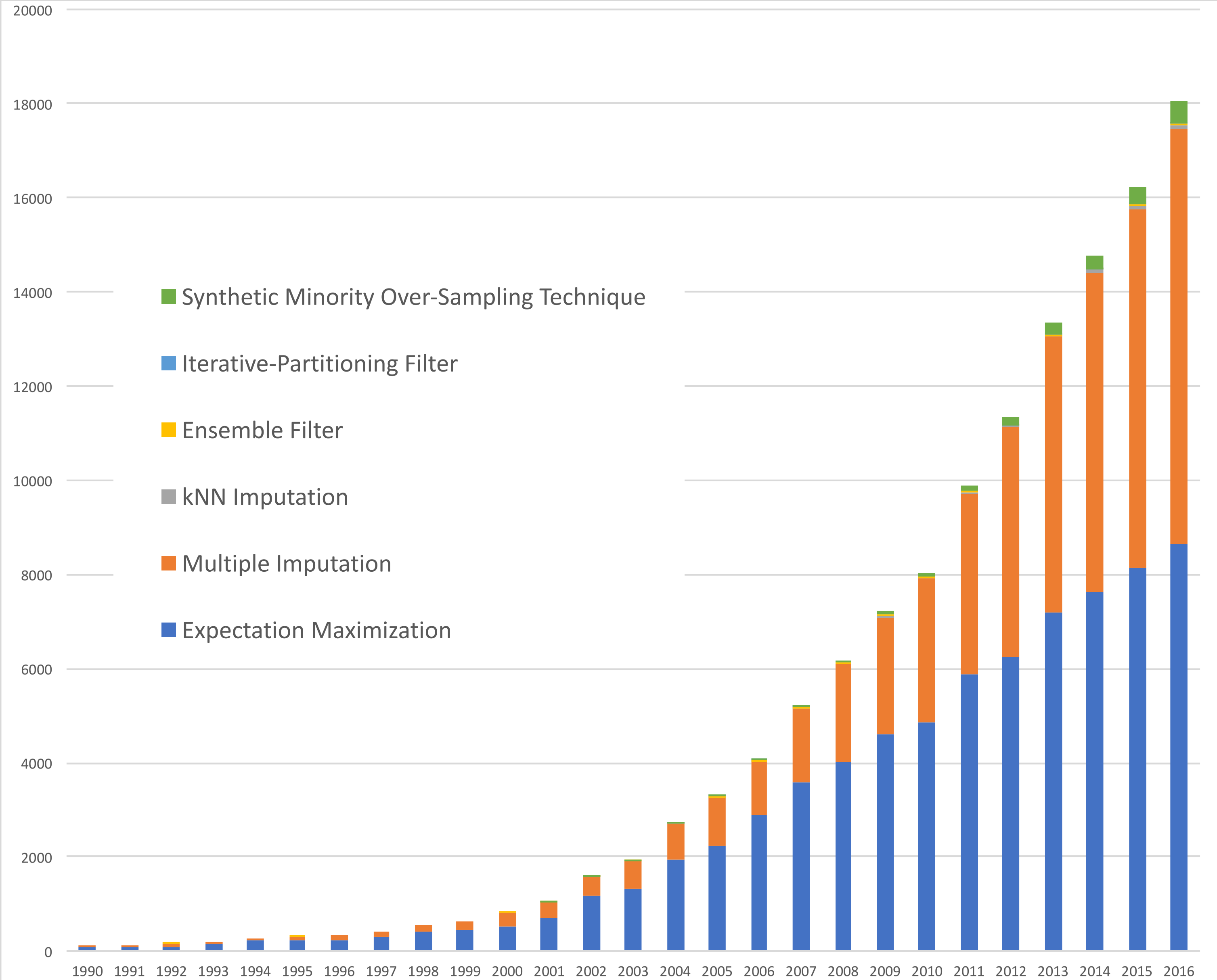


Figure 4.2: Evolution of academic publications concerning the data preprocessing frameworks.

4.2 Imperfect data

In health care or health informatics, the data is rarely clean or complete, imperfect data are prevalent due to missing or noise. Therefore, data analysis and modeling techniques in healthcare applications should be able to remove the noisy data and fill in the missing values. In this section, the widely used preprocessing techniques for imputing the missing values and for filtering the noise will be identified.

4.2.1 Missing Data Imputation

As a definition, a missing value is an attribute that has not been sampled in the data set, or that was never recorded, for any reason. The accuracy of applications in health care system depends on a data set that is supposedly complete. To support the health care tasks, the application expects to process sequences of complete instances sampled collected from sensors. However, due to various reasons such as equipment errors, incorrect measurements, limitations in the data acquisition process or faulty sampling, missing data are a common incidents. The presence of missing value makes the conduct of data analysis complicated, and it usually poses severe troubles for scientists. Improper handling of the missing value in the analysis can have a critical consequence of the decision that can be deduced from the data. We can point out three types of issue are usually associated with missing value: (1) loss of efficiency in the extraction process; (2) complications in handling and analyzing the data; and (3) bias due to incomplete data [58]. Therefore, cleaning and preparing the data is usually a required preprocessing stage in order to be helpful to and adequately clear for the knowledge extraction process.

To diminish the negative influence of missing data, there are usually three different approaches:

- The simplest, very usual however rarely practical method is to delete the samples that contain them. Only when a relatively low number of samples with missing values in data and analysis carried out over the remaining complete examples will not lead to serious bias during the inference, this simplest method can be practical.

- Traditional approaches to missing data come from statistics which are based on the applying of maximum likelihood procedures, where the parameters of a model for the complete data are estimated, while the probability functions are sampled to impute the missing value.

- The final goal of the missing value imputation is to fill in the missing value with estimated ones. In most cases, a data set's attributes are not independent of each other. Therefore, missing values can be determined through the identification of relationships among values.

Among a broad family of existing missing value imputation methods, we focus our attention on maximum likelihood and imputation methods which are widely used by researchers.

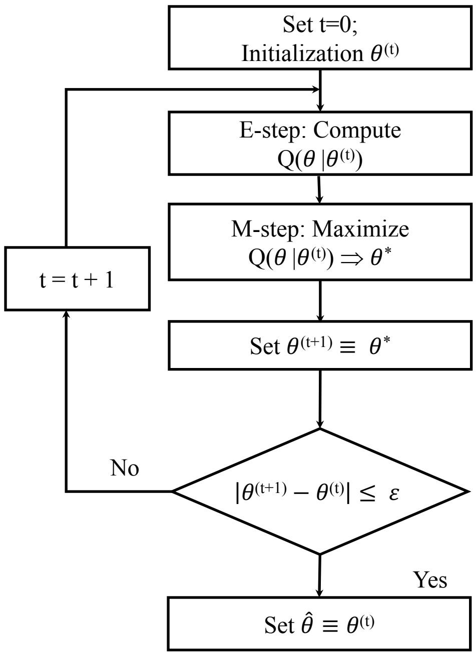


Figure 4.3: Flow chart for the expectation maximization algorithm

Regularized Expectation-Maximization (EM) [59] is a meta-algorithm applied to optimize the maximum likelihood of data. The given estimates of the mean and of the covariance matrix are revised in three steps in an iteration of the EM algorithm.

- First, the regression parameters of the variables with MVs among the variables with the available attribute are computed from the estimates of the mean and of the covariance matrix for each record with MVs.
- Second, the MVs in a record are filled in with their conditional expectation values given the available values and the estimates of the mean and the covariance matrix, the conditional expectation values being the product of the available values and the estimated regression coefficients.
- Third, the mean and the covariance matrix are re-estimated, the mean is considered as the sample mean of the completed data set and the covariance matrix as the sum of the sample covariance matrix of the completed data set and an estimate of the conditional covariance matrix of the imputation error.

The EM algorithm starts with initial estimates of the mean and the covariance matrix and cycles through these steps until the imputed values and the estimates of the mean and the covariance matrix stop changing appreciably from one iteration to the next. Since its proposal by Dempster et al. [59], the citation accounted by in Google Scholar is more than

Algorithm 1 EM algorithm.

function EM(T – data set with MVs, $stagtol$ – stagnation tolerance parameter across iterations)
  initialize: μ, Σ
  while $norm(x_{mis}^{it} - x_{mis}^{it-1}) \leq stagtol * norm(x_{mis}^{it-1})$ **do**
    //E-step
    Compute B for the current μ and Σ
    //Impute the MVs by using B
    $x_{mis} = \mu_{mis}^{it} + (x_{obs} - \mu_{obs}^{it})B + e$,
    //M-step. Reestimate μ and Σ
    $\widehat{B} = \widehat{\Sigma}_{obs,obs}^{-1}\widehat{\Sigma}_{obs,mis}$
    $\widehat{C} = \widehat{\Sigma}_{mis,mis} - \widehat{\Sigma}_{mis,obs}\widehat{\Sigma}_{obs,obs}^{-1}\widehat{\Sigma}_{obs,mis}$
    $\widehat{\mu}^{(it+1)} = \frac{1}{n}\sum_{i=1}^{n} x_i$
    $\widehat{\Sigma}^{(it+1)} = \frac{1}{\tilde{n}}\sum_{i=1}^{n}\left[\widehat{S}_i^{(t)} - (\widehat{\mu}^{(t+1)})\widehat{\mu}^{(t+1)}\right]$
  end while
  return T
end function

51,359 citations. Therefore, for missing value imputation, EM is one of the first successful solutions which applies maximum likelihood as a guaranteed approach. Thanks to its effective, the repercussion of EM has spread to numerous fields, especially in health care area. In 2016, there was 5500 research applied EM in health care application which was indexed by Google Scholar. Figure 4.3 illustrates the flow chart for the expectation maximization algorithm

Multiple Imputation (MI) [60] is a statistical algorithm for handling incomplete data sets. There are three required steps for the application of this algorithm: imputation, analysis, and pooling, which are depicted in figure 4.4.

- Imputation: in this step, the missing entries of the incomplete data sets are imputed for m times (m=3 in the figure). Imputed values, which can be different for each missing entry, are drawn from a distribution. The result of this step is m completed data sets. Because missing values can appear anywhere in the data set or missing entries could be related its value, the construction of this m completed data sets is the most challenging problem of MI.

- Analysis: in this step, each of the m completed data sets is analyzed. The result of this step is m analyses.

- Pooling: a final result is combined with the m analysis results thanks to the proposed simple rules.

The MI algorithm a prevalent method which was cited and compared in thousands of research articles. Since its proposal by Rubin et al. [60], the impact of this method is notable in the literature with 14813 citations while over 8820 research projects applied MI in health care applications accounted by Google Scholar.

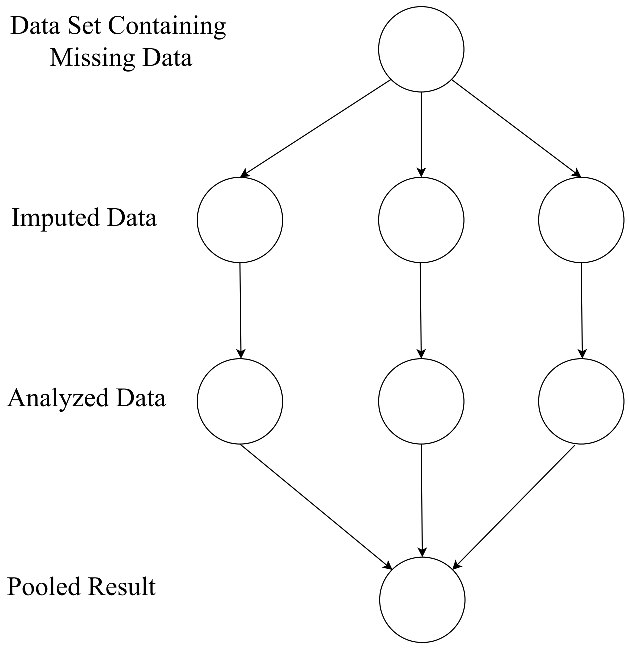


Figure 4.4: Flow chart for the multiple imputation algorithm.

Table 4.1: The number of publication referring to three widely used imputation data algorithms in healthcare applications from 2010 to 2016 by Google Scholar.

	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**
Expectation Maximization	4840	5880	6230	7190	7630	8150	8660
Multiple Imputation	3060	3830	4890	5850	6780	7600	8820
kNN Imputation	21	34	38	33	55	69	59

kNNI: kNN Imputation or K-Nearest Neighbor Imputation (KNNI) [61] Despite being extremely simple, the KNNI algorithm is extensively applied thanks to two basic properties:

- Firstly, it can adapt to any problem because only a dissimilarity function among two samples is needed. In this method, the k nearest neighbors are computed, and value from them is imputed.
- Secondly, both numerical and nominal values can handle with kNNI. The most general value among all neighbors is taken for nominal values, and for numerical values, the average value is applied. Therefore, a proximity measure between instances is needed for it to be defined.

Algorithm 2 MI algorithm.

```
function MI(T – data set with MVs)
    initialize: T^1 = T, T^2 = T, ..., T^l = T
    Obtain μ and Σ from EM(T)
    for 1 to k do
        for i = 1 to l do
            x^i_mis = μ_mis + (x^i_obs − μ_obs)B + e,
        end for
        Update θ = μ, Σ using the Bayesian posterior distribution
    end for
    for i = 1 to l do
        Obtain the model R^i from T^i
    end for
    R̄ = (1/m) Σ_{i=1}^{m} R̂^i
    Ū = (1/m) Σ_{i=1}^{m} U^i
    V = (1/(m−1)) Σ_{i=1}^{m} (R̂^i − R̄)^2
    Q = Û + (1 + 1/m)V
    return T
end function
```

Many pieces of research have widely accepted kNNI as a simple but effective method for dealing with the missing value. From machine learning point of view, the utilization of kNNI as an imputation algorithm has been a perceptive solution. In spite of being cited and compared in hundreds of research projects, the application of kNNI, especially in the healthcare field, is still tiny compared with EM and MI algorithm. In 2016, there are only 59 projects which applied EM to solve the problem in health care. Therefore, according to its good behavior, there is still a vast potential of applying kNNI method in the application of healthcare. Table 4.1 demonstrates the number of publication referring to three widely used imputation data algorithms in healthcare applications from 2010 to 2016 by Google Scholar.

4.2.2 Noise filtering

As we have previously mentioned, the quality of the results obtained in an application is mainly determined by the data quality. Most of the algorithms consider that there are no disturbances in the data which is a precise example of the underlying distribution. Nevertheless, there is a truth that the data collected by sensors is rarely proper and it regularly suffers from corruptions. The noise affects not only input but also output features. When noise affects in the input attribute, it is commonly indicated as attribute noise. Besides, when the noise exists the output attribute, it becomes the worst case; consequently, it leads to the consequence that the bias introduced will be greater. Therefore, noise filtering, a process of removing noise from an underlying signal, is another typical challenge.

Algorithm 3 kNNI algorithm.

function kNNI(T – data set with MVs, k – number of neighbors per instance to be chosen, $D(x_i, x_j)$ – a distance or dissimilarity function of x_i and x_j, S – the imputed version of T)
  initialize: $S = \{\}$
  for each instance x_i in T **do**
    $\widehat{x_i} \leftarrow x_i$
    if x_i contains any missing value **then**
      Find set I_{Ki} with the k nearest instances to x_i from T using D
      for each missing value in attribute h of x_i **do**
        if h is numerical **then**
          $\widehat{x}_{ih} = \left(\sum_{j \in I_{Kih}} x_{jh}\right) / (|I_{Kih}|)$
        else
          $\widehat{x}_{ih} = mode(I_{Kih})$
        end if
      end for
    end if
    $S \leftarrow \widehat{x}_{ih}$
  end for
  return S
end function

Accompanying with the objective of increasing the classification accuracy, there are three regularly main approaches to deal with noise:

- Robust learners: The classifiers, which are less influenced by noise, are applied or adapted to create the robust learners. C4.5, the algorithm benefits pruning strategies to reduce overfitting caused by noise, is a traditional method of robust learners.

- Data polishing techniques: it is an ideal option to correct noisy instances unless a robust learner method can be applied. These methods are very time to consume, therefore, they are limited to small data sets. However, researchers claim that even partial noise correction enhances the performance results.

- Noise filters: it is the most well-known method which benefits noise filters to analyze and remove instances in the training set. Noise filters are faster than adopting data polishing techniques and do not require the learning algorithm to be changed.

Comparing with Missing Data Imputation, researchers pay much less attention to Noise Filtering in the healthcare area. Table 4.2 presents the number of publication referring to two widely used noise filtering algorithms in healthcare applications from 2010 to 2016 by Google Scholar. Since data polishing techniques are not as popular as noise filters in preprocessing, we focus our attention on these noise filtering methods.

Table 4.2: The number of publication referring to two widely used noise filtering algorithms in healthcare applications from 2010 to 2016 by Google Scholar.

	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**
Ensemble Filter	18	19	18	23	25	30	31
Iterative-Partitioning Filter	1	0	2	1	5	7	13

Ensemble Filter (EF) [62] is a popular filter algorithm. The simplest solution for a filter is to discard the entries that are being misclassified by a classifier. However, this technique can be incorrect or invalid for other models or classifiers because of being biased by the nature of the classifier. Instead of using only a single learning algorithm, EF applies a set of learning methods over several subsets of the training data to support the aim of improving the filtering process. However, there is a disadvantage of the EF filter that it does not impose on the new cleaned training data set which created by itself.

The EF filter is the first to successfully come up with the use of an ensemble to filter noisy data and to attain good behavior for many classifiers used later over the cleaned data set. The influence of this algorithm is distinguished since the seminal work by Brodley et al. [62] achieved 450 citations in Google Scholar in 2016. However, Google Scholar also figures out that EF remains its infancy in health care area since there is just 31 EF health care related research during the year 2016.

Iterative-Partitioning Filter (IPF) is a kind of noise filtering technique, which discards noisy instance in multiple iterations till a termination criterion is met [63]. Based on the Partitioning Filter, the main idea of the Iterative-Partitioning Filter is obtain more accurate filters, that can be applied to clean the dataset iteratively. The method can also identify and remove mislabeled attributes in large training sets. IPF can be more or less conservative by fluctuating the number of filtering repetition. IPF is affected by the voting schemes, majority, and consensus.

According to the advance, IPF has started to be broadly used from imbalanced domains to semi-supervised issues, especially in healthcare applications.

4.3 Imbalanced Data Preprocessing

In the classification, an imbalanced dataset is not only a common but also serious problem. The reason of this scenario is a skewed distribution of data between classes. The samples of one class extremely exceed the sample of the other. Unfortunately, the most critical approach of the data is commonly represented by the minority class. Since this positive class might be related to extraordinary and essential instances or the data collection of these samples is expensive, it is tough to recognize. Besides, there usually the association between the imbalanced class issue and binary classification along with various minority classes, the problem become more and more difficult. As a result, most of the standard techniques suppose balanced class distribution. Consequently, these algorithms fail to accurately represent distributive characteristics of data when presented with solid imbalanced data sets. In other to overcome the imbalanced data, there are two approaches can be applied on data and algorithm levels:

- On data level, resampling techniques to get the balanced distribution was applied in most of the studies. Undersample the majority class, oversample the minority class or both of them can be coincidently used. Synthetic minority over-sampling technique (SMOTE) is a most common method on the data level.

- On algorithm level, boosting techniques can be applied. Boosting techniques or adjust misclassification cost is a task of constructing strong classifier from some weak classifiers which are called baseline classifiers. The common baseline classifiers are SVM and KNN algorithms.

Synthetic Minority Oversampling Technique or SMOTE algorithm [64] a statistical technique for dealing with imbalanced data. Since its first proposal in 2002, SMOTE becomes sophisticated methods and the most renowned approaches in the field. Concisely, the main idea behinds SMOTE is to form new positive class instances by interpolating between several positive class instances that lie together for oversampling the training data. This positive class is structured by taking each minority class instance and applying synthetic examples along the line segments joining any or even all of the k minority class nearest neighbors. Depending on the amount of over-sampling required, neighbors from the k nearest neighbors are randomly chosen.

Figure 4.5 illustrates process of creating the synthetic data points in the SMOTE technique, where x_i is the selected point, x_{i1} to x_{i4} are some selected nearest neighbors and r_1 to r_4 the synthetic data points created by the randomized interpolation.

In order to generate the synthetic samples based on the difference between the typical example under consideration and its nearest neighbor. Then, this difference is multiplied by a random number between 0 and 1. Under consideration, the result is added to the characteristic vector leading to the selection of a random point along the line segment between two particular features. In order to become more general, the method effectively determines the decision region of the positive class.

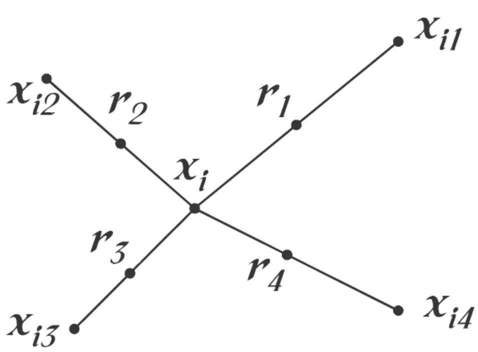


Figure 4.5: An illustration of how to create the synthetic data points in the SMOTE technique [65]

Algorithm 4 SMOTE

function SMOTE(T - Number of minority class samples; N% - Amount of SMOTE; k - Number of nearest neighbors)

If N is less than 100%, randomize the minority class samples as only a random percent of them will be SMOTEd.

if $N < 100$ **then**

Randomize the T minority class samples

$T = (N/100)T$

$N = 100$

$N = (int)(N/100)$ (The amount of SMOTE is assumed to be in integral multiples of 100.)

k = Number of nearest neighbors

$numattrs$ = Number of attributes

$Sample[][]$: array for original minority class samples

newindex: keeps a count of number of synthetic samples generated, initialized to 0

$Synthetic[][]$: array for synthetic samples

(Compute k nearest neighbors for each minority class sample only.)

for $i \leftarrow 1$ to T **do**

Compute k nearest neighbors for i, and save the indices in the $nnarray$

$POPULATE(N, i, nnarray)$ (Function to generate the synthetic samples.)

function POPULATE(N, i, nnarray)

while N $\neq 0$ **do**

Choose a random number between 1 and k, call it nn. This step chooses one of the k nearest neighbors of i.

for $attr \leftarrow 1$ to $numattrs$ **do**

$dif = Sample[nnarray[nn]][attr]$ - $Sample[i][attr]$

$gap = random$ number between 0 and 1

$Synthetic[newindex][attr] = Sample[i][attr] + gap * dif$

$newindex$++

$N = N$ - 1

return

Although the entire dataset is set as an input of SMOTE, only the percentage of minority class increases. For example, suppose there is an imbalanced dataset where 99% of the cases have the value A, just 1% of the cases have the target value B which is the minority class. To increase the percentage of minority cases to twice the previous percentage, we would enter 200 for SMOTE percentage in the module's properties.

