Health Care Expert System using IOT (Internet of things)

MS COMPUTER SCIENCE

Asra Anees

MSCS-F17-013

Session: Fall [2017 – 2019]

DEPARTMENT OF COMPUTER SCIENCE

INFORMATION TECHNOLOGY THE SUPERIOR COLLEGE

LAHORE, PAKISTAN

Health Care Expert System using IOT (Internet of things)

A thesis submitted in completely fulfillment the requirements for the

Degree of Master of Science in

Computer Science

Student Name: Asra Anees

Supervisor Name: Dr. Arfan Jaffer

DECLARATION

This thesis is a presentation of my original research work. Wherever contributions of others are involved, every effort is made to indicate this clearly, with due reference to the literature, and acknowledgement of collaborative research and discussions. I also declare that this work is the result of my own investigations, except where identified by references and free from plagiarism of the work of others.

Signature: ………….

Asra Anees

Date: ……………….

The undersigned hereby certify that they have read and recommend the thesis entitled “Health Care Expert System using IOT (Internet of things)” by Asra Anees for the degree of Master of Science in Computer Science.

____

Dr. Arfan Jaffer, Thesis Supervisor

____

, Thesis Committee Member

____

, Thesis Committee Member

____

, Head of Department

ACKNOWLEDGEMENTS

The success and final outcome of this thesis project required a lot of guidance and assistance from many people and I'm extremely fortunate to have got this all along the completion of my thesis work. Whatever I have done is only due to such guidance and assistance and I would not forget to thank them.

First, I would like to thank Allah for facilitating the obstacles that I have been through during my master degree. Secondly, I would like to express my gratitude towards my family & friends for their kind co-operation and encouragement which helped me to complete this thesis.

I would like to send special thanks for Dr. Arfan Jaffer for the support during this project, I would not pass it without their help and support. Last not least, thanks for everyone's help, support, or motivate me either by word action, or even emotion.

Thank you,

Asra Anees

DEDICATION

I would like to dedicate my thesis to my beloved teacher and friends. A special feeling of gratitude to my loving parent whose words of encouragement.

I also dedicate this thesis to my brothers and sisters specially who have supported me throughout the process. I will always appreciate all what they have done, especially for helping me develop my technology skills.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Today, people have connected through the internet anywhere, anyplace and anytime. The innovation of facilities like smart homes and cities, healthcare application and much more are connected through the network of heterogeneous IoT (internet of things) include smart devices, smart sensor, etc. and enable M2M (machine to machine) S2M (sensor to machine) M2D (machine to devise) and S2D (sensor to device) communications. The connection is built by using the devices, smart IoT objects, and end-users.

As the demand for IoT increases day by day the health care sectors gain more attention some chronic diseases are needed permanent monitoring. These diseases consist of several infirmities in the form of high blood pressure, diabetes, high cholesterol, Obesity, smoking, mental health illness, cardiovascular diseases, chronic lung, cancers, and stroke disease. Mainly these diseases lead to elderly peoples and dealing with this disease is a big challenge in the development and undeveloped areas. The person suffered from these diseases need frequently monitoring but the staff in the hospital are much less than the number of patients need regular checkup and treatment so there is the possibility of error in the monitoring of any patient. Due to the lack of resources in the hospital, our proposed system provides the facility to the patient for remote monitoring from home especially for the elderly and disabled peoples. This new technology enables repeated monitoring on regular bases from home.

In past years, the e-health system developed which deal with the massive amount of patient records and data. The main target is to make an expert system using machine learning technique and IOT devises such as sensors, which automatically classified, predict sensitive data and provides the proper and complete information to medical professionals. This health care management system design to combine patient monitoring devices with low cost and this technology mainly used for continuous monitoring of patients. The objective includes continue health monitoring of the patient, life-save emergency report, early-stage diagnose and more physician can diagnose and decide, view medication side effects, errors and prescribe accordingly.

The Artificial intelligence technique and algorithms use to enhance the feature of data. In the health care department, there is a massive amount of information stored in the electronic health care system. The aim is to apply the data mining technique on multiple diseases for data classification, predict the accuracy and error between medical data. The three classifiers J48 of decision tree algorithm, MLP (Multi-Layer-Perception) and NB (Naïve-Bayes) used with the Confusion sets based on 20, 50, 80-fold cross-validation method. The results conducted on diabetes