Table 4.3: The number of publication referring to Synthetic Minority Over-Sampling Technique - SMOTE in health care applications from 2010 to 2016 by Google Scholar.

	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**
SMOTE	94	117	171	261	286	383	484

SMOTE one of the first popular and comprehensive algorithm for preprocessing data in imbalanced learning. The original paper has archived more than 2400 citations in Google Scholar so far. SMOTE become one of the most promising models of reducing the existence of imbalanced classes by over-sampling data thanks to its good behavior:

- Simplicity and ease of implementation in any application
- Guaranteed improvement in performance,
- Generality with any learning algorithm
- Adaptability and extensibility in the novel approach.

In healthcare area, SMOTE has started to be widely applied in the application of the fields. Table 4.3 presents the number of publication referring to SMOTE algorithm in healthcare applications from 2010 to 2016 by Google Scholar. Figure 4.6 presents the SMOTE flow chart.

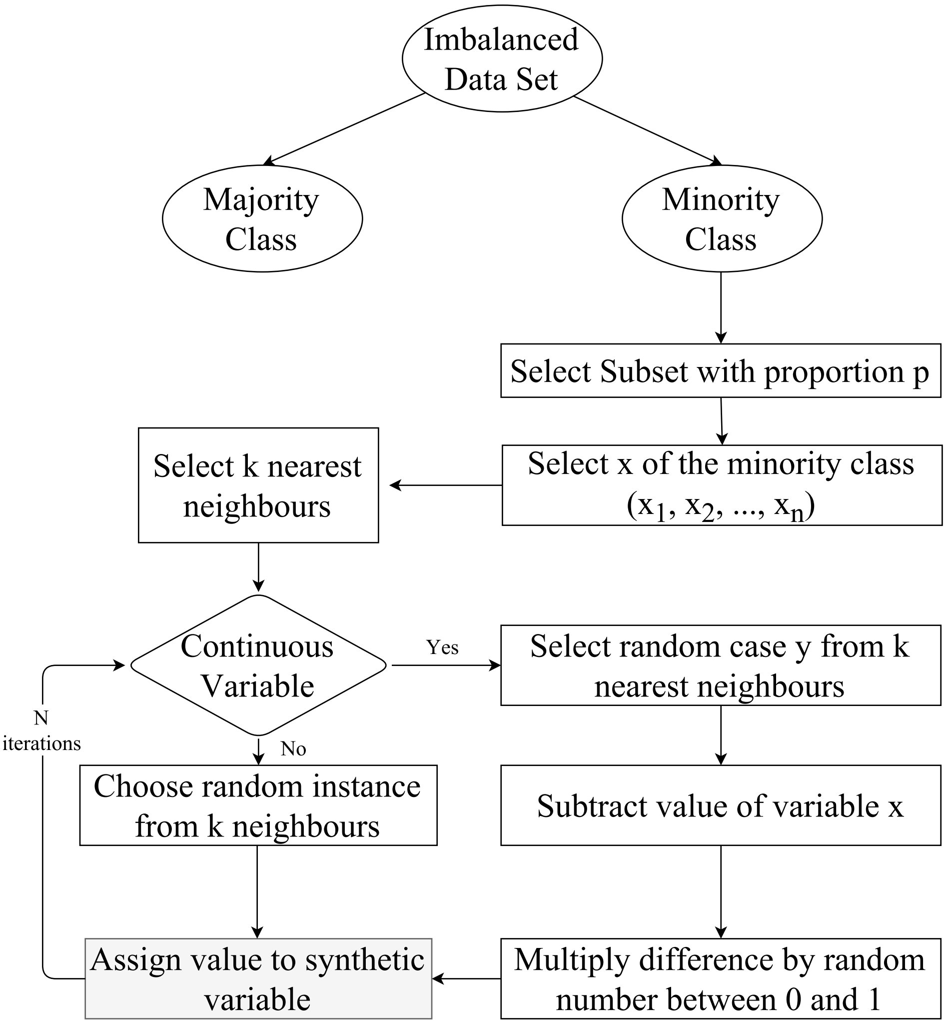


Figure 4.6: Flow chart for SMOTE.

Chapter 5

Healthcare Applications

5.1 Introduction

IoT-based healthcare systems can be utilized to a diverse array of fields, including care for pediatric and elderly patients, the supervision of chronic diseases, and the management of private health and fitness, among others.

Combined with sensor and communication technologies, the computational methodologies or algorithms enable systems' designers to develop enhanced healthcare applications. For instance, there are a lots of algorithms for activity recognition that vary depending on the underlying sensor technologies that are used to monitor activities, the alternative machine learning algorithms that are used to model the activities, and the complexity of the activities that are being modeled [70][71]. For Anomaly Detection, there are standard statistical methods to automatically detect and analyze anomalies including the box plot, the chart, and the CUSUM chart [72]. Anomaly detection is most accurate when it is based on behaviors that are frequent and predictable. For Decision Support, If-then rule can be applied to propose a set of suitable actions. Besides, various machine learning algorithms such as decision trees represent methodologies for learning healthcare and clinical knowledge [73].

There are abundant applications for healthcare that have been developed in academia and industry. In the scope of this research, we focus on three kinds of application to ensure the wellness [103]. With the development of sensor technology, the application for Health Monitoring also increase. For example, wearable personal ECG monitoring device (PEM) for early detection of cardiac events, which recognizes and records irregularities by creating various warning levels. Another instance is AMON which is in the form of a wristband and measures various physiological signals. Today, there are several commercially available health monitoring devices, such as HealthBuddy by Bosch, TeleStation by Philips HealthGuide by Intel, and Genesis by Honeywell. Fall detection applications rely on several technologies: wearable devices, ambient sensors, and cameras. Wearable fall detection systems measure position and movement utilizing sensors such as accelerometer and gyroscope and by measuring orientation and acceleration. Indoor localization study has earned a big deal of attention from both academia and commerce, with various systems being proposed using a diversity of technologies, for example, Ubiquitous Networking Robotics in Urban Settings (URUS) project, “Smart floor technology”, Open Mobile Alliance, etc. [74].

5.2 Algorithms

Nowadays, there is an exponential expansion in concern in healthcare area along with greater diversity in the algorithms used. However, parts of this system are carried out in low-resource processing algorithms suffering from data explosion. To overcome the problems, there is a need of a direct incorporate between ultralow power micro-controllers, advanced sensor technologies, and analytical algorithms. The complexity of typical operations applied in machine learning and signal processing, for example, matrix inversions and decompositions, is estimated $O(n^3)$. This complexity implies that if there are 100 samples, a simple algebraic operation needs around 1 million inner loop instructions. Note that accounting additional operator instructions such as floating point products are not included. [7]. However, the basic and light-weight algorithms have been implemented for basic processing components of online techniques, like noise filters, feature extraction, and peak detectors. Concurrent advances in Graphics Processing Units (GPUs) have made it able to handle high volumes of data in large repositories more proficiently. Besides, this technology also immensely support intensive transformations and operations to be executed in parallel, which benefit many algorithms.

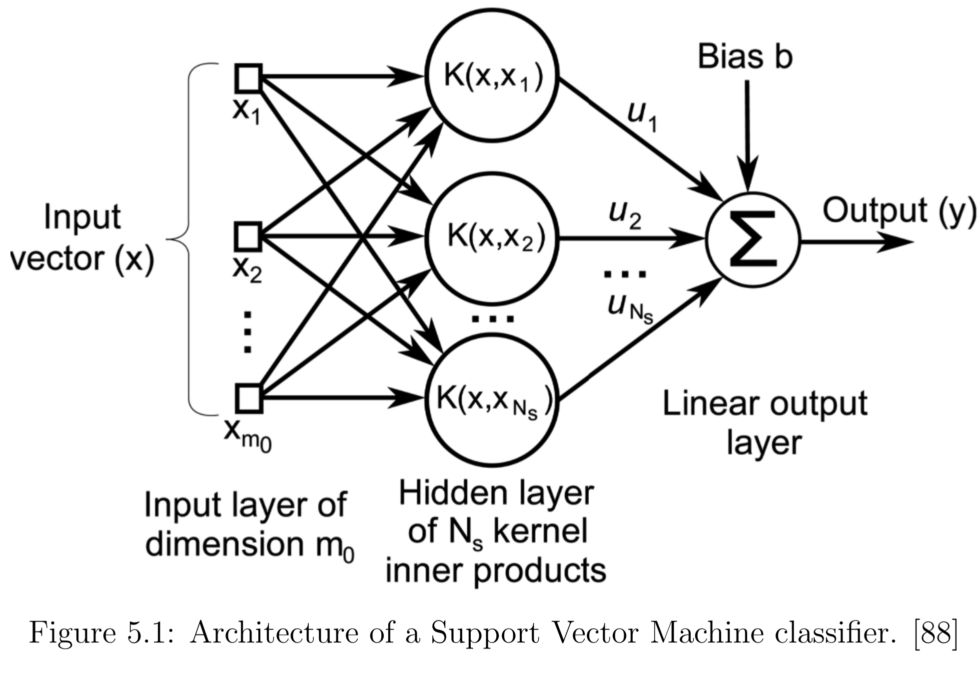


Figure 5.1: Architecture of a Support Vector Machine classifier. [88]

Figure 5.4 depicts this research trends in algorithms applied for healthcare decision support systems since 1990 to 2016 while Table 5.1 is the number of publication referring to several algorithms in healthcare applications from 2010 to 2016. Publication statistics are acquired from Google Scholar; the search query is defined as the subfield name of algorithms and at least one of medical or health appearing, e.g., “support vector machine” AND medical OR health. In this section, start from support vector machine algorithm, we identify plausible architectures in healthcare area.

Table 5.1: The number of publication referring to several algorithms in healthcare applications from 2010 to 2016 by Google Scholar.

	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**
Deep Autoencoder	2	3	8	15	21	53	146
Deep Boltzmann Machine	5	9	13	23	60	124	207
Deep Belief Network	36	99	118	160	304	561	1040
Recurrent Neural Network	925	1020	1210	1610	1510	1740	2960
Deep Neural Network	32	36	96	199	502	1210	3390
Convolutional Neural Networks	87	135	182	250	479	1230	3920
Semantic Web	5440	5730	6370	6850	6460	6290	5490
Bayesian Inference	3870	4400	5020	5800	6120	6660	7500
Random Forest	1150	1580	2260	3380	4380	6100	8870
Fuzzy Logic	7010	8080	9690	10700	11400	12900	13200
Genetic Algorithm	9360	10500	13200	14500	14700	16200	15900
Decision Tree	8820	10400	12200	13600	14500	15400	17000
Support Vector Machine	10900	13000	15200	16000	16300	16500	17000

Support Vector Machines - SVM [80], which was first proposed by Cortes and Vapnik in 1995, is a supervised machine learning algorithm applied for both classification or regression problems. Figure 5.1 defines the architecture of a Support Vector Machine classifier. Inner product kernels, K(* , *) denote the m_0-dimensional kernel inner product of the input vector with each of the N_s Support Vectors. At its very fundamental level, a support vector machine is a classification or prediction method and forms by the combination of statistical and machine learning technique. This powerful discriminative classifier formally classifies cases by identifying and optimizing a separating boundary called hyperplane. There are abundant advantages of SVM approach. This classifier can overcome the high dimensionality problem, one of the typical problems in machine learning when there is a large number of input variables corresponding to the number of available observations. Another advantage of SVM is that less training data needs to be inputted as compared to other algorithms while errors and complexity can be minimized. The approach has demonstrated high performance in solving classification problems from detecting diseases [75] to real time health monitoring [76] in clinical settings and bioinformatics. It original work [80] succeeded 26894 citations by Google Scholar with over 17000 related research in healthcare area in 2016.

Fuzzy logic (FL) is an approach based on the concept of “degree of truth” between 0.0 and 1.0 instead of the usual “true or false” (1 or 0) boolean logic on which the current computer is based. The method was first introduced by Prof. Lofti Zahed in 1965 [82].

It is a technique of decision making that simulates human decision making. This method allows the application enable the input to interpolate between the crisp set of classical logic, allowing a soft transition of the degree from "false" to "true". In the inference of clinical decision support systems, the method allows formalizing vagueness. Figure 5.2 presents the architecture of a fuzzy logic approach. In general, there are 2 types of fuzzy logic: type 1 - fuzzy logic and type 2 - fuzzy logic. In type 1 fuzzy set, researchers determine the degree of achieving the characteristics of the object. For instance, if there are 3 different green balls. The first is green by 67%; second is green 80%, Third is green 95%. However, in type 2 fuzzy set, researchers can just determine the range of the degree achieving the characteristics. For example, if there are 3 different green balls. The first is green by 50%-60%; second is green 70%-80%, third is green 95%-100%.

Both type 1 and type 2 fuzzy systems have been widely applied in applications of pervasive health from health monitoring [83] to diagnosis a disease [84].

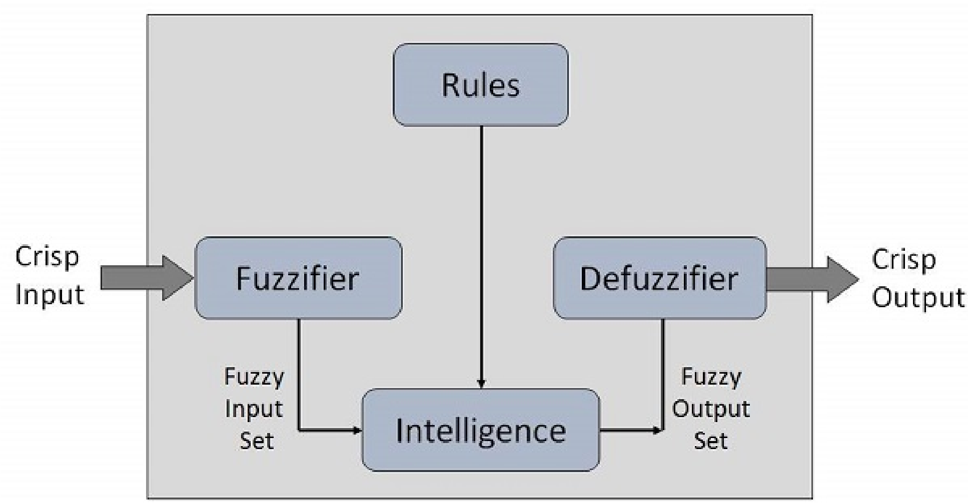
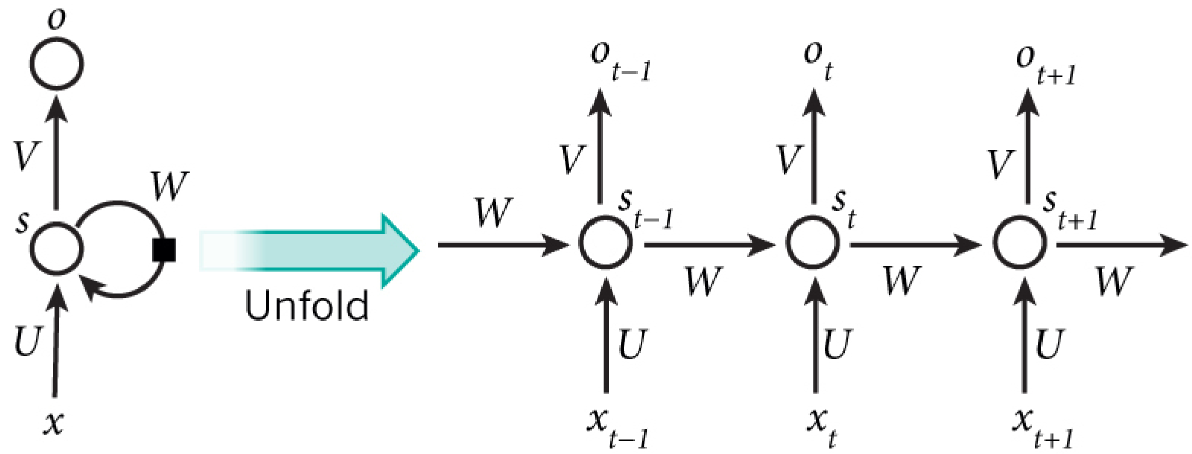


Figure 5.2: Architecture of a Fuzzy logic approach.

Recurrent Neural Networks or RNNs [85] are the families of the neural network which contains hidden layers for processing sequential data. Recurrent Neural Networks have a "memory" which captures information about what has been calculated so far and execute the same task for every element of a sequence, with the output being depended on the previous computations. This "memory" is meaningful in many applications where is a dependency between the input and output values such as the analysis of text, speech or DNA sequences. RNN can not only learn the local and long-term temporal dependencies in the data but can also accommodate input sequences of variable lengths. Figure 5.3 depicts a recurrent neural network and the unfolding in time of the computation involved in its forward computation. The RNN is usually fed with training sets that have strong inter-dependencies and a significant representation to keep information about what resulted in all the previous time steps. The outcome obtained by the network at time t-1 affects the choice at time t. This is the "memory" of RNN. Although RNN is a simple and powerful

Figure 5.3: A recurrent neural network and the unfolding in time of the computation involved in its forward computation.



algorithm, it suffers from the vanishing and exploding gradients problem. In other to combat these problems, the especially design of RNN was proposed by S. Chochreiter et al. in 1997, it is Long Short-Term Memory (LSTM) [86]. Currently, RNN and LSTM have earned the concern of researchers in the healthcare field. In 2016, nearly 2960 research applied RNN in healthcare system had been conducted. The methods are further detailed in Appendix B.

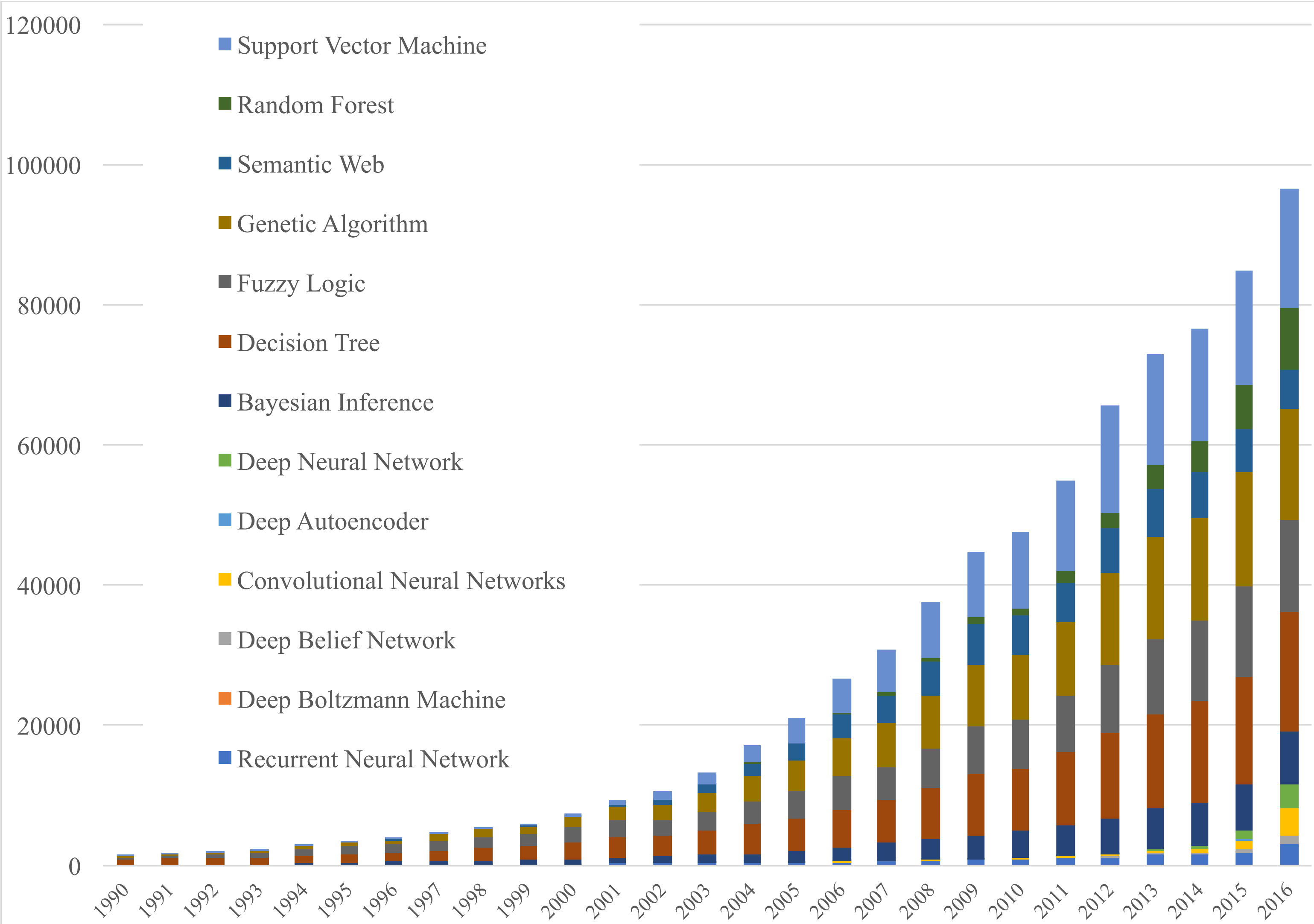


Figure 5.4: Evolution of academic publications concerning the methods in healthcare applications.

5.3 Applications

Among abundant IoT-based healthcare applications, we can classify them into two main categories: pervasive monitoring and medical informatics applications. Table 5.2 summarizes the feature IoT-based healthcare applications.

5.3.1 Pervasive Monitoring

The advance in sensor technology such as wearable, implantable, and ambient sensors allow health and wellbeing can be continuously monitored. For elderly who lives independently or suffered from chronic diseases, the application can be utilized to improve quality of care by anomaly detection or continuously monitor their biological parameters. Indoor location monitoring applications can track the location of an individual to ensure the safety inside the house. Indoor environmental information, another important aspect can be easy collect and monitor to ensure the comfort, wellness safety of the individual. Thanks to the applications in assistive devices and human activity recognition, the care of patients with disabilities and patients undergoing rehabilitation can also be improved. For patients in critical care, continuous monitoring of vital signs, such as blood pressure, respiration rate, and body temperature, is required for enhancing treatment results by carefully investigating the patient's health.

5.3.2 Medical Informatics - Prediction

Medical Informatics is the combination of information science, computer science, and healthcare. Reinforcement for the final goal of enhancing and advancing clinical decision support systems or assess medical data for quality insurance and accessibility of healthcare services, medical informatics mainly pay attention to the analysis of big, aggregated data in healthcare environments. Benefit from the using of advanced machine learning algorithms such as Recurrent Neural Network, Deep Neural Network or Deep Belief Network in big medical dataset, many applications can assist to predict a disease or even the next activity of human.

Applications	*Data Required*	*Technologies*	*Infirmity*	*References*
		Pervasive Monitoring		
Anomaly detection	Vital Signs EEG ECG PPG	Deep Belief Network	Diabetes Emergency Assistance	[89] - [92]
Biological parameters monitoring	Biological signals	Support Vector Machine	Heart rate monitoring, BP monitoring Body temperature monitoring Oxygen saturation monitoring	[93]
Indoor location monitoring	Accelerometer and Compass	Maximum Likelihood Estimation	Wheelchair management	[113]
Indoor environmental parameters monitoring	Environmental data	Temperature monitoring	Support Vector Machine	[94] - [95]
Human activity recognition	Accelerometer and Compass Video Embedded sensor	Deep Belief Network Convolutional Neural Network Support Vector Machine	Rehabilitation system	[87]
		Medical Informatics - Prediction		
Prediction of disease	Big medical dataset	Recurrent neural network Deep Neural Network Deep Belief Network	Cough detection Medication Management Wound analysis for advance diabetes patients	[77] - [79]
Human activity prediction	Embedded sensor Video	Support Vector Machine	Alzheimer's prediction	[96]

Table 5.2: Summary of IoT-based healthcare applications

Name	**License**	**Platform**	**Written in**	**Interface**	**Supported techniques**	Cloud Computing
Caffe	FreeBSD	Linux, Win, OSX, Android	C++	C++, Python, MATLAB	RNN, CNN, SVM	
Microsoft Cognitive Toolkit	MIT	Linux, Win	C++	Command line	RNN, CNN	
Deeplearning4jK	Apache 2.0	Linux, Win, OSX, Android	C++, Java	Java, Scala, Clojure	RNN, CNN, DBN	
Wolfram Math	Proprietary	Linux, Win, OSX, Cloud	C++	Java, C++	CNN, DBN	x
TensorFlow	Apache 2.0	Linux, OSX	C++, Python	Python	RNN, CNN, DBN, SVM	
Theano	BSD	Cross-platform	Python	Python	RNN, CNN, DBN	
Torch	BSD	Linux, Win, OSX, Android, iOS	C, Lua	Lua, LuaJIT, C	RNN, CNN, DBN	
Keras	MIT license	Linux, Win, OSX	Python	Python	RNN, CNN, DBN	
Neon	Apache 2.0	OSX, Linux	Python	Python	RNN, CNN, DBN	x

Table 5.3: Popular software packages that provide algorithms implementation.

Smartphone Applications	*Descriptions*
Google Fit	Track any activity such as walking, running, and cycle throughout the day.
Samsung Health	Record and analyze user daily activities and habits to help maintain successful diet and lead healthy lifestyle.
Empatica Embrace	Real-time Seizure Alerts to caregivers.
Huawei Health	Provide professional sports guidance for user sport
Empatica Embrace	Real-time Seizure Alerts to caregivers.
ASUS ZenFit	Track user fitness activities on your ZenWatch
Empatica Embrace	Real-time Seizure Alerts to caregivers.
Mi Fit	Track user exercises and analyze your sleep and activity data.
Healthy Children	Get trusty advice from the nation's leading child health experts.
Health Assistant	Keep track of a wide range of health parameters such as weight, body water and fat, waist size, height, blood pressure, body, temperature, lipids, blood sugar (glucose), smoking, physical activity, taken medicines.
Calorie Counter	Keep track of food consumed by the user as well as the user's weight and measurements, etc.
Heart Rate Monitor	Real time monitor heart rate.
ECG Self-Monitoring	An application developed for ECG registration and automatic ECG interpretation from the convenience of user's home
Cardio Mobile	An application for mini wireless mobile ECG monitoring devices
Body Temperature	Measure the fever by using user finger.
Eye Care Plus	Test and monitor vision.
iOximeter	Connect to oximeter machine by BLE and then calculate the pulse rate and SpO_2
SkinVision	Keep track of the user's skin health and enables the early discovery of any skin disorder.
Fall Detection	An application is designed to detect and generate alerts when a device fall occurs.
Hearing Test	Allows for the self-assessment of hearing.

Table 5.4: Smartphone application for healthcare.

Chapter 6

Open Challenges and Issues

6.1 Introduction

Despite of numerous accomplishments in smart healthcare research field, it remains various challenges and limitations in the development, implementation, and evaluation of this smart healthcare system at the present. In this chapter, based on the general architecture of smart healthcare systems, we recognize three main vital issues (1) the problem of collecting data by sensors (2) the low accuracy of indoor location monitoring systems (3) the design of smart healthcare system. These issues are necessarily to be considered during design, develop and implement processes of the smart healthcare system. Figure 6.1 depicts the vital issues based on the general smart health care system architecture.

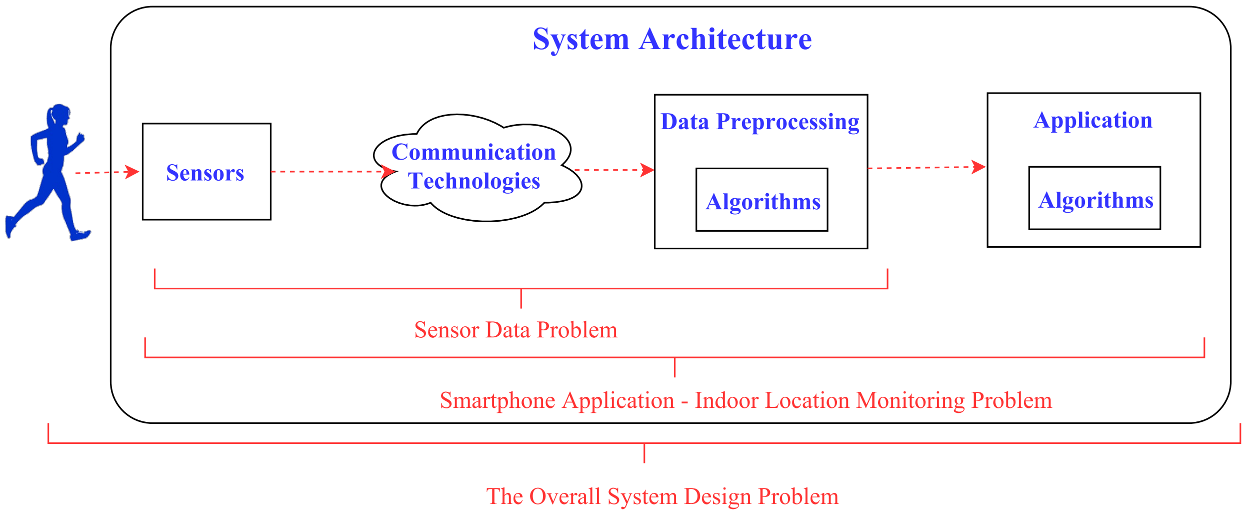


Figure 6.1: The vital issues based on the general smart healthcare system architecture.

6.2 Sensor Data Problem

Missing sensor data is the primary key issue in healthcare sensor data prolems. In spite of the fact that advanced technologies have been developed through time, there are still a distance between values from reading processes and actual values of measuring parameters. Advanced technologies has been used to perform well in some controlled situation; however, they failed to get high accuracy levels in real-life scenarios[121].

There are several reasons behind this uncertainty and missing including (1) internal inaccuracies of sensor operations, (2) the effect of human activities in sensor deployment locations and (3) the impact of biological fouling upon the sensors over time [98].

- There are abundant sensor data manufacturers with their technologies and algorithms in sensor industry. However, each applied algorithm have their strength and weakness. As a result, this weekness of the adopted algorithm involves internal inaccuracies of sensor operations. Consequently, the collected data fail to get accuracy or even not defined (or missing data).

- Sensors send the collected data to servers and other nodes. Datastreams is the continuous flow of data measuring from a sensor into the network. Due to many reasons such as power outage at the sensor's node, a high bit error rate of the wireless radio transmissions, this sensor datastream can be lost or corrupted. The human's activity in sensor deployment locations can be consider as random occurrences of local interferences which is the reason of missing sensor data.

- When the sensor is deployed in-situ, due to the environmental aspects, it can be drifted by time. Once its critical surface is covered in biofouling, sensor's ability to collect accurate readings is compromised. As a consequence, these drifted sensors render the data collected inaccurate or even do not render data in some cases. The accuracy of sensors, especially the conductivity sensor, is rapidly degraded. Meanwhile, the sampling rate of the sensor is also compromised, which leads to the problem of missing sensor data.

The research of Acuna Edgar [99] has presented problems on missing datas that sophisticated handling methods are essentially required to get a better accuracy if there is more than 5% of missing samples. Especially in smart healthcare area, the accuracy of application is a critical issue. Fail to get accuracy in the application can lead to the serious consequence to the patient. Therefore, the quality of monitoring data becomes a major concern in sensors deployments of health monitoring systems along with other factors such as power consumption, context awareness, security and patient comfortable. Many techniques were applied to collect high quality data, however in some studies, data was collected just in short periods of time or the quality of data was substandard.

Along with missing sensor data challenge, there are several other issues such as lack of data standards, data privacy, data formatting, data normalization and data synchronization.

Although several standardization efforts have been made through time, however, few medical sensor manufacturers adhere to these standards. Many vendors that manufacture a wide range of healthcare sensors while new manufacturers start their business in this promising technologies race. The sensor data manufacturers tend to design proprietary data models and protocols to externalize sensed signals. Due to this lack of adoption standards, the format of collected data is different from each system. For example, there are at least five major venders of sensor electronic health record system in the USA. However, they all have different standards and different ways of storing data, which prevents the systems from sharing data. The task of mining data from multiple healthcare systems faces severe challenges. As a result, there must be a design of custom solution specific to each sensor data, which is a waste of time and increasing the system cost.

Moreover, to cope with differences in the sensing process, a semantic normalization is often required. For example, in some systems, daily reported body temperature capture may represent to a daily average body temperature, while in other systems it may correspond a body temperature average captured every night before the patient goes to bed. The result yields from comparing these values in a healthcare application may be incorrect, particularly if they are not semantically analyzed. Consequently, immediate efforts are required to address this standardization problem.

Data synchronization is another difficulty of sensor data. Based on the internal clock, sensors report the data to servers. Unfortunately, these sensor's clocks are often not synchronized; it is another challenge in the task of analyzing data across sensors. Also, due to power consumption, the sampling rate may be different from each type of sensors. Therefore, assumptions and alignment policies need to be thoroughly sketched.

Another key challenge involes data privacy and the security of captured health data from diverse sensors and device from illegal access. Healthcare data is usually more valuable to thieves than financial data. Anything that is connected to the Internet can become a potential portal for hackers to get in and get data out. In this regard, the regulations should be defined to share data with authorized users, organizations and application. In regard to the technical aspect, there is a need of optimal methodologies for collaboration between security, detection and response assistance to prevent different attacks, threat, and vulnerabilities.

In conclusion, the challenges associated with data collected by sensor specific for smart healthcare system are classified as follow:

- **Missing Sensor Data.**
- Lack of data standards.
- Data privacy and security.
- Data formatting.
- Data normalization.
- Data synchronization.

6.3 Smartphone Application - Indoor Location Monitoring Problem

Recent years have seen an expanded adoption of smartphones in smart healthcare area, which highlights the increase of smartphones as a driver of the IoT. However, battery life, small screen size, potentially erroneous data input, viruses, potentially inefficient patient-physician interactions, loss or theft, data privacy, and security have risen as the emergent issues of smartphone healthcare applications.

Due to the memory size and operating system, smartphone applications cannot deal with massive algorithms. Therefore, server-based service for smartphone application is a requirement. On smartphones using a stylus, data input is much slower and probably inaccurate. For security reason, antivirus software must be used to protect smartphones from viruses and spyware. However, this antivirus application is not only made the devices slow but also waste the battery power. Another problem is the risk of electromagnetic interference with healthcare devices when using smartphone application.

In spite of abundant of smartphone-based healthcare applications accessible in online application stores (e.g., Google Play store, Apple's App Store, etc.), most of them have not been discussed in the medical literature. There is a need for medical literature review specific to these applications.

The reliability and accuracy in the real-world setting of almost recent technologies such as activity recognition and indoor location monitoring are still required to be improved. Furthermore, some simplifying theories should be relaxed, such as the theories concerning individual homes and availability of labeled data. Additionally, a need for standard benchmark datasets is another issue.

Indoor location is another important concept of smart health monitoring system. The collected location data can be used to monitor activities of elderly at home for detecting different health conditions. Many studies have achieved their success in estimating steps via acceleration parameters to navigate locations inside the house. For example, the average distance estimation error for indoor 16-step straight-line walking experiments was 5.5% with a maximum error of 2.05 m with the combination of using smartphone built-in sensors and the map of the floor [100]. Some studies even used two smartphones at one time for tracking locations, which is not only drain a smartphone's battery but also impractical [101]. Therefore, there is a need of a lightweight and accurate localization algorithm. The algorithm has to satisfy three main conditions: being suitable for execution in a smartphone, avoiding the adoption of detailed wireless signal maps and requiring inexcessive hardware installations.

The current problems of application and indoor location monitoring system can be summarized as follow:

- Low battery life.
- Small screen size.
- Potentially erroneous data input.

- Data privacy and security problems.
- Inaccuracy of indoor location monitoring system.

6.4 The Overall System Design Problem

Home-Based Health Care System is an integration idea from ubiquitous computing and ambient intelligence applied to the health, wellness and well-being concepts, including the amount of data collected by sensors and actuators to monitor, predict and improve the patient's physical and mental condition. Therefore, the task of improving the quality of collected data and accuracy of indoor location monitoring system and Smart Health Platform are the parts of Smart Health System. There are three main problems that need to be considered when designing an IoT Healthcare System.

- A lack of standardization.
- Smart Health Platform.
- Human in the loop (doctor in the loop) concept.

The first problem of Smart Healthcare System design is a lack of standardization. Although in smart healthcare context, efforts have been put in standardization bodies, for example, Health Level 7 and the Continua Health Alliance handle data modeling problems while several IEEE standard protocols address device interoperability issues. A diverse range of products and devices which support smart health system have been manufactured by various vendors. Besides, more and more new vendors proceed to enter this promising technological competition. Nevertheless, these products and devices are not followed any standard rules and regulations for compatible interfaces and protocols. Regardless of the user's location, smart healthcare system requires standardized rules for ambulatory environments that support point-of-care while assuring the user's privacy. To overcome this interoperability problem, there is a need of cooperation between dedicated groups such as the Information Technology and Innovation Foundation (ITIF), The European Telecommunications Standards Institute (ETSI) and Internet Protocol for Smart Objects Alliance (IPSO) to standardize IoT-base healthcare technologies. In this standardization, the wide range of aspects should be considered, for instance, communications layers, sample rate, data precision, protocol stacks, consisting of physical and media access control layers, device interfaces, data aggregation interface, device descriptions, and gateway interfaces.

Although many efforts have been made, there is a need for standard configuration of the Smart Health Platform [102]. Therefore, the second problem is the lack of standard Smart Health IoT Platform. In the Internet of Things field, IoT platform is basically software products, which provide comprehensive sets of application-independent functionalities that can be used to construct IoT applications. It is the central hub of domains that control the connectivity of connected IoT devices via the telecommunication network, provide hosting space and processing power for application and services. Smart Health Platform is the sub-platform of IoT platform that especially applies in the healthcare domain. However, the architecture of IoT-based healthcare device is usually more complicated than that of common IoT hardware and demands a real-time operating system with more stringent requirements. Therefore, there is a need for customized computing platform with run-time libraries. Various approaches can be applied to build an IoT Healthcare platform.

- The ideal concept to create a suitable platform is the use of Service-oriented approach (SOA) which is a procedure applied to build an architecture based on the utilization of services. By applying SOA, the services can be utilized by using various application package interfaces (APIs).

- Besides, as a part of the IoT Healthcare platform, appropriate framework and libraries, especially a particular class of disease-oriented library, should be created to support software developers, designers and researchers to take advantage of given documents, codes, classes, message templates also other useful data.

There is a lack of Home-Based Health Care System which is an integrated system that can not only ensure the wellness of the patient but also provide the APIs for the third-party application.

The third problem of system design is an integration of concept human in the loop with IoT Healthcare System. Large amounts of data collected by sensors are hard to manage manually; therefore they seem like useless to human. Nevertheless, they are essential for machine learning algorithms, as the more training data are available the better results are achieved. In the aspect of system design, a perfect match of both together is to include the concepts of human in the loop (HITL). HITL is a model that demands user interaction as well as enables the human to modify the result of an event or process. Therefore, it is useful for system training. Especially in the IoT Healthcare System, the cooperation between an authorized body, scientists, and doctors, association of medical experts is an essential requirement to ensure acceptable quality. The system design should allow the application to regularly update based on the due consideration of recent advances in medical science. However, there is few related work on the human in the loop approach in IoT Healthcare System. Doctors and nurses usually avoid learning and using new technologies while scientists and IT developers do not have comprehensive knowledge of the problem at hand, principally medicine and the human body. For instance, the extraction of useful features from sensors data requires a deep understanding of the problem and need to be driven by domain experts. Figure 6.2 is an example of the concept of human in the loop (doctor in the loop) for activity recognization. In this instance, a doctor-in-the-loop creates and modifies rules on demand to train the algorithm which developed by IT researchers.

Therefore, to ensure the quality of the IoT Healthcare System, the system design should support human in the loop concept.

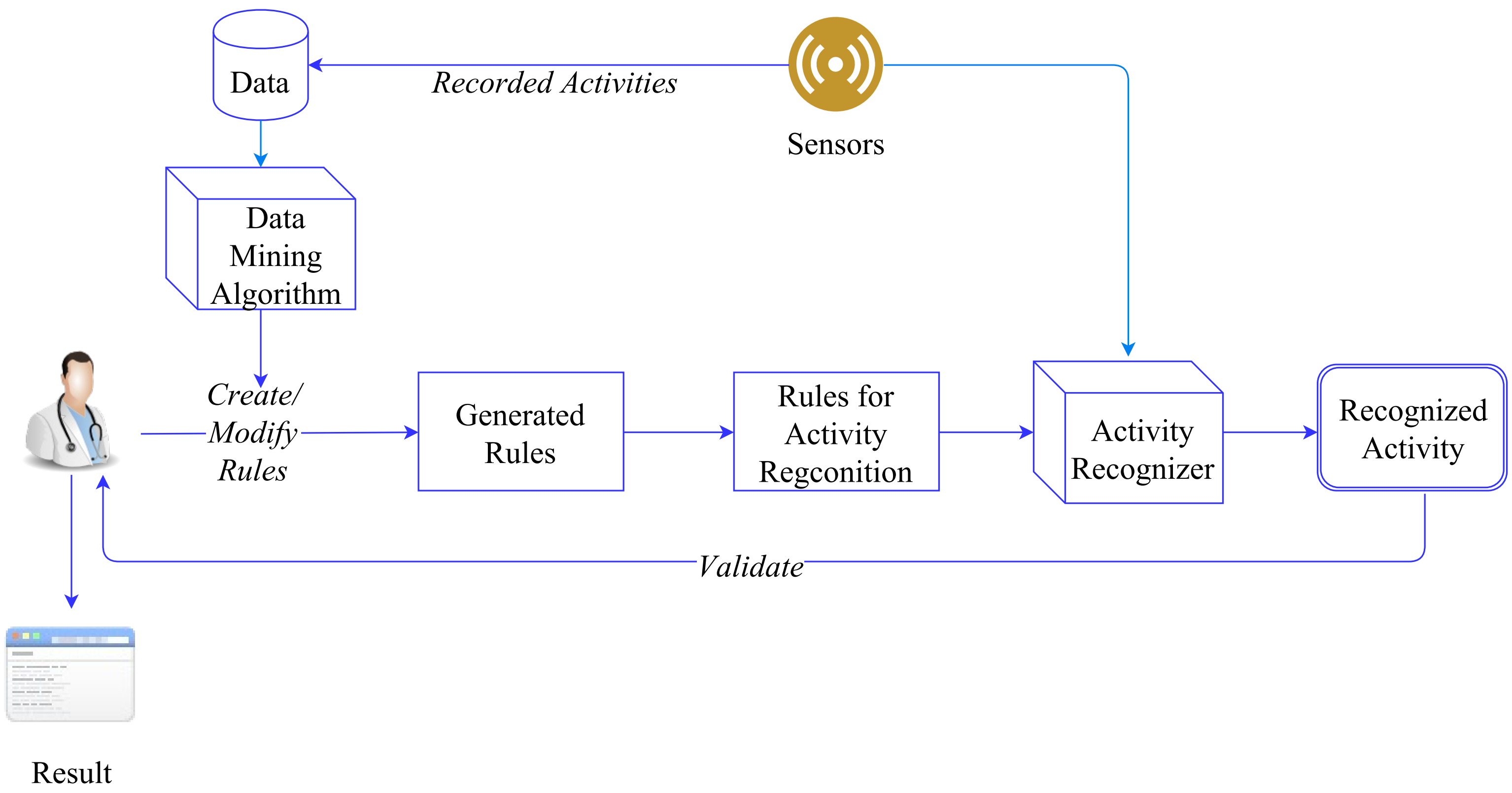


Figure 6.2: Human-in-the-loop (Doctor-in-the-loop) Concept [3]

Chapter 7

Implementation

7.1 Introduction

As a second study, based on the general system architecture as shown in Figure 1.2 in the first chapter of this research report, we conduct the implementation of the smart healthcare system. Figure 7.1 illustrate the procedure of the implemented health monitoring system which can be summarized as follow.

1. E4 Wristband wearable device is used to capture the various physiological signals such as BVP, IBI, HR, temperature, and accelerometer figures of the patient.

2. Through BLE technology, the collected data is transferred to the servers by Android smartphone gateways. There are two types of smartphone gateways in this implementation. The first is the S-Healthcare gateway application developed by our research team. The second type of smartphone gateway is Empatica RT application, provided by Empatica, who is the manufacturer of E4 Wristband device.

3. Relatively with the two types of smartphone gateways, there are two types of database servers employed in this implementation. The transferred data by S-Healthcare gateway was saved in Google Firebase Database Server while data collected by Empatica Application was stored in Empatica Connect - a Database Server provided by Empatica.

4. Sevice server accessed data from both Google Firebase Database Server and Empatica Connect Database Server, then performed pre-processing steps by applying different pre-processing algorithms.

5. The pre-processed data was analyzed in the final step of the implementation. The visualization, which is illustrated by the system, can be approached by doctors, researchers and family members of the patient. The result provided by the implemented system is significant for the future research.

The detail of each the components in our implementation process are explained in the following sections.

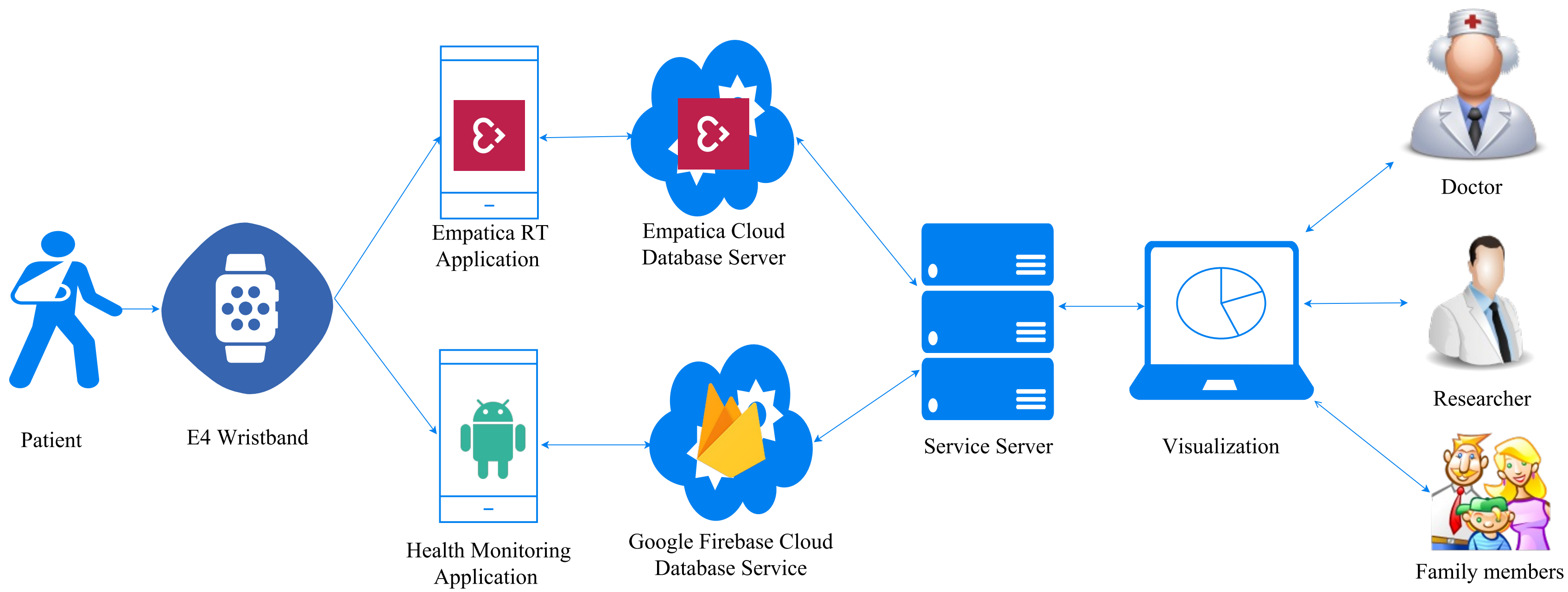


Figure 7.1: The Structure of the Implemented Health Monitoring System

7.2 E4 Wristband Wearable Device

Introduced by Empatica, co-founded by MIT professor and wearables pioneer Rosalind Picard, the E4 wristband [104] is a wearable wireless multi-sensor device designed for comfortable, continuous, real-time computerised biofeedback and data acquisition. There are four embedded sensors in E4 wristband:

- Photoplethysmography (PPG) sensor to read blood volume pulse (BVP). Heart rate, heart rate variability, and other cardiovascular features can be determined from this blood volume pulse.
- Electrodermal activity (EDA) sensor applied to provide sympathetic nervous system arousal. From this EDA, characteristics related to stress, engagement, and excitement can be derived.
- 3-axis accelerometer sensor is utilized to capture motion-based activity.
- Infrared thermopile sensor applied to measure skin temperature.

Figure 7.2: The E4 wristband overview [104]

The E4 wristband can operate either in streaming mode for real-time data processing utilizing a BLE interface or in in-memory recording mode employing its internal memory. In a comfortable and compact wearable fashion, the E4 device creates an ability to conduct healthcare study outside of the laboratory conditions by acquiring real-time data for mobile recording. The data quality of E4 wristband was carefully validated by Cameron McCarthy at [105]. The technical details of Empatica E4 wristband sensors are specified in Table 7.1.

The E4 wristband is implemented with a PPG sensor, scientifically as BVP signal, which illuminates the human skin and captures the reflected light. The sampling rate of PPG sensor is 64Hz. As its application, PPG signal is the input to the algorithm measuring heat beats and thereby calculating IBI signals for getting outputs. The equation 7.1 represents the processing of PPG in the E4 wristband.

$$GREEN, RED - - > [Algorithm1] - - > BVP - - > [Algorithm2] - - > IBI \tag{7.1}$$

In this process, instead of traditional Red/IR combination, E4 wristband utilizes the coupling of green and red lights to increase the detection efficiency of the pulse wave. The sensor is made of 4 LEDs operation wavelengths (2 green leds, 2 red leds), and 2 units of photodiodes ($14mm^2$ of sensitive area). The created lights by the Green and Red LEDs are oriented toward the skin and absorbed by the blood in different ways. Then the Light receiver is used to measure the portion of the reflected light. Time interval between two diastolic points, which are local minima of the BVP signal, is applied to calculate IBI. This procedure and sample BVP output signal are illustrated in the figure 7.3.

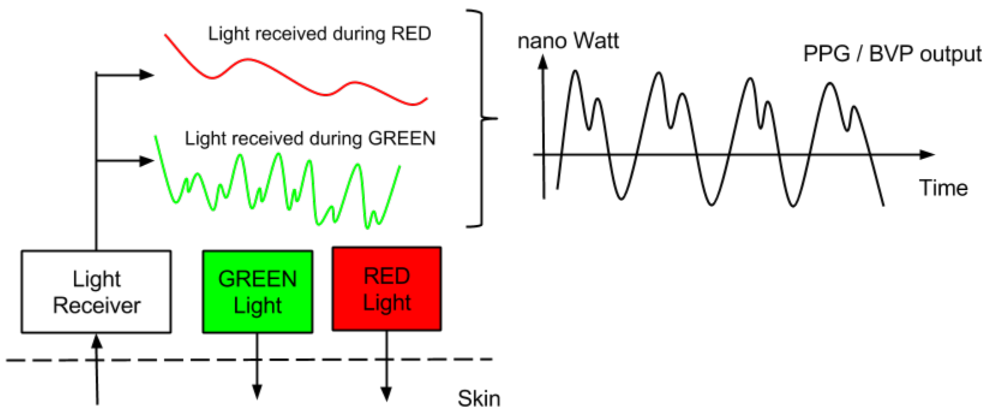


Figure 7.3: The the BVP signal utilization [104]

Specifications	*PPG*	*EDA*	*Temperature*	*Accelerometer*
Sampling Frequency	*64 Hz*	*4Hz*	*4 Hz*	*32 Hz*
Unit	*nanoWatt (nW)*	*microSiemen* (μS)	*Celsius* ($^{\circ}C$)	*G-forces (g)*
Resolution	*0,9 nW / Digit*	*1 digit to 900 pico Siemens*	$0.02^{\circ}C$	*8 bit of the selected range*
Range	-	*0,01– 100* μS	*-40 to 115*$^{\circ}C$	*-2 to +2g*
Consist of	*- Photodiodes (2 units - 14* mm^2 *sensitive area)* *-Green LED(2 Leds)* *- Red LED (2 Leds)*	*Electrodes - 2 units*	*Infrared thermopile*	*3-axes acceleration sensor #3, X,Y and Z*
Placement requirement	*None*	*Bottom wrist*	*None*	*None*
Available from Connect as CSV file of through the API	*Yes*	*Yes*	*Yes*	*Yes*
Sensor Replaceable	*No*	*Yes*	*No*	*No*

Table 7.1: Technical aspects of Empatica E4 wristband sensors

7.3 Android Gateway Application

For smartphone gateway development, Empatica provides developers SDKs for both Android and iOS platforms. In this research, due to the universal diversity of equipment available on the market, ease of deployment and widespread adoption of the application, we decided to apply the Android platform. The primary function of the Android gateway is to collect real-time physiological data from E4 Wristband and transfer to the specific database server. Besides, the smartphone application allows the patient to monitor real-time data through chart visualization. There are two types of smartphone gateways were implemented in this research: (1) S-Healthcare gateway developed by our research team and (2) Empatica RT provided by Empatica. The two smartphone gateways apply the same SDKs to collect the real-time data. Therefore, there is few difference between the efficiencies of the S-Healthcare gateway and Empatica RT. However, the captured data from Empatica RT application is uploaded to the server only after finishing the streaming session. This disadvantage prevents the user from remote real-time monitoring healthcare data. By contradiction, the collected data from S-Healthcare gateway is immediately streaming to the server in real-time, which enables to build a remote real-time health monitoring system. Furthermore, S-Healthcare gateway allows an ability to modify the data structure and custom functions, etc.

7.3.1 S-Healthcare gateway

In this project, we used Android Studio, the official Integrated Development Environment (IDE) for Android app development, based on IntelliJ IDEA, as an IDE. Table 7.2 details the applied Java method to retrieve the real-time data-streams from S-Healthcare gateway application. These methods were provided by Empatica developers SDKs.

The technique used for visualizing the real-time data is MPAndroidChart. This technique is powerful and usable easily. Besides, it supports Java libraries for Android Chart with real-time processing. MPAndroidChart library was proposed by Philipp Jahoda [106]. In order to use the library in the Android project, it is recommended to put the following code in root-level *build.gradle* file of Gradle Dependency.

```
allprojects {
        repositories {
                maven { url "https://jitpack.io" }
                }
}
```

Then the following code should be added to app level *build.gradle* file.

```
dependencies {
        implementation 'com.github.PhilJay:MPAndroidChart:v3.0.3'
}
```

To connect and save real-time data to the server, we integrated Google Firebase libraries [107] into our project. Based on its advance, Google Firebase provided API to create,

Return Type	*Method*	*Description*
void	*didReceiveAcceleration (int x, int y, int z, double timestamp)*	*Invoked when a new acceleration value is available*
void	*didReceiveBatteryLevel (float level, double timestamp)*	*Invoked when a new battery level value is available*
void	*didReceiveBVP (float bvp, double timestamp)*	*Invoked when a new BVP value is available*
void	*didReceiveGSR (float gsr, double timestamp)*	*Invoked when a new GSR value is available*
void	*didReceiveIBI (float ibi, double timestamp)*	*Invoked when a new IBI value is available*
void	*didReceiveTemperature (float temp, double timestamp)*	*Invoked when a new temperature value is available*

Table 7.2: The applied Java method to retrieve the real-time data-streams from S-Healthcare gateway application

read, write, and delete data in real time with just a few lines of code. The format of stored data is JSON which is accessible from all the platforms from Android, iOS and web applications. The further details of Google Firebase are discussed in the subsection 7.4.

In order to apply Google Firebase in the Android project, after successfully connecting the application to Firebase by Android Studio Firebase Assistant, the following code should be added in root-level *build.gradle* file of Gradle Dependency.

```
buildscript {
        dependencies {
                classpath ‘com.google.gms:google-services:3.1.1’
        }
}

allprojects {
        repositories {
                maven {
                        url ‘‘https://maven.google.com”
                }
        }
}
```

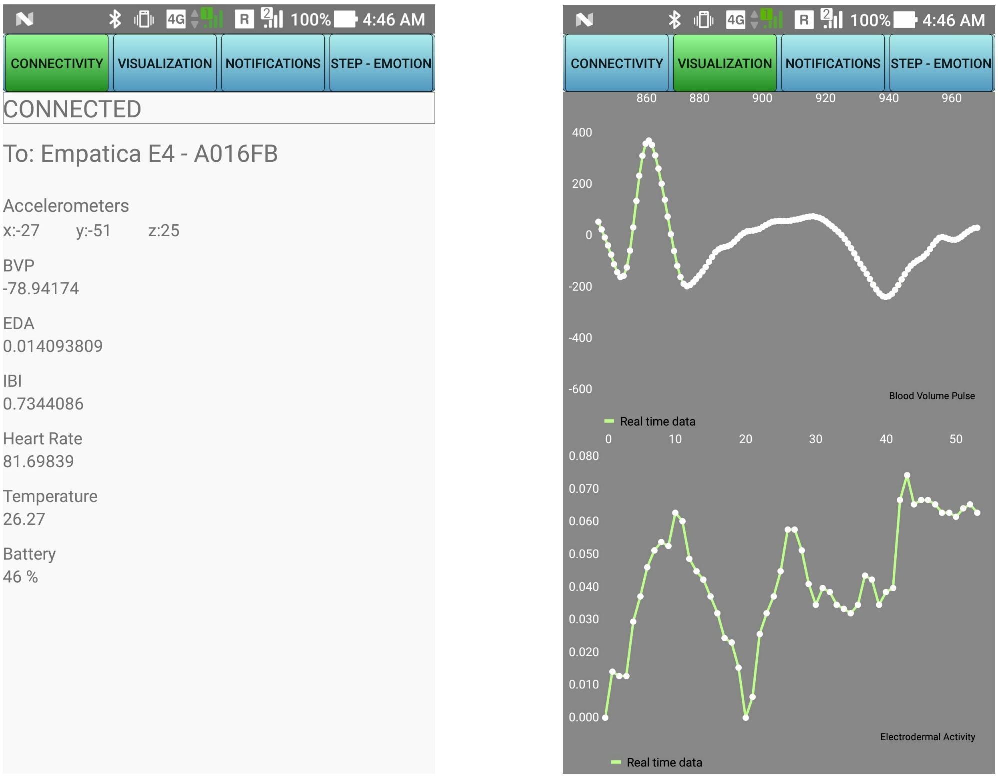
Then, at app level *build.gradle* file, add the following code.

```
apply plugin: 'com.android.application'

        android {
                // ...
        }

        dependencies {
                // ...
                compile 'com.google.firebase:firebase-core:11.8.0'
        }
        // ADD THIS AT THE BOTTOM
        apply plugin: 'com.google.gms.google-services'
}
```

Figure 7.4 illustres the Main Screen and Visualization Screen of the S-Healthcare gateway.



(a) (b)

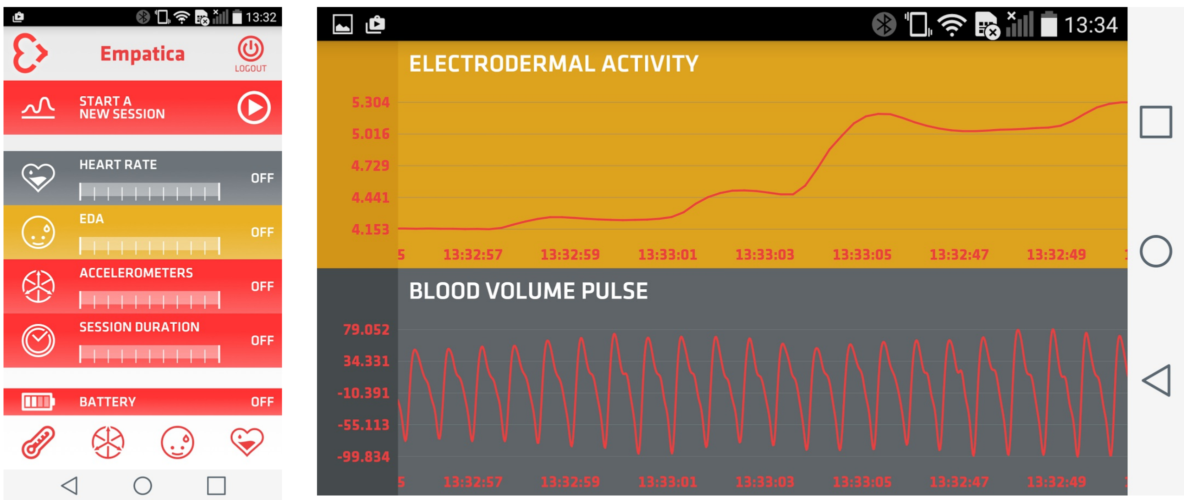
Figure 7.4: S-Healthcare gateway for collecting transferring data with (a) Main Screen (b) Visualization Screen.

7.3.2 Empatica RT

Empatica RT or can be called Empatica Realtime App is a research application developed by Empatica. Empatica RT enables the user to visualize real-time physiological signal data collected from Empatica E4 devices. The streamed data is saved and automatically backed up to Empatica Server only after finishing the streaming session. This is one of the limitations of Empatica RT which prevents the user from remote real-time monitoring healthcare data. Functions of Empatica RT can be summarized as follow:

- Provide the real-time compact view details about signals taken from the E4 device.
- Able to view the recorded data in detail by zooming and panning.
- Monitor the session duration details and the battery level of the device.
- Automatically upload the collected data to the Empatica Server after completing each streaming session.

Figure 7.5 represent the Main Screen and Visualization Screen of the Empatica RT application.



(a) (b)

Figure 7.5: Empatica RT application (a) Main Screen (b) Visualization Screen

7.4 Database Server

7.4.1 Google Firebase Database Server

Google Firebase Database Server has the benefit of uploading data to an online database and retrieves data from the cloud storage. This advantage plays an important role in protecting data from being lost unexpectedly and help to save researchers time and effort. Therefore, we chose it to apply for our implementation. Moreover, Google Firebase provides APIs for multiple platforms ranging from Android, iOS to the web application to operate with the server. When Google Firebase APIs is implemented into an application, with just a few simple line of code, this application can benefit from various firebase features. The Firebase provided features are described in 7.3.

Features	*Description*
Analytics	*Feature allows researchers to clearly understand how users are using the application.*
Authentication	*Only the authorised users can connect to the application. Authentication by (1) Email and password, (2) Federated identity provider integration including Google, Github, Twitter, Facebook (3) Phone number (4) The custom authentication by developers and even (5) Anonymous authentication are provided.*
Messaging	*The feature allows free cost messages deliver to different platforms.*
Real-time Database:	*The non-SQL cloud-based database supports storing and fetching data without time delay. Data is saved as JSON format and synchronized in real-time to all associated clients.*
Storage	*The feature allows the user to store and retrieve contents such as images, videos, audio, documents directly from client SDK. The processing of uploading and downloading the materials are invisible to the user. Only authorized user can access to the stored contents.*
Hosting	*The feature enables the user to deploy Firebase as hosting for website development.*
Crash reporting	*The feature enables to send reports of errors in the application after its release.*

Table 7.3: Features provided by Google Firebase Database Server

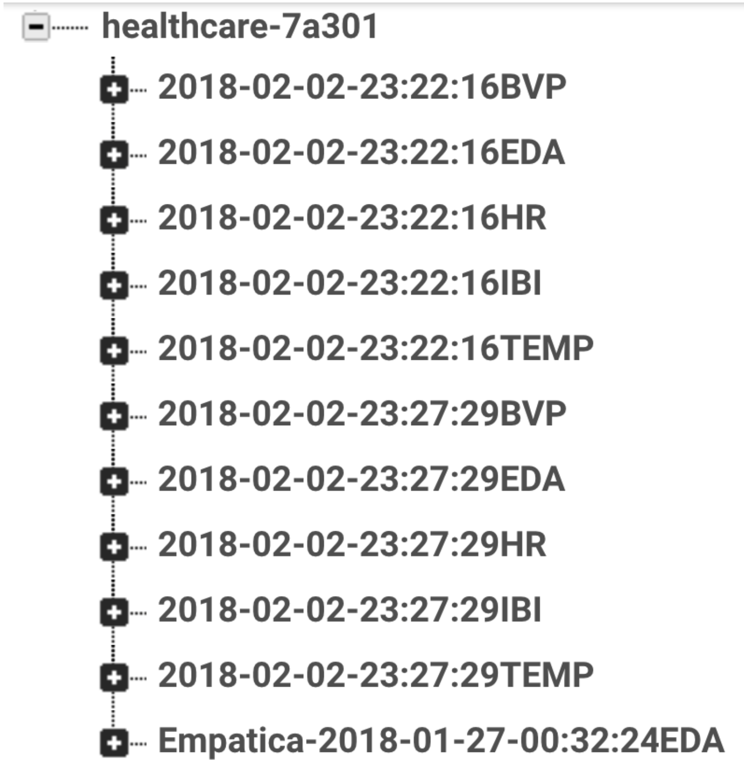


Figure 7.6: An example of JSON file name stored in Google Firebase Database Server

In order to support multiple platform applications access to the stored database, Google Firebase Server stores data as JSON. JSON is well-known as a lightweight, easy to parse and generate, data-interchange format. In this implementation, for each streaming session, the measured data including BVP, EDA, IBI, TEMP, HR are stored in the different JSON files. The JSON file name format is constructed as follow:

yyyy-mm-dd-HH:MM:ss + data_name

where 'yyyy-mm-dd-HH:MM:ss' is the initial time of the streaming session; data name is including BVP, EDA, IBI, TEMP, HR. For example, if streaming session starts on 23:22:16 February 2nd, 2018; the collected data is BVP; the JSON file name **2018-02-02-23:22:16BVP** will be saved on server for this streaming session.

To distinguish between data was collected by S-Healthcare gateway and Empatica RT, the JSON file name of data collected by Empatica RT is added "Empatica-" at the beginning. For example, **Empatica-2018-02-02-23:22:16BVP**. Figure 7.6 is an example of JSON file name stored in Google Firebase Database Server, where *healthcare-7a301* is the root project database automatically created by Firebase.

The sample format of JSON file can be seen in this example where the first line, for instance, "-L4LtTOLnianCVxfreVF" is the code initialed by Firebase; *data1* is the timestamp, and *data2* is the actual collected value.


```
{
        ``-L4LtTOLnianCVxfreVF" : {
        ``data1" : ``1.5175813594558294E9",
        ``data2" : ``0.67190576"
        },
        ``-L4LtTYzBKEeAvQ5OpFI" : {
        ``data1" : ``1.5175813601433609E9",
        ``data2" : ``0.6875315"
        },
        ``-L4LtTiYmEwsO9er-g_C" : {
        ``data1" : ``1.5175813608152666E9",
        ``data2" : ``0.67190576"
        },
        ``-L4LtTtEm_WYvnNfZf4G" : {
        ``data1" : ``1.5175813614871724E9",
        ``data2" : ``0.67190576"
        },
        ``-L4LtU3C1xog6sMIRIL3" : {
        ``data1" : ``1.5175813621903296E9",
        ``data2" : ``0.7031572"
        },
        ``-L4LtUBXxsTB_hVJZxL6" : {
        ``data1" : ``1.5175813628309839E9",
        ``data2" : ``0.6406543"
        },
        ``-L4LtULz4SPEPwCnxKF9" : {
        ``data1`` : ``1.5175813634560125E9",
        ``data2" : ``0.6250286"
        }
}
```


7.4.2 Empatica Connect

Empatica Connect is a secure cloud-based database server to analyze and achieve the data obtained by the E4 device. The interface consists of list and calendar views to navigate to an assigned session, a dashboard for data visualization. Besides, to support 3rd party applications accessing raw data, the Empatica Connect provides download links. Figure 7.7 depicts the interface of Empatica Connect.

The Empatica Connect significant features are summarized as follow:

- Store the archived from E4 devices securely.
- Securely access data from an internet browser.
- Export session packages to data in the user prefers analytic software suite (for example, Matlab, Weka, etc.)
- Provide the overview information for all stored sessions including duration, device serial number, and session start date-time
- Provide a searching method by session date time, equipment.

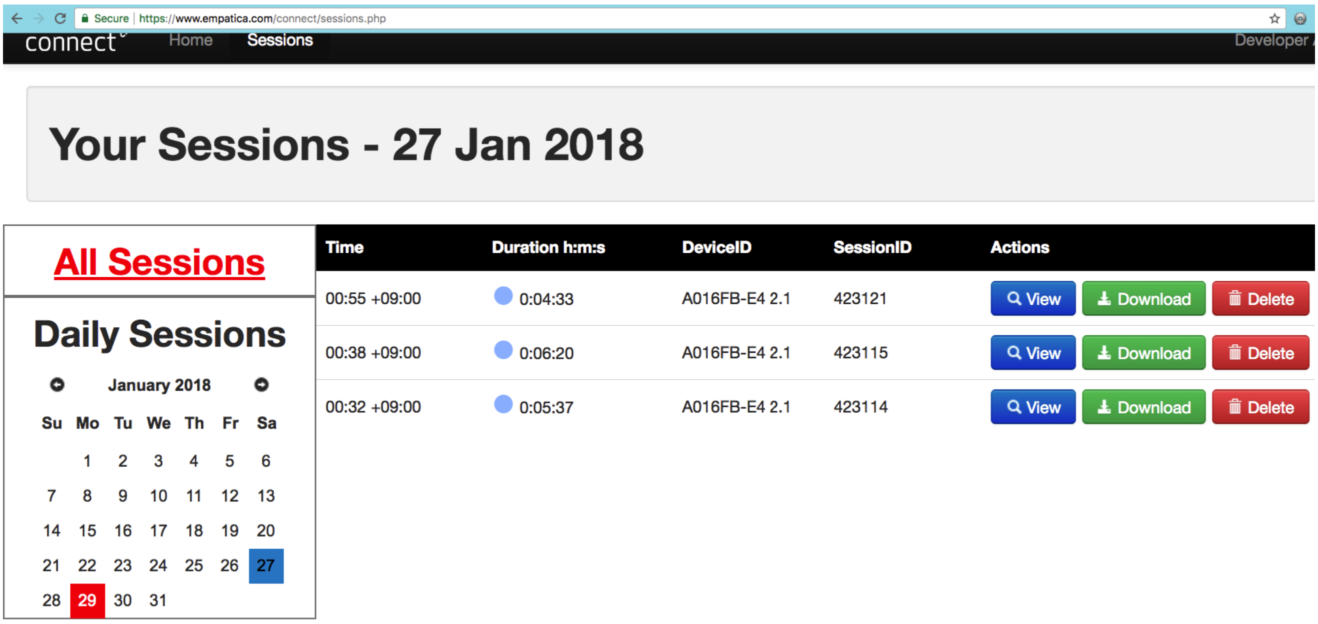


Figure 7.7: The Empatica Connect Interface.

7.5 Service Application

Unlike Google Firebase, Empatica Connect only provides a link for 3rd party application to download the data. Unfortunately, there is no available APIs to create, write, delete data for 3rd party application in Empatica Connect Server. Therefore, to deal with data collected by Empatica RT, at first, the service application must transfer data from Empatica Connect to Google Firebase Database Server. Then, the service application achieves real-time data from Google Firebase Server, applies different data preprocessing algorithms, visualize the information as final results.

Node.js is an open source, cross-platform runtime environment for developing server-side and networking services. Node.js allows developers to reuse the share resources freely by employing JavaScript across the stack which unifies the language and data format (JSON). The aim of Node.js to develop real-time network service with push ability, two-way connections, allow the client and server to initiate communication and exchange data efficiently. Therefore, we applied Node.js to transmit data from Empatica Connect to Google Firebase Database Server. The Node.js function used to transfer data is shown in the following.

```
let req = request.defaults({
        jar: request.jar(),
        headers: {'user-agent': 'Mozilla/5.0 (Macintosh;
        Intel Mac OS X 10_12_6)
        AppleWebKit/537.36 (KHTML, like Gecko)
        Chrome/63.0.3239.84
        Safari/537.36'
        }
});

await req({
        url: 'https://www.empatica.com/connect/caramba.php',
        method: 'POST',
        form: {
        username: 'username',
        password: 'password'
        }
});

let {
        body
} = await req
        ('https://www.empatica.com/connect/connect.php/users/
        17410/sessions?from=0&to=999999999999');
let data = JSON.parse(body);
for (let {id} of data) {
 if (!fs.existsSync(dataPath + id)) {
        let path = dataPath + id + '.zip';
        await new Promise(r => req.request('https://www.empatica.
                        com/connect/download.php?id=' +
                        id).pipe(fs.createWriteStream(path)).
                        on('close', r));
        fs.unlinkSync(path);
 }
}
```

Since its first released in 1991, Python programming language has attracted widespread attention from developers by its advancement. Python is a cross-platform, easy syntax, extensible, math-like, object oriented, open source, readability, and high level programming language with dynamic semantics. Although Python programs are usually supposed to run slower than Java programs, the general development time of a Python program is much less than a competitive Java program. It is one of the most recommended development languages for several years and remains to be a favorite among experienced developers. Python has particular tools that are incredibly essential in operating with machine learning systems by the abundant of frameworks and libraries, extensions. One of Python powerful frameworks is Django. Django is a free and open source, high-level Python Web framework that supports rapid development and clear, logical design. With the Python core, Django can take advantage by integrating Python built tools and libraries into the framework. Therefore, we decided to apply Django framework according with Python version 3.0 to build the core of Service Application. Django project can be started with the command:

```
$ django-admin startproject [projectname]
```

The structure of a Django project is illustrated in Figure 7.8

```
[projectname]/
├── [projectname]/
│   ├── __init__.py
│   ├── settings.py
│   ├── urls.py
│   └── wsgi.py
└── manage.py
```

Figure 7.8: Project Structure of a Django project

In this project, to ensure the quality of collected data, prevent the missing sensor data problem, we implemented various of Missing Sensor Data Preprocessing Algorithms such as EM, MI, KNNI which have been explained in Chapter 4. Impyute was written by E. Law at [108]. It is a super lightweight library of missing data imputation algorithms developed in Python 3. Impyute supports the features as follow.

- Imputation of Cross Sectional Data
  - Multivariate Imputation by Chained Equations.
  - Expectation Maximization.
  - Mean Imputation.
  - Mode Imputation.

  - Median Imputation.
  - Random Imputation.
- Imputation of Time Series Data
  - Last Observation Carried Forward.
  - Autoregressive Integrated Moving Average (WIP).
  - Expectation Maximization with the Kalman Filter (WIP).
- Dataset Generation
  - Datasets
    * MNIST.
    * Random uniform distribution.
    * Random gaussian distribution.
  - Missingness Corruptors
    * MCAR
    * MAR (WIP)
    * MNAR (WIP)
- Diagnostic Tools
  - Loggers.
  - Distribution of Null Values.
  - Comparison of imputations.
  - Little's MCAR Test (WIP).

Impyute can be easily installed by executing the following code:

```
$ pip install impyute
```

Once Impyute library implemented in the Django, the application can take advantage by applying a few line code to employ the missing sensor data imputation algorithms. The following is a simple example of EM algorithm by Impyute library.

```
>>> from impyute.datasets import random_uniform
>>> raw_data = random_uniform(shape=(5, 5), missingness=''mcar")
>>> print(raw_data)
        [
        [   1.    0.    4.    0.    1.]
        [   1.   nan    6.    4.   nan]
        [   5.    0.   nan    1.    3.]
        [   2.    1.    5.    4.    6.]
        [   2.    1.    0.    0.    6.]
        ]
```

```
>>> from impyute.imputations.cs import em_imputation
>>> complete_data = em_imputation(raw_data)
>>> print(complete_data)
        [
        [ 1.     0.     4.     0.     1.   ]
        [ 1.     0.5    6.     4.     4.   ]
        [ 5.     0.     3.75   1.     3.   ]
        [ 2.     1.     5.     4.     6.   ]
        [ 2.     1.     0.     0.     6.   ]
        ]
```

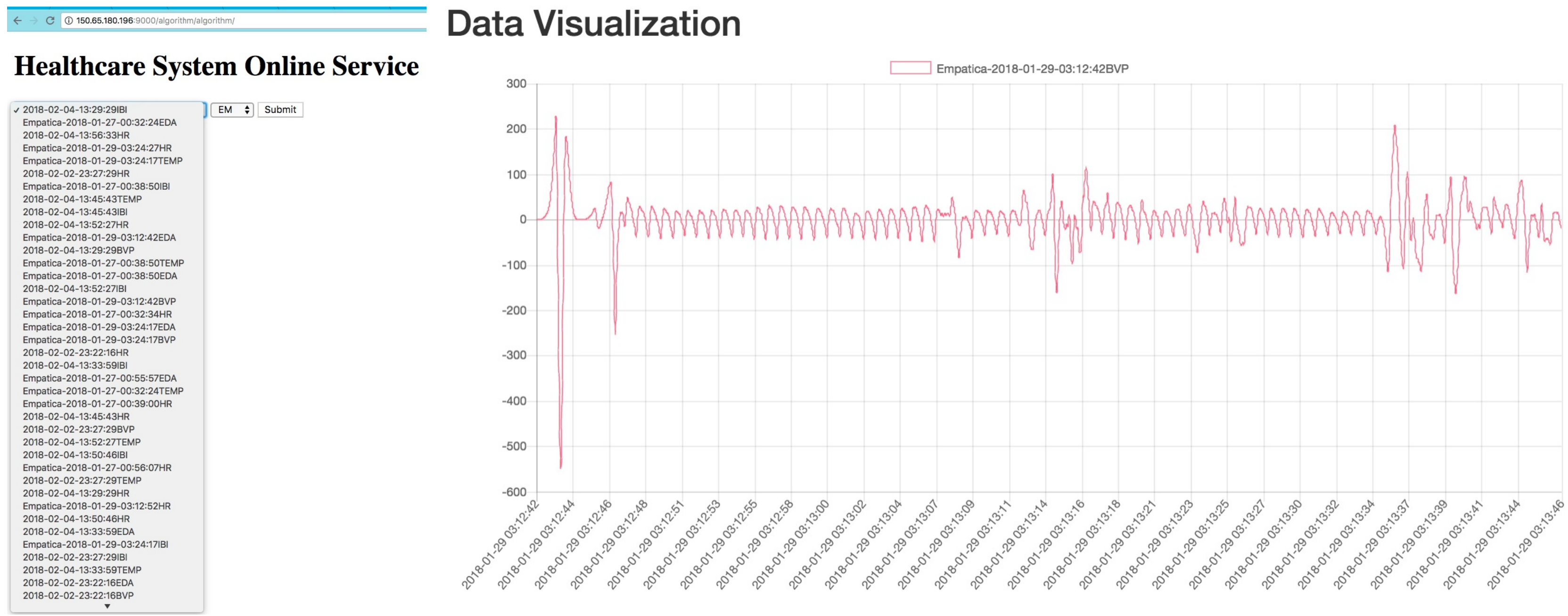
In order to visualize the real-time data, we chose to apply Chart.js [109]. It is an open source Javascript library greatly support real-time chart visualization for website application. The following is an simple example of Chart.js.

```
var ctx = document.getElementById('myChart').getContext('2d');
var myChart = new Chart(ctx, {
        type: 'line',
        data: {
          labels: ['M', 'T', 'W', 'T', 'F', 'S', 'S'],
          datasets: [{
                label: 'label',
                data: [12, 19, 3, 17, 6, 3, 7],
                backgroundColor: ''rgba(153,255,51,0.4)"
                }, {
                label: 'oranges',
                data: [2, 29, 5, 5, 2, 3, 10],
                backgroundColor: ''rgba(255,153,0,0.4)"
                }]
                }
});
```

The techniques utilized in the Service Application can be reviewed as follow:

- Python 3 Programming Language.
- Django Framework.
- Impyute Missing Sensor Data Library
- Chart.js Library.

Figure 7.9 represents a screenshot of the data visualization process (a) Healthcare System Online Service main screen (b) Data Visualization.



a) b)

Figure 7.9: The screenshot of the data visualization process (a) Healthcare System Online Service main screen (b) Data Visualization.

Chapter 8

Conclusion

In this chapter, we conclude our research survey, the contribution made to the field. Besides, future works are provided for further study and extension.

8.1 Concluding Remarks

We have solved the research objectives defined in Chapter as:

- Investigating and summarizing the existing knowledge and state-of-the-art technologies in the field.
- Provide an exhaustive understanding of the successful case studies on the IoT in the healthcare context which is expected to be useful for further research.
- Identify the current problems of smart healthcare systems based on the view of general system architecture. Besides, we will implement a smart healthcare system which applies methodologies to solve the current issues of the field.

Among the abundant researchers regarding the potential and implementation of IoT in healthcare already existed, this research gives a survey based on various aspects of IoT-based healthcare systems from necessary collected data, sensors and communication technologies, the preprocessing algorithms, the algorithms applied to applications and the typical applications in the healthcare area. In addition, the research provides detailed research progress concerning how the IoT can support healthcare. The technologies have been compared with each other while the trend of methods has been illustrated in the report. The discussion in **Chapter 6** on several critical problems such as sensors data problems, the accuracy of indoor location monitoring system and the system design is expected to facilitate fundamental understanding for further studies in the field. Moreover, the smart healthcare system was implemented, as a basic system that support future research.

- In Chapter 2, the necessary information needs to be collected in order to ensure the quality of life, safety and well-being are identified. This information was classified into Physiological information, the environmental aspects, indoor location information. The role of each data in smart healthcare system was also discussed in detail.

- In Chapter 3, the advance in infrastructures and technologies has been reviewed. Based on the necessary information needs to be collected detailed in Chapter 2, the sensors have been organized into Physiological Sensors; Environmental Sensors and Location Sensors. Technical detail of each type of sensor is explained in detail. Moreover, the advantage of Body Area Network Architecture (BAN) in the smart healthcare system is also described. Furthermore, wireless communication protocols are compared with each other, which provide the comprehensive view for the researchers in the task of determining Supporting Infrastructure and Technology for their future research.

- The state-of-art of data preprocessing and data processing steps are sequentially identified in Chapter 4 and Chapter 5. EM, KNNI and MI are the most well-known data preprocessing step while Genetic Algorithm, Decision Tree and Support Vector Machine are widely applied by the researcher in processing step for smart healthcare application. The recent research also distinguished the promising ability of RNN and LSTM for smart healthcare system. Besides, smart healthcare applications are categorized into Pervasive Monitoring and Medical Informatics - Prediction applications. Data requirements, as well as technologies aspects of each application, are addressed. Finally, industrial IoT applications in healthcare and smartphone apps for general healthcare are detailed.

- Lack of standardization, missing sensor data, smart health platform, human in the loop concept, indoor location accuracy and variety of other issues of the smart healthcare system are identified in Chapter 6. Thanks to the discussion of this Chapter, the researchers can have a comprehensive view of the current problems of the field.

- In Chapter 7, we have implemented the smart healthcare system. By utilizing various advanced technologies such as modern sensors, BLE, algorithms, Python language, Nodejs language, this system promises to be a core for our further research in smart healthcare system.

8.2 Contributions

One of the benefits of this research project is that it draws from the existing smart healthcare surveys. The current research have paid attention to each aspect of the field. Hence, this research survey revisited these studies and contributed the comprehensive smart health system by considering algorithms and implementation. Our research survey reveals and insights the potential opportunities in the research line of smart healthcare and its computing technologies. The main contributions of this survey will provide more particurlaly benefits to whom interested in the smart healthcare area.

Through the survey, **researchers**, who are the target readers of this research, can get enough understanding, save their time in searching articles and reduce repetitious work for supporting the task of designing and developing the smart healthcare system. The survey contributes by:

- Investigating and summarizing the existing knowledge and state-of-the-art technologies in the field.
- Presenting an overview of research conducted in different sectors of IoT-based healthcare system.
- Providing an extensive survey of IoT-based healthcare services and applications.
- Providing an exhaustive understanding of the successful case studies on the IoT in the healthcare context which is expected to be useful for further research.
- Classifying existing IoT-based healthcare technologies ranging from sensors, networks, algorithms, applications and presenting a summary of each.
- Highlighting the most comprehensive recent compilation of IoT-based healthcare projects and publications.
- Explaining in-depth insight into challenges and open issues that must be addressed to make IoT-based healthcare technologies robust. Suggest solutions to tackle out these challenges and open issues.
- Implementing a core system for smart healthcare. The physiological data collected from the implemented system is expected to be valuable for the data scientists in their future research.

For **stakeholder** such as doctors, nurses, caregivers, hospitals, pharmaceutical companies, who are new to IoT-based healthcare technologies, the research survey provides an overview of the field as well as the encouraging potential of applying smart healthcare in supporting the task of caregiving. Based on our implemented system, doctors, nurses, and caregivers can remotely monitor real-time data of the patients. Hospitals can apply IoT in the task of managing the victims while pharmaceutical companies can apply IoT to reduce the medical cost.

For **healthcare manufacturers**, a lack of standardization issues identified in the research is expected to encourage the cooperation between healthcare manufacturers to introduce the widely adopted standardization for IoT healthcare. This cooperation will not only benefit the patients but also the healthcare manufacturer itself.

For **family members**, the implemented system in this research provides ability for remote monitoring patient health. Besides, the extensive variety of healthcare applications classified in the survey suggests family members a wide range of selection in healthcare applications for the patient.

For **patients**, who are the end-user and also the one who gets the most benefit from the research. The utilization of this survey benefits patient with better medical assistance, shorten treatment time, lower medical costs and more satisfying health care services.

8.3 Future work

The vision of this research is building an Integrated Home-Based Healthcare Monitoring System that can ensure wellness for the individual who is living inside the house. In this system, the data collected by health sensors, environment sensors and location sensors will be pre-processed and sent to the Smart Health Platform through the IoT gateway. The Smart Health Platform not only stores the data but also controls the connectivity of device and provides the APIs for the third-party application. The Smart Health Applications can provide several safety applications such as fall detection and emergency assistance.

The proposed Home-Based Healthcare Monitoring System is consisting of three main points:

- Firstly, we will deploy the sensors including personal health sensors, environment sensors and the location sensors. Missing-data imputation is the major concern in sensors deployments. Although Long short-term memory (LSTM), a kind of Recurrent Neural Networks, is applied to successfully solve many problems of time series data such as speech recognition video summarization, music recognition; the technology is quite new in the missing-data imputation field. Thanks to their advantages, the application of LSTM in the smart healthcare field promise to solve the missing data imputation problem. The research at [110] presented the first study to empirically evaluate the ability of LSTMs to recognize patterns in multivariate time series of clinical measurements. Therefore, our research will apply LSTM with the improvement methods such as Kalman filters to impute the missing data of sensors.

- Secondly, it is the Indoor Location Monitoring System. Starting by using a small lightweight accelerometer and compass BLE sensor tags attached to the user's clothing, we will develop a step detector to estimate the step and step's length of the user. The collected data by sensors will be pre-processed as the result of the first step. Received signal strength indication (RSSI) data received from the wireless access point can be used to identify the location of the human inside the house, however, the accuracy is not high due to multipath fading, indoor shadowing and interference

[111]. Therefore, we need to apply Monte Carlo or Kalman filter methods for constructing RSSI-based localization algorithms. Combine the output of step detector and RSSI estimated location, we can define the location of the human inside the house. The performance of the proposed method is tested using K-Nearest Neighbors, Decision Tree, Random Forest and Support Vector Machine classifiers to get the higher accuracy.

- Finally, it is the Smart Health Platform. Apache Hadoop software library is a framework that allows for the distributed processing of large data sets across clusters of computers using simple programming models. It is designed to scale up from single servers to thousands of machines, each offering local computation and storage. Therefore, we will apply Apache Hadoop in the data storage layer. The proposed Smart Health Platform provide the connectivity and device management methods, especially we will build the APIs that can allow the third-party application easy to implement in our system.

For exploring and evaluating our proposed Integrated Home-Based Healthcare System, we will develop several applications for ensuring the safety such as fall detection and emergency assistance inside the house. The application will be based on the high-quality data collected by sensors, the estimated location of the human inside the house. The performance result of each step will be compared with previous experiments.

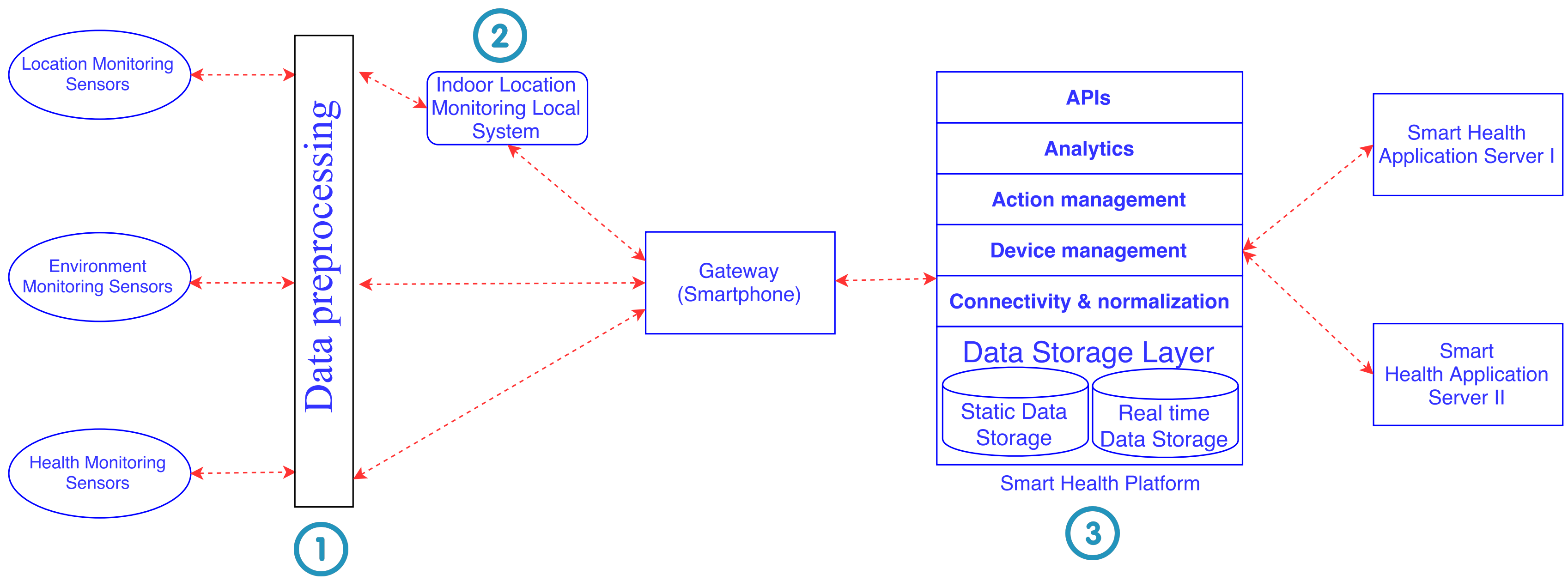


Figure 8.1: The proposed integrated home-based healthcare system

Appendix A

Review Selected Indoor Location Systems

A.1 CASAS: A Smart Home in a Box [112]

Strength	*Weakness*
Simple, small, easy to implement	Just collect the environment features
CASAS-AR learns an activity model based on training data from multiple smart home sites and thus generalize for an application to new smart home with no training data	The battery life of sensors is short.

Table A.1: Strength and Weakness of the research

"Smart home in a box" which is a lightweight smart home design that is easy to install and provides SH capabilities out of the box with no customization or training needed.

In the architecture of CASAS, the physical layer contains hardware (sensors, actuators) which communicates with each other by ZigBee Wireless. Every component of the CASAS communicates via a XMPP bridge to the Publish/subscribe manager. The CASAS architecture uses lightweight APIs, therefore it is easily maintained.

All the CASAS components fit into a single small box. They can install a new smart home in just two hours and remove the equipment in 30 minutes.

For the activity recognition, CASAS use the software called AR which implement the Support Vector Machine (SVM) method. AR provides the learning problem as one of mapping the sequence of the k most recent sensor events to a label indicating the activity corresponding to the most recent event in the sequence. For example, the sequence of sensor events consisting of

2011-06-15	03:38:23.271939	BedMotionSensor	ON
2011-06-15	03:38:28.212060	BedMotionSensor	ON

2011-06-15 03:38:29.213955 BedMotionSensor ON
could be mapped to a Sleep activity label.

For the activity discovery for unlabeled data, they use the unsupervised learning algorithm.

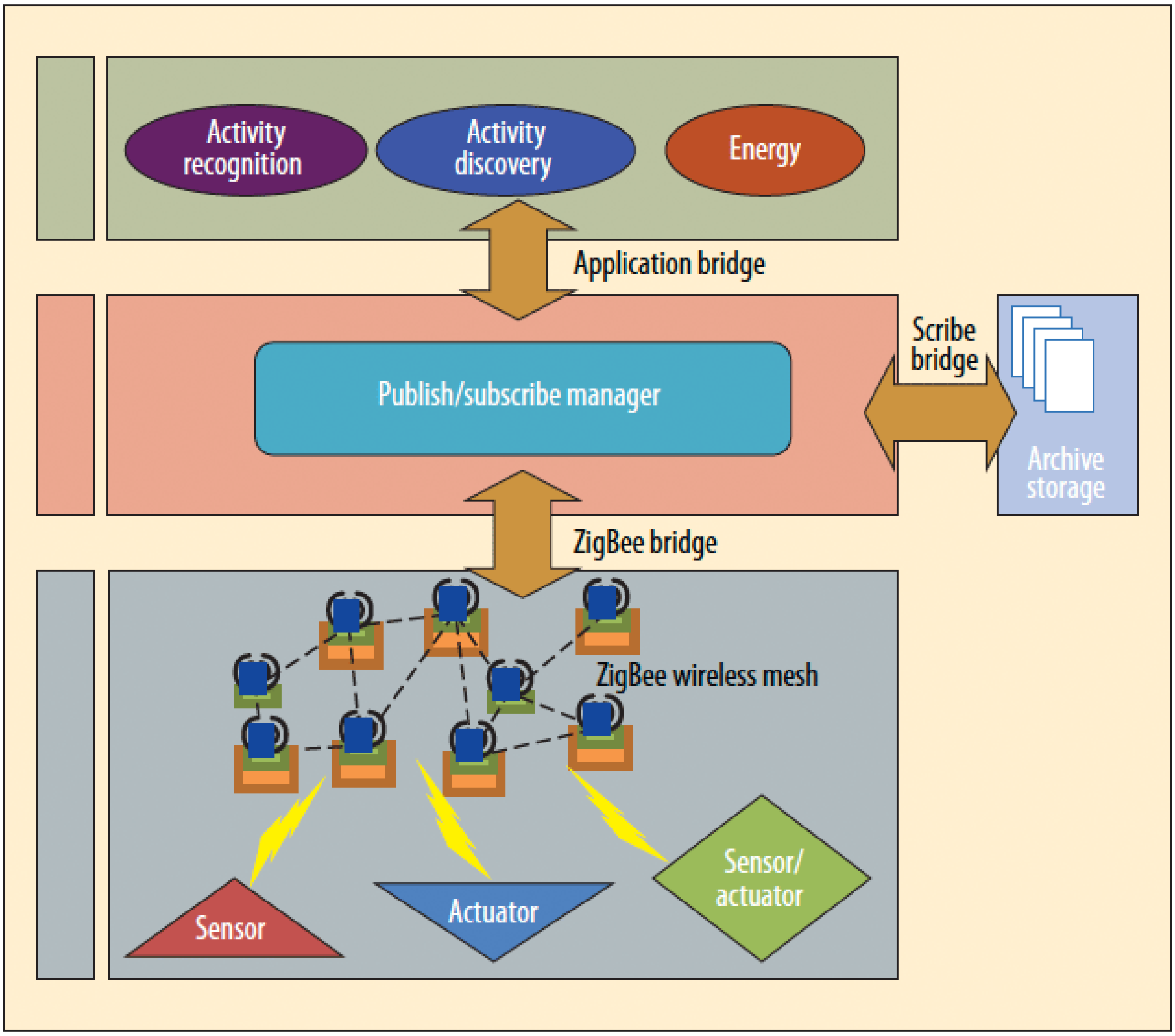


Figure A.1: CASAS smart home components.

A.2 Smartphone-Based Real-Time Indoor Location Tracking With 1-m Precision [113]

Problem: The actual identification of activities is often difficult as the accuracy of existing indoor tracking sensors and the corresponding analysis software is still in its infancy. Indoor Movement tracking is not always easy to accurately detect, especially in the smart health field.

Goals: to develop a lower cost, reduced intrusiveness, and higher mobility, deployability, and portability localization system that can identify a patient's real-time location in a home environment with maximum estimation error of 2 m at a 95% confidence level.

Background: Many studies have been done to estimate the step and location inside the house. For example, the average distance estimation error for their indoor 16-step straight-line walking experiments was 5.5% with a maximum error of 2.05 m with the combination of using smartphone and the map of the floor. Some studies even used two smartphones at one time for tracking location which is not only drain a smartphone's battery but also impractical. Some studies could achieve an overall estimation error of about 2 m. There is a need of a lightweight but accurate localization algorithm suitable for execution in a smartphone, avoid adopting detailed wireless signal maps, as well as excessive hardware installations.

Technical Aspect: First develop a step detector using the accelerometer and compass of an iPhone 5, then create a radio-based localization subsystem using a Kalman filter and signal strength indication (RSSI) data received from three triangular deployed BLE sensors (i.e., TI SensorTags) to estimate the location and movement of a tracked target who held a User Agent in their hand and walked around a small concrete-walled office (9 m x 6 m). The three BLE sensors (denoted as S0, S1, and S2) were placed against three different walls of the office at 1.1 m height from the floor.

For evaluation, the four mechanisms step detection, maximum-likelihood estimation, follow-step-detection, and tight-coupling sensor fusion, were implemented to track the user's real-time movement. With both the maximum estimation error well under 2 m and an average estimation error of 0.47 m for the proposed mechanism - step detection (and 0.56 m for tight-coupling sensor fusion mechanism) achieved the highest accuracy.

Result: The study has presented a proof-of-concept localization system for real-time indoor patient movement pattern telemonitoring with high accuracy.

Strength: What makes this research so compelling is the fact that only one form of sensor data was used, which opens the possibility of further improvements by combining sensor technologies. In fact, the authors plan to extend this work by adding other sensor data from force sensitive resistors around the location but this could easily be extended further with Wi-Fi, IR, sound, and video data. This work was part of a larger project to implement a patient telemonitoring solution.

Weakness: use the phone sensor, therefore the patient needs to carry the phone anytime. It can be improved by using a small lightweight sensor tag (e.g., TI SensorTag with on-board inertial sensors) attached to the user's clothing to minimize the intrusiveness.

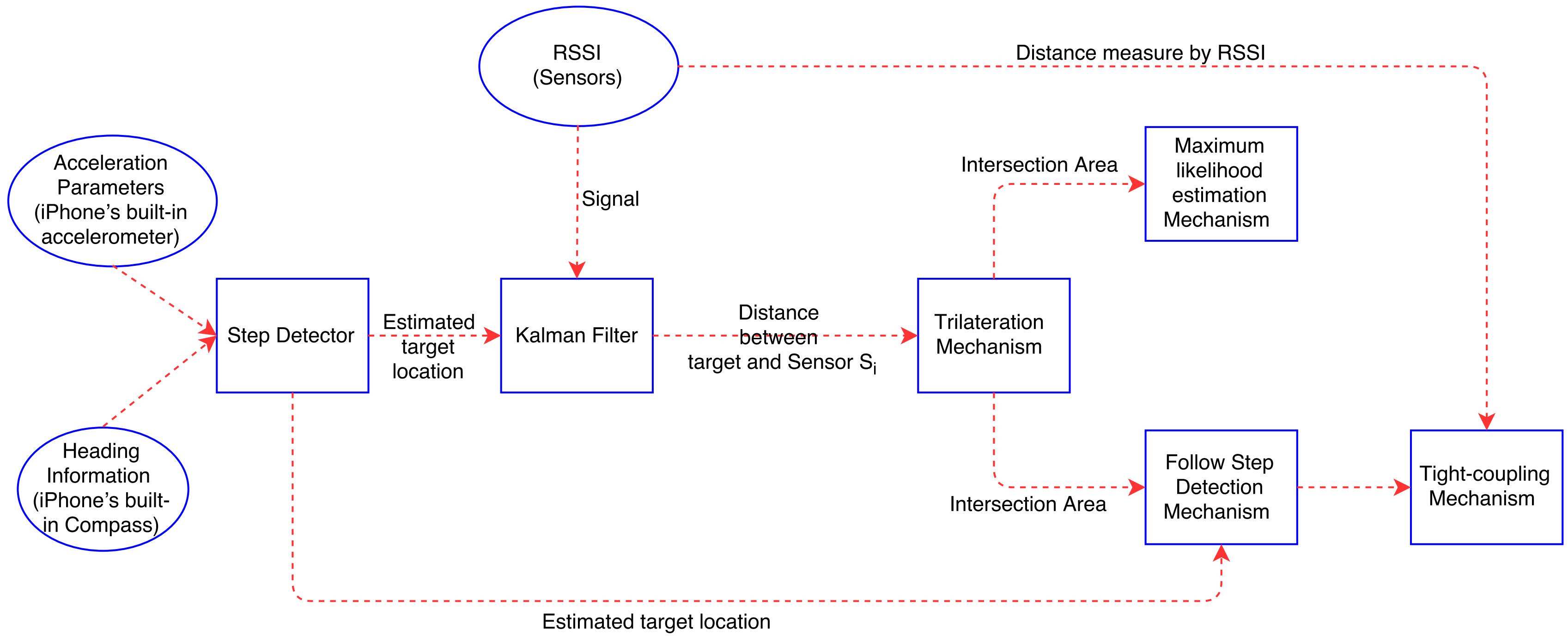


Figure A.2: Smartphone-Based Real-Time Indoor Location Tracking With 1-m Precision

Appendix B

A Study on RNN and LSTM Algorithm for Smart Healthcare System

B.1 Overview

Nowadays, sequences are everywhere in our world. They are not only speech of actions in videos, video summarization, music recognition but also forecasting like weather, markets and politics. Furthermore, physiological data collected by sensors is time series data, a special case of sequential data where sequence is based on time. Since Feed-Forward Neural Network shows their limitation of dealing with sequences, we need a new kind of machine learning algorithm or a new neural network that can learn sequences. This kind of neural network must be able to remember the past, especially pay attention to the important parts of the past. Recurrent neural networks or RNN (Rumelhart et al., 1986a) are family of neural network that have shown great promise in many NLP tasks, mainly for processing sequential data. In this research, based on WildML's RNN tutorial [114], Machine Learning by Andrew Ng [115] and Neural Networks for Machine Learning by Geoffrey Hinton [116], we will review the architecture and learn how to train RNN, come up with the vanishing and exploding gradient problem of RNN which can be solved by LSTM, a special kind of RNN. Besides, the Google's Tensorflow Framework, one of the well-known libraries supporting RNN and LSTM will be introduced. Finally, the advantages of applying RNN and LSTM in smart healthcare field are clearly explored.

B.2 Feed-Forward Neural Network

In Feed-Forward Neural Network which is a simple type of Neural Network architecture, we have layers' structure where the output of one layer is being fed as the input to the following layer and each unit does a relatively straightforward computation. It takes the input x multiplies by a weight W, performs a sum with bias b and then passes all through an activation function g to yield the output y:

$$y_i = g(\sum_j W_{ij}x_i + b_i) \tag{B.1}$$

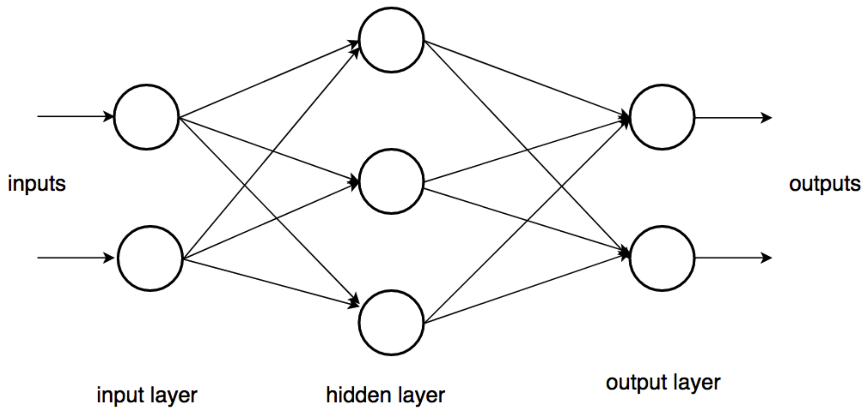


Figure B.1: Feed-Forward Neural Network

The above structure can also be converted into vector notation structure:

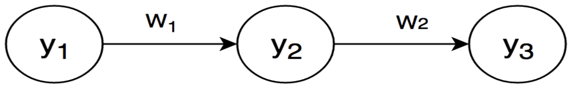


Figure B.2: Feed-Forward Neural Network by vector notation

We take the output from the previous layer y_{k-1} multiply by weight matrix W, add a bias b and then pass all of this through an activation function g to yield the output vector y_k. The function (1) becomes:

$$y_k = g(\sum_j W y_{k-1} + b_k) \tag{B.2}$$

To measure of "how good" a neural network did with its given training sample and the expected output, we use cost function *C(output, truth)*. A cost function takes two vectors output and truth as the input and returns the distance how far is the output away from our desired ground truth. We want to be able to adjust our weights such that the output vector looks exactly like the ground truth. We take this cost and pass it back through the network and obtain for each weight element ΔW which is a gradient that tells us well how are we going to adjust the weight in order to decrease the cost. An algorithm called back propagation can compute these gradients.

We want to compute the derivative $\frac{\partial C}{\partial W^*}$ to adjust the weights to decrease the cost. This derivation tells us how quickly the cost changes when we change the weight or bias.

$$C(y, truth) = C(g(\sum_j W_{ij}x_i + b_i)) \tag{B.3}$$

If $\sum_j W_{ij}x_i + b_i = a(W_{ij}x_i)$
We have
$C(y, truth) = C(g(a(W_{ij}x_i)))$
Therefore, we have

$$\frac{\partial C}{\partial W^*} = \frac{\partial C}{\partial g} \cdot \frac{\partial g}{\partial a} \cdot \frac{\partial a}{\partial W^*} \tag{B.4}$$

One of the problems in Feed-Forward Neural Network is that we have to fix the length of the input layer. The input layer size is fixed to the size of the input for example size of an image must be 28 by 28 pixels across all the different training examples. However, in many cases especially for sequence data, the length of the input sequences can vary from example to example up to almost an order of magnitude. The other problem of Feed-Forward Neural Network is the problem of independence. In a traditional neural network, we assume that all inputs and outputs are independent of each other, for example, the different training examples such as images are independent of each other. However, many tasks show that it is not the good idea for independence. If we want to predict the next word in a sentence we better know which words came before it. Therefore, researchers introduce the new technology to make use of sequential information called Recurrent Neural Networks (RNNs).

B.3 Recurrent Neural Networks (RNNs)

Recurrent Neural Networks or RNNs are the families of neural network for processing sequential data. Recurrent Neural Networks have a "memory" which captures information about what has been calculated so far and execute the same task for every element of a sequence, with the output being depended on the previous computations.
Many researchers have taken advantage of RNN to earn great success in many NLP tasks from Language Modeling and Generating Text, Machine Translation, Speech Recognition to Generating Image Description. However, these powerful sequence models do not enjoy widespread use because it is extremely difficult to train them properly.

RNN can not only learn the local and long term temporal dependencies in the data but can also accommodate input sequences of variable lengths.
A recurrent neural network at time t it takes input as x_t multiplies by a weight matrix U and then this unit is passed as the input of the unit at time *t+1*. In this structure, s_t, which is calculated based on the previous hidden state and the input at the current step, becomes the memory of the network. s_t need the activation function to transform the inputs of the layer into the outputs. To fit nonlinear hypotheses, *tanh*, the sigmoid

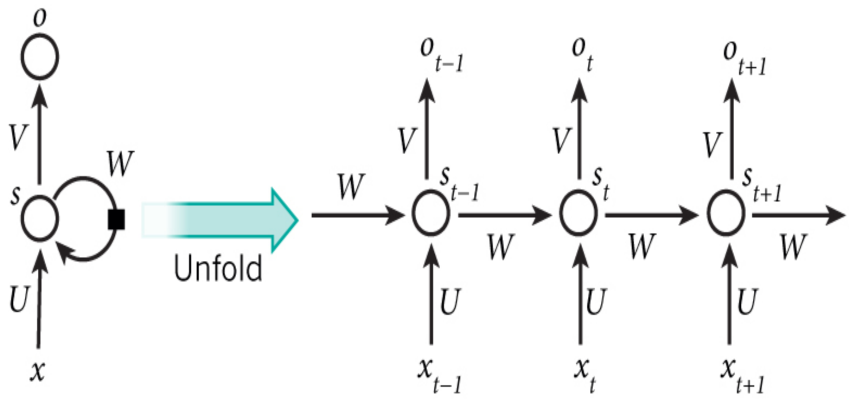


Figure B.3: A recurrent neural network and the unfolding in time of the computation involved in its forward computation. [114]

Table B.1: The notations have been used in this report

Neural layer	Description
x_t	The input value at time step t.
s_t	The hidden state at time step t.
o_t	The output value.
b, c	Bias vectors along with weight matrices U, V and W.
U	The weight matrix from input layer to hidden layer.
W	The weight matrix from previous hidden layer to the hidden layer
V	The weight matrix from hidden layer to output layer.

function or $ReLUs$ are one of the most common activation functions that have been used. Because of the derivate of *tanh* can be computed easily by using the original value that $(tanhx)' = 1 - tanh^2x$, we will use *tanh* as the activation function for s_t.
For the activation function of the output layer, we can use the softmax function as it is the simply way to convert raw scores, probabilities and its generalization to multiple classes.
Forward propagation can start with a specification of the initial state *s(0)*. Then for $t = 1$ to $t = \tau$, we apply the following update equations:

$$s_t = tanh(Ux_t + Ws_{t-1}) \tag{B.5}$$

$$\hat{y} = softmax(Vs_t) \tag{B.6}$$

The RNN computes a sequence of hidden states s_t and a sequence of outputs $\hat{y}_t$ by the following key steps: We need a loss function L to measure the errors to train our network.

for t from 1 to τ:

1. Compute hidden activations of time t with current input and hidden activations for *(t-1)*.
2. For all j in the output units compute the net_j (dot product of V with s_t).
3. Apply the softmax function on the net_j and get the probability distribution for time t.

The final goal is find the parameter U, V and W that can minimize L for our training data.
During back-propagation training, depending on the target values, the output node values should be either 1.0 or 0.0. When we apply Mean Squared Error (MSE), the gradient contains a term of *(output) * (1 – output)*. As the computed output gets closer and closer to either 0.0 or 1.0, the value of *(output) * (1 – output)* gets smaller and smaller into 0. However, when we use Cross-entropy error, there is not a term (output) * (1 – output) in the formula. The weight changes do not get smaller and smaller and so training is not likely to stall out. Therefore, cross entropy loss is the most common choice for the loss function.

$$L(y, \hat{y}) = -\frac{1}{N}\sum_t y_t log \hat{y}_t \tag{B.7}$$

For computational straightforwardness later, we define:

$$\begin{aligned} E_t &= -y_t log \hat{y}_t \\ &= -y_{t_i} log \hat{y}_{t_i} \end{aligned} \tag{B.8}$$

A gradient-based technique for training certain types of recurrent neural networks are Backpropagation through time (BPTT). The key steps of Backpropagation through time are: To understand how can we calculate the gradient, we should look on the very basic math concept called *Einstein Summation* which is useful for dealing with the chain rule and matrix derivatives. For example, suppose we have a function $f(x, y)$ where x, $y \in R^N$ and x, y are functions of $r \in R$, $x = x(r)$ and $y = y(r)$. Then,

$$\frac{\partial f}{\partial r} = \frac{\partial f}{\partial x_i}\frac{\partial x_i}{\partial r} + \frac{\partial f}{\partial y_j}\frac{\partial y_j}{\partial r} \tag{B.9}$$

U, V and W are the parameter of RNN that need to compute the gradient of loss function [3].

for t from 1 to τ:

1. Compute the delta at the output.
2. Compute Δw_{ji} where w is the (softmax) weight matrix W^S.
3. Determine the bias terms.
4. Backpropagate and compute delta for hidden layer.
5. Compute the updates to weight matrix V and W.
6. Perform BPTT by computing the error to be propagated to the previous layer.

B.3.1 V

The parameter V is present only in the function $\hat{y}$. If $q_t = Vs_t$, we have:

$$\frac{\partial E_t}{\partial V_{ij}} = \frac{\partial E_t}{\partial \hat{y}_{t_k}} \frac{\partial \hat{y}_{t_k}}{\partial q_{t_l}} \frac{\partial q_{t_l}}{\partial V_{jk}} \tag{B.10}$$

From (B.8), we have:

$$\frac{\partial E_t}{\partial \hat{y}_{t_k}} = -\frac{y_{t_k}}{\hat{y}_{t_k}} \tag{B.11}$$

From (B.6), $\hat{y}$ is the sigmoid function, therefore,

$$\begin{aligned} \frac{\partial \hat{y}_{t_k}}{\partial q_{t_l}} &= \begin{cases} -\sigma(q_t)_k \sigma(q_t)_l, k \neq l \\ \sigma(q_t))_k(1 - \sigma(q_t)_k), k = l \end{cases} \\ &= \begin{cases} -\hat{y}_{t_k}\hat{y}_{t_l}, k \neq l \\ \hat{y}_{t_k}(1 - \hat{y}_{t_k}), k = l \end{cases} \end{aligned} \tag{B.12}$$

Putting (B.11) and (B.12), together give us a sum over all values of k to calculate $\frac{\partial E_t}{\partial q_{t_l}}$:

$$\begin{aligned} -\frac{y_{t_l}}{\hat{y}_{t_l}} \left(\hat{y}_{t_l} (1 - \hat{y}_{t_l}) + \sum_{k \neq l} \left(-\frac{y_{t_k}}{\hat{y}_{t_k}} \right) (-\hat{y}_{t_k}\hat{y}_{t_l}) \right) &= -y_{t_l} + y_{t_l}\hat{y}_{t_l} + \sum_{k \neq l} y_{t_k}\hat{y}_{t_l} \\ &= -y_{t_l} + \hat{y}_{t_l} \sum_k y_{t_k} \end{aligned} \tag{B.13}$$

If yt are are all one-hot vector, then $\sum_k y_{t_k} = 1$, then

$$\frac{\partial E_t}{\partial q_{t_l}} = -y_{t_l} + \hat{y}_{t_l} \tag{B.14}$$

We have $q_t = Vs_t$, so $q_{t_l} = \; V_{l_m} s_{t_m}$, then

$$\begin{aligned}\frac{\partial q_{t_l}}{\partial V_{ij}} &= \frac{\partial (V_{l_m} s_{t_m})}{\partial V_{ij}} \\ &= \delta_{il}\delta_{jm} s_{t_m} \\ &= \delta_{il} s_{t_j}\end{aligned} \tag{B.15}$$

Therefore, from (B.14) and (B.16), we have:

$$\begin{aligned}\frac{\partial E_t}{\partial V_{ij}} &= (-y_{t_l} + \hat{y}_{t_l}) s_{t_j} \\ &= (- \; y_{t_l} + \; \hat{y}_{t_l}) \bigotimes s_t\end{aligned} \tag{B.16}$$

where $\bigotimes$ is the outer product.

B.3.2 W

The parameter W is present in the function $\hat{y}$ and s_t. From (B.5), let $z_t = Ux_t + Ws_{t-1}$ then $s_t = tanh(z_t)$.

$$\frac{\partial E_t}{\partial W_{ij}} = \; \frac{\partial E_t}{\partial \hat{y}_{t_k}} \frac{\partial \hat{y}_{t_k}}{\partial q_{t_l}} \frac{\partial q_{t_l}}{\partial s_{t_m}} \frac{\partial s_{t_m}}{\partial W_{ij}} \tag{B.17}$$

We have

$$\begin{aligned}\frac{\partial q_{t_l}}{\partial s_{t_m}} &= \frac{\partial (V_{lb} s_{tb})}{\partial s_{t_m}} \\ &= V_{lb}\delta_{bm} \\ &= V_{lm}\end{aligned} \tag{B.18}$$

Since there is an implicit dependence of s_t on W_{ij} through s_{t-1}, therefore we have:

$$\frac{\partial s_{t_m}}{\partial W_{ij}} \; \rightarrow \; \frac{\partial s_{t_m}}{\partial W_{ij}} + \frac{\partial s_{t_m}}{\partial s_{t-1_n}} \frac{\partial s_{t-1_n}}{\partial W_{ij}} \tag{B.19}$$

It is also clear that on RNN, s_{t-1} depends on s_{t-2}, this process continues until we reach *s-1*, which is an initialized to a vector of zero. The equation (B.19), can be rolled as

$$\frac{\partial s_{t_m}}{\partial W_{ij}} \; \rightarrow \; \frac{\partial s_{t_m}}{\partial W_{ij}} + \frac{\partial s_{t_m}}{\partial s_{t-1_n}} \frac{\partial s_{t-1_n}}{\partial W_{ij}} + \frac{\partial s_{t_m}}{\partial s_{t-1_n}} \frac{\partial s_{t-1_n}}{\partial s_{t-2_p}} \frac{\partial s_{t-2_p}}{\partial W_{ij}} \tag{B.20}$$

$$\frac{\partial s_{t_m}}{\partial W_{ij}} = \; \frac{\partial s_{t_m}}{\partial s_{r_n}} \; \frac{\partial s_{r_n}}{\partial W_{ij}} = \sum_{r=0}^{t} \frac{\partial s_{t_m}}{\partial s_{r_n}} \; \frac{\partial s_{r_n}}{\partial W_{ij}} \tag{B.21}$$

Combine all above together yields:

$$\frac{\partial E_t}{\partial W_{ij}} = (- \; y_{t_l} + \; \hat{y}_{t_l}) \, V_{lm} \sum_{r=0}^{t} \frac{\partial s_{t_m}}{\partial s_{r_n}} \; \frac{\partial s_{r_n}}{\partial W_{ij}} \tag{B.22}$$

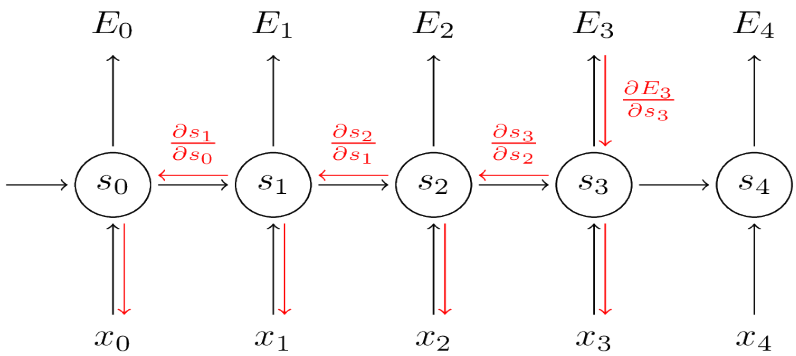


Figure B.4: Example in case of 3 layers, W is used in every step up to the output we care about, we need to backpropagate gradients from $t = 3$ through the network all the way to $t = 0$. [114]

B.3.3 U

The task of computing the gradient of U is similar to doing it for W because they are both present in the function y and s_t. Since the value of $\frac{\partial s_{r_n}}{\partial U_{ij}}$ and $\frac{\partial s_{r_n}}{\partial V_{ij}}$ are different, U and W are different.

$$\frac{\partial E_t}{\partial U_{ij}} = \frac{\partial E_t}{\partial \hat{y}_{t_k}} \frac{\partial \hat{y}_{t_k}}{\partial q_{t_l}} \frac{\partial q_{t_l}}{\partial s_{t_m}} \frac{\partial s_{t_m}}{\partial U_{ij}} \tag{B.23}$$

$$\frac{\partial E_t}{\partial U_{ij}} = (-\ y_{t_l} +\ \hat{y}_{t_l})\, V_{lm} \sum_{r=0}^{t} \frac{\partial s_{t_m}}{\partial s_{r_n}} \frac{\partial s_{r_n}}{\partial U_{ij}} \tag{B.24}$$

B.4 The vanishing and exploding gradients problem of RNNs

For many years, researchers found that they could never be able to train the RNNs over long time periods. For example, from the equation (B.17) and in case the gradient of E_3:

$$\frac{\partial E_3}{\partial W} = \sum_{k=0}^{3} \frac{\partial E_3}{\partial \hat{y}_3} \frac{\partial \hat{y}_3}{\partial q_3} \frac{\partial q_3}{\partial s_k} \frac{\partial s_{t_k}}{\partial W} \tag{B.25}$$

We can see that $s_3 = tanh(Ux_t + Ws_2)$ depends on s_2, which depends on W and s_1, so on. Therefore, to compute gradient we have to back propagate all the way to time T equals zero across for every single time step. For example, we have a recurrent neural network trained on a long sequence around 100 time steps. To calculate the derivative at the hundredth time step we have to multiply a bunch of derivatives all the way back

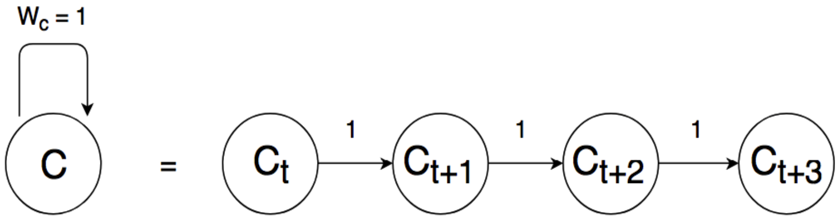


Figure B.5: Memory cells of LSTM

to t equals zero in order to properly capture the influence of the weight matrix W. The magnitude of the update is going to scale with the size of these weight matrices because we are multiplying all these derivatives repeatedly. In the case that the weights are small, the gradients will shrink and become exponentially small to zero. This is the vanishing problem, the state at zero gradient steps end up not learning long-range dependencies. Similarly, if the weights are big, we could get exploding instead of vanishing gradient which means the gradients will get huge and wipe out all our knowledge. That is called the exploding gradient problem. It seems like that good initial weight can be the idea for solving the problem. However, a good initial weight matrix is just can reduce the effect of vanishing and exploding gradients. It is hard to detect the dependency of the current output or an input from many time-steps ago. RNNs have difficulty dealing with long-range dependencies. The derivative of $ReLU$ is a constant of either 0 or 1, the using of $ReLU$ becomes a more preferred solution instead of *tanh* or *sigmoid activation function*. The most common approach is to use some of these other architectures that are still recurrent neural networks however have been especially designed to combat this vanishing gradient problem. It is Long Short-Term Memory (LSTM).

B.5 Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) was first introduced in 1997 by Hochreiter and Schmidhuber [4] in in an effort to solve the problem of getting a recurring neural network to remember things for a long time. The core of a LSTM model is very simple with a memory cell C that has a recurrent weight with itself of 1. This memory cell C is simply taking its input from the previous time step and then copy it over to the next time step. By designing a memory cell that used logistic and linear units with multiplicative interactions, the LSTM could remember things for hundreds of time steps. In LSTM, instead of propagating through directly from one neuron to the next which cause the problem of vanishing gradient, the state go through the structure called gates. Gates are use to protect and control the cell state. Only if a logistic gate is turned on, the information can gets into the memory cell. The sigmoid function at the gate produces the values between 0 and 1, which describe how much the information can get through that gate. A value 0 mean that "nothing can get through" while 1 mean "anything can get through". There are three of these gates in an LSTM which all have the same dimensions

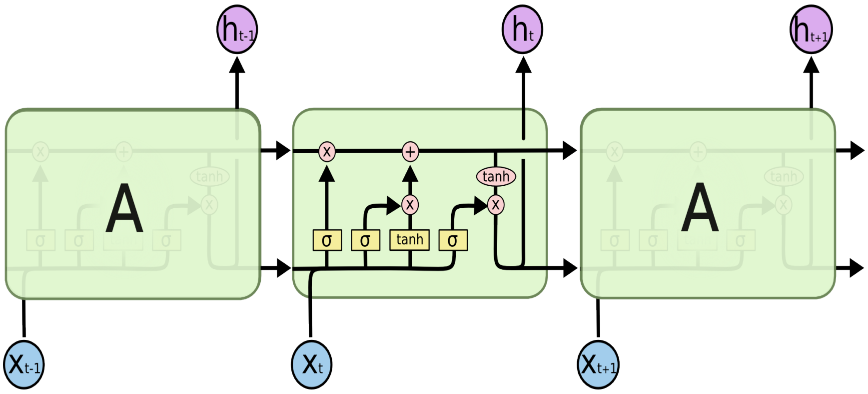


Figure B.6: The repeating module in an LSTM contains four interacting layers.

ds (the size of the hidden state), they are forget f, input i and output o gate layers. The input gate defines how much of the newly computed state for the current input can be get through. The forget gate decide what information can be throwed away from the cell state. The output gate defines how much of the internal state we want to expose to the external network (higher layers and the next time step). Table B.2 describes step by step of LSTM Algorithm describes the step by step procedure of LSTM Algorithm.
By changing the value of f, i and o, LSTM can control the amount of information to get into the next layer as well as the dependency on individual inputs. They forget gate f is the key value for this architecture. If the forget gate is on (f is close to 1.0), the gradient does not vanish. Because f is never larger than 1.0, the gradient can not explore either. Therefore, LSTM can help prevent the vanishing and exploding gradient problems in a recurrent neural network.

Example of LSTM: Suppose we have a memory cell that have one unit and it can only take the value of 0 or 1 to denote the gender of a speaker in the text. Such as when the input is Tim that indicates a contextual change in the gender of the speaker. We first multiply the memory cell by 0 in order to flush everything that came before and then input a 1 into the memory cell to indicate that this is a male speaker for the next time step. This can be important for a predictive model because if the input is "Tim ran to", we want to use the fact that the speaker is currently male in order to output the correct prediction which is "his". When the model hits a input such as Sally it will want to then again flush the memory cell and write in a new value to reflect the speaker.

B.6 Tensorflow

B.6.1 Overview

Tensorflow is a library which was developed by Google Brain Team. Originally created for tasks that require heavy numerical computations, TensorFlow was geared towards the problem of machine learning, and especially deep neural networks. Since it was released as an open-source package under the Apache 2.0 license in November 2015, Tensorflow has become one of the most popular machine learning library.
The main idea behind TensorFlow is that the numeric computation is expressed as a computational graph or directed graph which has two basic units: graph nodes and graph edges. Graph nodes are functions/computations which have any number of inputs and outputs while graph edges are numbers, matrices or tensors which flow between nodes. Tensors are the standard way of representing data in Tensorflow. In practice we can simply define the tensors as n-dimensional matrix. Therefore, a 2 dimensional tensor is the same as a standard matrix. The advantage of using flow graphs as the backbone of deep learning framework is that it allows to build complex models in terms of small and simple operations. TensorFlow offers several advantages for an application. Thanks to a

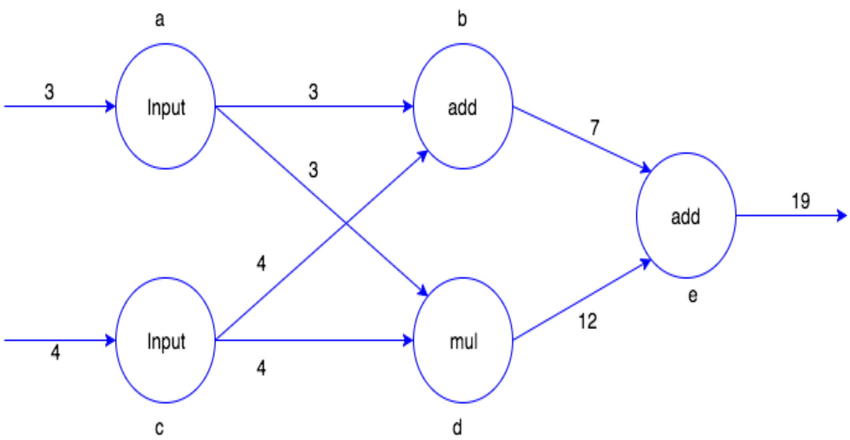


Figure B.7: Tensorflow Data flow graph example.

C C++ backend, TensorFlow is able to run faster than pure Python code. It provides both a Python and a C++ API. However, the Python API is more complete and it's generally easier to use. TensorFlow also has great compilation times in comparison to the alternative deep learning libraries. And it supports CPUs, GPUs, and even distributed processing in a cluster.

B.6.2 Install Tensorflow

Before getting started to exploiting Tensorflow, we need to install Tensorflow on the machine. Fortunately, there is a full, step-by-step tutorial to installing TensorFlow onto

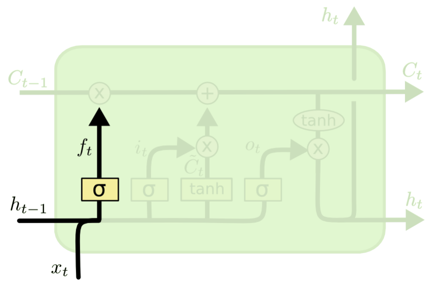
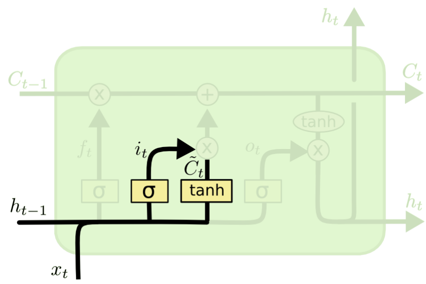
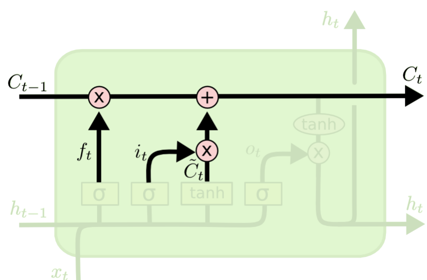
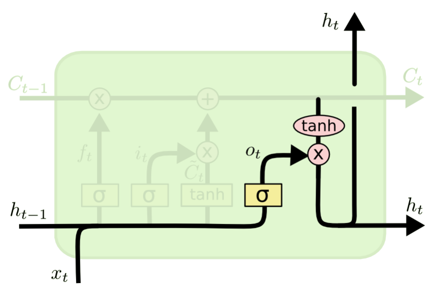
Step	Model	Description
1		At the first step, forget gate layer looks at h_{t-1} and x_t to output the number between 0 and 1 for each number in the previous cell state. $f_t = (W_f \bullet [h_{t-1}, x_t] + b_f)$
2		In the next step, the sigmoid layer called the input gate will define the values to update: $i_t = \delta(W_f \bullet [h_{t-1}, x_t] + b_i)$ After that, the tanh layer add a vector of new values t to the $\tilde{C}_t$ state: $\tilde{C}_t = tanh(W_C \bullet [h_{t-1}, x_t] + b_C)$ Then, we can define which new information is going to store in the cell state.
3		In this step, we multiply the old state by f_t to forget which have been decided before. Then we add the new prospect values $i_t * \tilde{C}_t$ which scaled how much have been decided to update the state value. At this step, we can update the cell state C_{t-1} into cell state C_t. $C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$
4		Finally, we run the sigmoid function to decide which part of the cell state can be outputted, then we put that cell state to tanh function and multiply it by the output of sigmoid gate. As the result, we have the output to the next layer. $o_t = (W_f \bullet [h_{t-1}, x_t] + b_o)$ $h_t = o_t * tanh(C_t))$

Table B.2: Step by step of LSTM Algorithm.

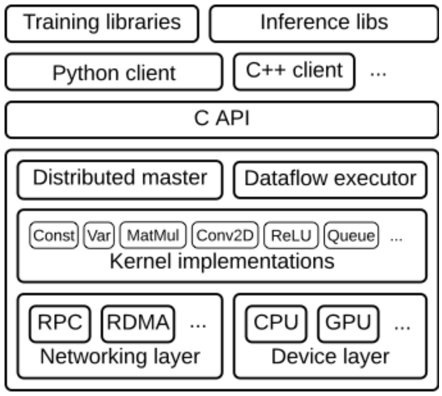


Figure B.8: Tensorflow Architechture. [120]

Linux, Mac OS and even Windows computers on Tensorflow official website [120]. On the scope of this research, we will implement the installation of Tensorflow on Linux environment by using virtualenv, a tool to create isolated Python environments. A folder which contains all the necessary executables to use the packages that a Python project would need will be created by virtualenv. During the installation process, not only TensorFlow but also all the packages that TensorFlow requires will be installed.

1. Install pip and virtualenv by issuing one of the following commands:

```
$ install python-pip python-dev python-virtualenv
# for Python 2.7
$ sudo apt-get install python3-pip python3-dev
  python-virtualenv
# for Python 3.n
```

2. Create a virtualenv environment

```
$ virtualenv --system-site-packages targetDirectory
# for Python 2.7
$ virtualenv --system-site-packages -p python3
  targetDirectory
# for Python 3.n
```

3. Activate the virtualenv environment

```
$ source ~/tensorflow/bin/activate
# bash, sh, ksh, or zsh
$ source ~/tensorflow/bin/activate.csh
# csh or tcsh
```

4. Install TensorFlow in the active virtualenv environment

```
(tensorflow)$ pip install --upgrade tensorflow
# for Python 2.7
(tensorflow)$ pip3 install --upgrade tensorflow
# for Python 3.n
(tensorflow)$ pip install --upgrade tensorflow-gpu
# for Python 2.7 and GPU
(tensorflow)$ pip3 install --upgrade tensorflow-gpu
# for Python 3.n and GPU
```

5. For testing TensorFlow Installation, we can use the command

```
$ python
# Python
import tensorflow as tf
hello = tf.constant('Hello, TensorFlow!')
sess = tf.Session()
print(sess.run(hello))
```

The output of the script should be

```
Hello, TensorFlow!
```

TensorFlow has been successfull installed on the system.

B.7 Conclusions and future works

In this report, we have finished a fundamental reason that is learning and researching about the architecture of RNN, especially the vanishing and exploding gradients problem of RNNs which leading to the need of LSTM. By using the special structure called cell, LSTM has shown its great success in long term learning algorithm. Moreover, we learnt about TensorFlow Framework and understood the runnable application based on TensorFlow. For the future works, which the basic understanding about RNN, LSTM, we will try to implement as a part of research topic "An Integrated Home-Based Health Care System".

In spite of the fact that the technology has been developed through time, there are still distance between the readings and the actual values of measuring parameters. Some researches show that although the advanced technology may perform well when tried in a controlled situation, it failed to get accuracy when apply in the real-life scenarios [121]. There are several reasons behind this uncertainty including internal inaccuracies of sensor operation, the effect of human activ- ity in sensor deployment locations and the impact of biological fouling upon the sensors over time [122]. The research at [123] showed that problems with more than 5% of missing samples may require sophisticated handling methods. Therefore, along with power consumption, context awareness, security, patient comfortable, the quality of monitoring data become a major concern in sensors deployments of the health monitoring systems. Many techniques were applied to collect high quality data, however in some studies, data was collected just in short period of time or need data quality was substandard.

Although LSTM is applied to successfully solve many problems of time series data such as speech recognition video summarization, music recognition; the technology is quite new in the missing-data imputation field. The research at [124] presented the first study to empirically evaluate the ability of LSTMs to recognize patterns in multivariate time series of clinical measurements. Therefore, it shows a great ability of applying LSTM with the improvement methods such as Kalman filters to impute the missing data of sensors in smart healthcare field.

Bibliography

[1] "United Nations Population Information Network (POPIN)," United Nations. [Online]. Available: http://www.un.org/popin/index.html. [Accessed: 08-May-2017].

[2] Mousavi, Saba, et al. "Depression, Chronic Diseases, and Decrements in Health: Results from the World Health Survey." The Lancet 370 (2007): 851-858.

[3] Holzinger, Andreas, Carsten Röcker, and Martina Ziefle. "From smart health to smart hospitals." Smart health. Springer International Publishing, 2015. 1-20.

[4] Ullah, Kaleem, Munam Ali Shah, and Sijing Zhang. "Effective ways to use Internet of Things in the field of medical and smart health care." Intelligent Systems Engineering (ICISE), 2016 International Conference on. IEEE, 2016.

[5] Manyika, J., et al. "Disruptive Technologies: Advances that will transform life, business, and the global economy, May 2013."

[6] López, Pablo, et al. "Survey of internet of things technologies for clinical environments." Advanced Information Networking and Applications Workshops (WAINA), 2013 27th International Conference on. IEEE, 2013.

[7] Andreu-Perez, Javier, et al. "From wearable sensors to smart implants - Toward pervasive and personalized healthcare." IEEE Transactions on Biomedical Engineering 62.12 (2015): 2750-2762.

[8] Baig, Mirza Mansoor, and Hamid Gholamhosseini. "Smart health monitoring systems: an overview of design and modeling." Journal of medical systems 37.2 (2013): 9898.

[9] Ravì, Daniele, et al. "Deep learning for health informatics." IEEE journal of biomedical and health informatics 21.1 (2017): 4-21.

[10] Flores, Jason R. "The Importance of Vital Signs." vital signs 3 (2016): 5.

[11] Hooker, Edmond A., et al. "Respiratory rates in emergency department patients." The Journal of emergency medicine 7.2 (1989): 129-132.

[12] Abbate, Stefano, Marco Avvenuti, and Janet Light. "Usability study of a wireless monitoring system among Alzheimer's disease elderly population." International journal of telemedicine and applications 2014 (2014): 7.

[13] Chakraborty, Suryadip, et al. "Detecting mobility for monitoring patients with Parkinson's disease at home using RSSI in a wireless sensor network." Procedia Computer Science 19 (2013): 956-961.

[14] Belgacem, Noureddine, and Fethi Bereksi-Reguig. "Bluetooth portable device for ECG and patient motion monitoring." Nature & Technology 4 (2011): 19.

[15] Henry, J. Craig. "Electroencephalography: basic principles, clinical applications, and related fields." Neurology 67.11 (2006): 2092-2092.

[16] Alberti, Kurt George Matthew Mayer, and PZ ft Zimmet. "Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus. Provisional report of a WHO consultation." Diabetic medicine 15.7 (1998): 539-553.

[17] Gilbert, Chris. "Biochemical aspects of breathing." Recognizing and Treating Breathing Disorders E-Book (2014): 61.

[18] Klepeis, Neil E., et al. "The National Human Activity Pattern Survey (NHAPS): a resource for assessing exposure to environmental pollutants." Journal of Exposure Science and Environmental Epidemiology 11.3 (2001): 231.

[19] Yang, Zhengguo, Azman Osman Lim, and Yasuo Tan. "Event-based home safety problem detection under the CPS home safety architecture." Consumer Electronics (GCCE), 2013 IEEE 2nd Global Conference on. IEEE, 2013.

[20] Bluyssen, Philomena M., et al. "Assessment of wellbeing in an indoor office environment." Building and environment 46.12 (2011): 2632-2640.

[21] Arif, Mohammed, et al. "Impact of indoor environmental quality on occupant well-being and comfort: A review of the literature." International Journal of Sustainable Built Environment 5.1 (2016): 1-11.

[22] Patel, Shyamal, et al. "A review of wearable sensors and systems with application in rehabilitation." Journal of neuroengineering and rehabilitation 9.1 (2012): 21.

[23] Gu, Yanying, Anthony Lo, and Ignas Niemegeers. "A survey of indoor positioning systems for wireless personal networks." IEEE Communications surveys & tutorials 11.1 (2009): 13-32.

[24] Patel, Shyamal, et al. "A review of wearable sensors and systems with application in rehabilitation." Journal of neuroengineering and rehabilitation 9.1 (2012): 21.

[25] Arampatzis, Th, John Lygeros, and Stamatis Manesis. "A survey of applications of wireless sensors and wireless sensor networks." Intelligent Control, 2005. Proceedings of the 2005 IEEE International Symposium on, Mediterrean Conference on Control and Automation. IEEE, 2005.

[26] Moore, G. "Cramming more components onto integrated circuits." Electronics 38 (8) (1965): 114–117.

[27] Baldini, Francesco. "Invasive sensors in medicine." Optical Chemical Sensors (2006): 417-435.

[28] Vashist, Sandeep Kumar. "Non-invasive glucose monitoring technology in diabetes management: A review." Analytica chimica acta 750 (2012): 16-27.

[29] Bonhomme, V., et al. "Physiological signal processing for individualized anti-nociception management during general anesthesia: a review." Yearbook of medical informatics 10.1 (2015): 95.

[30] Paradkar, Neeraj, and Shubhajit Roy Chowdhury. "Cardiac arrhythmia detection using photoplethysmography." Engineering in Medicine and Biology Society (EMBC), 2017 39th Annual International Conference of the IEEE. IEEE, 2017.

[31] Al-Naimi, Ibrahim, and Chi Biu Wong. "Indoor human detection and tracking using advanced smart floor." Information and Communication Systems (ICICS), 2017 8th International Conference on. IEEE, 2017.

[32] "Digital tools for next generation brain science." NeuroPro, www.neuropro.ch/.

[33] Xu, Sheng, et al. "Soft microfluidic assemblies of sensors, circuits, and radios for the skin." Science 344.6179 (2014): 70-74.

[34] Ng, Kim-Gau, et al. "Evaluation of the cadi thermosensor wireless skin-contact thermometer against ear and axillary temperatures in children." Journal of pediatric nursing 25.3 (2010): 176-186.

[35] "Health Monitoring Devices, Wearable Blood Pressure Monitor." IHealth, ihealthlabs.com.

[36] Ruedy, Katrina, et al. "Accuracy of the Dexcom (TM) G4 Platinum CGM System (R) during Six Months of Home Use." DIABETES. Vol. 66. 1701 N BEAUREGARD ST, ALEXANDRIA, VA 22311-1717 USA: AMER DIABETES ASSOC, 2017.

[37] "Sensorcon." Sensorcon - Sensing Products by Molex, sensorcon.com

[38] Razzaque, M. A., Muta Tah Hira, and Mukta Dira. "QoS in Body Area Networks: A Survey." ACM Transactions on Sensor Networks (TOSN) 13.3 (2017): 25.

[39] Onofrei, Elena, et al. "Textile sensor for heat flow measurements." Textile Research Journal 87.2 (2017): 165-174.

[40] Chen, Min, et al. "Body area networks: A survey." Mobile networks and applications 16.2 (2011): 171-193.

[41] Acampora, Giovanni, et al. "A survey on ambient intelligence in healthcare." Proceedings of the IEEE 101.12 (2013): 2470-2494.

[42] Alliance, ZigBee. "IEEE 802.15. 4, ZigBee standard." (2009).

[43] Alliance, ZigBee. "ZigBee wireless sensor applications for health, wellness and fitness," 2010."

[44] Herrala, Sami, Jari Kylmänen, and Jaakko Niemi. "Bluetooth networking." U.S. Patent No. 9,648,662. 9 May 2017.

[45] "Mobile Telephony Market". Bluetooth Special Interest Group. Retrieved January 16, 2014

[46] Gomez, Carles, Joaquim Oller, and Josep Paradells. "Overview and evaluation of bluetooth low energy: An emerging low-power wireless technology." Sensors 12.9 (2012): 11734-11753.

[47] Bradley, Peter D. "An ultra low power, high performance medical implant communication system (MICS) transceiver for implantable devices." Biomedical Circuits and Systems Conference, 2006. BioCAS 2006. IEEE. IEEE, 2006.

[48] Islam, Mohd Noor, and Mehmet R. Yuce. "Review of Medical Implant Communication System (MICS) Band and Network." ICT Express 2.4 (2016): 188-194.

[49] Chávez-Santiago, Raúl, et al. "Architecture of an ultra wideband wireless body area network for medical applications." Applied Sciences in Biomedical and Communication Technologies, 2009. ISABEL 2009. 2nd International Symposium on. IEEE, 2009.

[50] Sow, Daby, Deepak S. Turaga, and Michael Schmidt. "Mining of sensor data in healthcare: a survey." Managing and mining sensor data. Springer US, 2013. 459-504.

[51] Xie, Sai, and Zhe Chen. "Anomaly Detection and Redundancy Elimination of Big Sensor Data in Internet of Things." arXiv preprint arXiv:1703.03225 (2017).

[52] Tasnim, Samia, Niki Pissinou, and S. S. Iyengar. "A novel cleaning approach of environmental sensing data streams." Consumer Communications & Networking Conference (CCNC), 2017 14th IEEE Annual. IEEE, 2017.

[53] Mao, Yi, et al. "An integrated data mining approach to real-time clinical monitoring and deterioration warning." Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2012.

[54] Apiletti, Daniele, et al. "Real-time analysis of physiological data to support medical applications." IEEE transactions on information technology in biomedicine 13.3 (2009): 313-321.

[55] Hu, Fei, et al. "Robust medical ad hoc sensor networks (MASN) with wavelet-based ECG data mining." Ad Hoc Networks 6.7 (2008): 986-1012.

[56] Schmitt, Peter, Jonas Mandel, and Mickael Guedj. "A comparison of six methods for missing data imputation." Journal of Biometrics & Biostatistics 6.1 (2015): 1.

[57] García, Salvador, Julián Luengo, and Francisco Herrera. "Tutorial on practical tips of the most influential data preprocessing algorithms in data mining." Knowledge-Based Systems 98 (2016): 1-29.

[58] Luengo, Julián, Salvador García, and Francisco Herrera. "On the choice of the best imputation methods for missing values considering three groups of classification methods." Knowledge and information systems (2012): 1-32.

[59] Dempster, Arthur P., Nan M. Laird, and Donald B. Rubin. "Maximum likelihood from incomplete data via the EM algorithm." Journal of the royal statistical society. Series B (methodological) (1977): 1-38.

[60] Rubin, Donald B. Multiple imputation for nonresponse in surveys. Vol. 81. John Wiley & Sons, 2004.

[61] Batista, Gustavo EAPA, and Maria Carolina Monard. "An analysis of four missing data treatment methods for supervised learning." Applied artificial intelligence 17.5-6 (2003): 519-533.

[62] Brodley, Carla E., and Mark A. Friedl. "Identifying mislabeled training data." Journal of artificial intelligence research 11 (1999): 131-167.

[63] Chen, XiaoShuang, et al. "A novel under-sampling algorithm based on Iterative-Partitioning Filters for imbalanced classification." Automation Science and Engineering (CASE), 2016 IEEE International Conference on. IEEE, 2016.

[64] Chawla, Nitesh V., et al. "SMOTE: synthetic minority over-sampling technique." Journal of artificial intelligence research 16 (2002): 321-357.

[65] López, Victoria, et al. "An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics." Information Sciences 250 (2013): 113-141.

[66] Liu, Haiguang, and John CH Spence. "The indexing ambiguity in serial femtosecond crystallography (SFX) resolved using an expectation maximization algorithm." IUCrJ 1.6 (2014): 393-401.

[67] Noor, Norazian Mohamed, and Mohd Lutfi Zainudin. "A review: Missing values in environmental data sets." Proceeding of International Conference on Environment. 2008.

[68] Meliones, Apostolos, and Demetrios Sampson. "Indoor Blind Navigator: A Use Case for Self-Guided Tours in Museums." Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments. ACM, 2017.

[69] Gentili, Michele, Sara Hajian, and Carlos Castillo. "A Case Study of Anonymization of Medical Surveys." Proceedings of the 2017 International Conference on Digital Health. ACM, 2017.

[70] Chen, Liming, et al. "Sensor-based activity recognition." IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 42.6 (2012): 790-808.

[71] Rashidi, Parisa, et al. "Discovering activities to recognize and track in a smart environment." IEEE transactions on knowledge and data engineering 23.4 (2011): 527-539.

[72] Tukey, John W. "Exploratory data analysis." (1977): 2.

[73] Raza Perwez, Mohammad, Naveed Ahmad, and Muhammad Sajid Javaid. "A critical analysis on efficacy of clinical decision support systems in health care domain." Advanced Materials Research. Vol. 383. Trans Tech Publications, 2012.

[74] Yassin, Ali, et al. "Recent advances in indoor localization: A survey on theoretical approaches and applications." IEEE Communications Surveys & Tutorials 19.2 (2016): 1327-1346.

[75] Rajalakshmi, R. "Identifying Health Domain URLs using SVM." Proceedings of the Third International Symposium on Women in Computing and Informatics. ACM, 2015.

[76] Abid, Anam, et al. "Real time health monitoring of industrial machine using multiclass support vector machine." Control and Robotics Engineering (ICCRE), 2017 2nd International Conference on. IEEE, 2017.

[77] Miotto, Riccardo, et al. "Deep patient: An unsupervised representation to predict the future of patients from the electronic health records." Scientific reports 6 (2016): 26094.

[78] Nie, Liqiang, et al. "Disease inference from health-related questions via sparse deep learning." IEEE Transactions on Knowledge and Data Engineering 27.8 (2015): 2107-2119.

[79] Mehrabi, Saaed, et al. "Temporal pattern and association discovery of diagnosis codes using deep learning." Healthcare Informatics (ICHI), 2015 International Conference on. IEEE, 2015.

[80] Cortes, Corinna, and Vladimir Vapnik. "Support-vector networks." Machine learning 20.3 (1995): 273-297.

[81] Chen, Jiangyong, and Xianghan Zheng. "A System Architecture for Smart Health Services and Applications." International Workshop on Multi-disciplinary Trends in Artificial Intelligence. Springer International Publishing, 2015.

[82] Rawat, Kavita, and Kavita Burse. "A soft computing genetic-neuro fuzzy approach for data mining and its application to medical diagnosis." Int J Eng Adv Technol 3 (2013): 409-411.

[83] Santos, Flávio Arthur O., et al. "Smart health: Using fuzzy logic in the monitoring of health-related indicators." Telematics and Information Systems (EATIS), 2016 8th Euro American Conference on. IEEE, 2016.

[84] Ohri, Kalyani, Harsukhpreet Singh, and Anurag Sharma. "Fuzzy expert system for diagnosis of Breast Cancer." Wireless Communications, Signal Processing and Networking (WiSPNET), International Conference on. IEEE, 2016.

[85] Medsker, L. R., and L. C. Jain. "Recurrent neural networks." Design and Applications 5 (2001).

[86] Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory." Neural computation 9.8 (1997): 1735-1780.

[87] Edel, Marcus, and Enrico Köppe. "Binarized-BLSTM-RNN based Human Activity Recognition." Indoor Positioning and Indoor Navigation (IPIN), 2016 International Conference on. IEEE, 2016.

[88] Ruiz-Gonzalez, Ruben, et al. "An SVM-based classifier for estimating the state of various rotating components in agro-Industrial machinery with a vibration signal acquired from a single point on the machine chassis." Sensors 14.11 (2014): 20713-20735.

[89] Jia, Xiaowei, et al. "A novel semi-supervised deep learning framework for affective state recognition on eeg signals." Bioinformatics and Bioengineering (BIBE), 2014 IEEE International Conference on. IEEE, 2014.

[90] Yan, Yan, et al. "A restricted Boltzmann machine based two-lead electrocardiography classification." Wearable and Implantable Body Sensor Networks (BSN), 2015 IEEE 12th International Conference on. IEEE, 2015.

[91] Wang, Aosen, et al. "Selective and compressive sensing for energy-efficient implantable neural decoding." Biomedical Circuits and Systems Conference (BioCAS), 2015 IEEE. IEEE, 2015.

[92] Wulsin, Drausin, et al. "Semi-supervised anomaly detection for EEG waveforms using deep belief nets." Machine Learning and Applications (ICMLA), 2010 Ninth International Conference on. IEEE, 2010.

[93] Mao, Yi, et al. "An integrated data mining approach to real-time clinical monitoring and deterioration warning." Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2012.

[94] Awad, Yara Abu, et al. "A spatio-temporal prediction model based on support vector machine regression: Ambient Black Carbon in three New England States." Environmental Research 159 (2017): 427-434.

[95] Lita, Ioan, et al. "Indoor environmental parameters monitoring for building automation systems." Electronics, Computers and Artificial Intelligence (ECAI), 2016 8th International Conference on. IEEE, 2016.

[96] Koiwa, Yuichiro, Jun Miura, and Koki Nakagawa. "Human activity modeling and prediction for assisting appliance operations." Industrial Electronics Society, IECON 2015-41st Annual Conference of the IEEE. IEEE, 2015.

[97] Baig, Mirza Mansoor, and Hamid Gholamhosseini. "Smart health monitoring systems: an overview of design and modeling." Journal of medical systems 37.2 (2013): 9898.

[98] Rahman, Ashfaqur, Daniel V. Smith, and Greg Timms. "A novel machine learning approach toward quality assessment of sensor data." IEEE Sensors Journal 14.4 (2014): 1035-1047.

[99] Acuna, Edgar, and Caroline Rodriguez. "The treatment of missing values and its effect on classifier accuracy." Classification, clustering, and data mining applications (2004): 639-647.

[100] Yun, Xiaoping, et al. "Self-contained position tracking of human movement using small inertial/magnetic sensor modules." Robotics and Automation, 2007 IEEE International Conference on. IEEE, 2007.

[101] Jin, Yunye, et al. "A robust dead-reckoning pedestrian tracking system with low cost sensors." Pervasive Computing and Communications (PerCom), 2011 IEEE International Conference on. IEEE, 2011.

[102] Perera, Charith, Chi Harold Liu, and Srimal Jayawardena. "The emerging internet of things marketplace from an industrial perspective: A survey." IEEE Transactions on Emerging Topics in Computing 3.4 (2015): 585-598.

[103] Islam, SM Riazul, et al. "The internet of things for health care: a comprehensive survey." IEEE Access 3 (2015): 678-708.

[104] E4 Wristband User's Manual 20150608., Empatica, Milano, Italy, 2015, pp. 5-16. Taken from https://www.empatica.com

[105] McCarthy, Cameron, et al. "Validation of the Empatica E4 wristband." Student Conference (ISC), 2016 IEEE EMBS International. IEEE, 2016.

[106] Jahoda, Philipp. "MPAndroidChart." Accessed January 20 (2018).

[107] Singh, Navdeep. "Study of Google Firebase API for Android." International Journal of Innovative Research in Computer and Communication Engineering 4.9 (2016): 16738-16743.

[108] E. Law, "Impyute: Library of the different imputation algorithms; methods for dealing with ambiguity and handling missing data." 2017.

[109] Downie, Nick. "Chart.js— Open source HTML5 Charts for your website." Chart.js (2017).

[110] Yassin, Ali, et al. "Recent advances in indoor localization: A survey on theoretical approaches and applications." IEEE Communications Surveys & Tutorials 19.2 (2016): 1327-1346.

[111] Zachary C Lipton, David C Kale, Charles Elkan, and Randall Wetzell. "Learning to diagnose with LSTM recurrent neural networks". ICLR, 2016.

[112] Cook, Diane J., et al. "CASAS: A smart home in a box." Computer 46.7 (2013): 62-69.

[113] Liang, Po-Chou, and Paul Krause. "Smartphone-based real-time indoor location tracking with 1-m precision." IEEE journal of biomedical and health informatics 20.3 (2016): 756-762.

[114] "Recurrent Neural Networks Tutorial, Part 1 – Introduction to RNNs," WildML, 08-Jul-2016. [Online]. Available: http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns/. [Accessed: 31-May-2017].

[115] "Machine Learning - Stanford University — Coursera." [Online]. Available: https://www.coursera.org/learn/machine-learning/. [Accessed: 31-May-2017].

[116] "Neural Networks for Machine Learning," Coursera. [Online]. Available: https://www.coursera.org/learn/neural-networks. [Accessed: 31-May-2017].

[117] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[118] "Understanding LSTM Networks," Understanding LSTM Networks – colah's blog. [Online]. Available: https://colah.github.io/posts/2015-08-Understanding-LSTMs/. [Accessed: 31-May-2017].

[119] R. Atienza, "LSTM by Example using Tensorflow – Towards Data Science – Medium," Medium, 17-Mar-2017. [Online]. Available: https://medium.com/towards-data-science/lstm-by-example-using-tensorflow-feb0c1968537. [Accessed: 03-Aug-2017].

[120] R. Atienza, “LSTM by Example using Tensorflow – Towards Data Science – Medium.” Medium, 17-Mar-2017. [Online]. Available: https://medium.com/towards-data-science/lstm-by-example-using-tensorflow-feb0c1968537. [Accessed: 06-Aug-2017].

[121] Baig, Mirza Mansoor, and Hamid Gholamhosseini. “Smart health monitoring systems: an overview of design and modeling.” Journal of medical systems 37.2 (2013): 9898.

[122] Rahman, Ashfaqur, Daniel V. Smith, and Greg Timms. “A novel machine learning approach toward quality assessment of sensor data.” IEEE Sensors Journal 14.4 (2014): 1035-1047.

[123] Acuna, Edgar, and Caroline Rodriguez. “The treatment of missing values and its effect on classifier accuracy.” Classification, clustering, and data mining applications (2004): 639-647.

[124] Zachary C Lipton, David C Kale, Charles Elkan, and Randall Wetzell. “Learning to diagnose with LSTM recurrent neural networks”. ICLR, 2016.

Publications

[1] Cu PHAM, Tan LE, Yuto LIM, Yasuo TAN, "An Architecture for Supporting RAS on Linux-based IoT Gateways", IEEE 6th Global Conference on Consumer Electronics (IEEE GCCE 2017 OS-ICE), Nagoya, October 2017.

Awards

[1] Tan LE, "JAIST President Award 2016", March 2017.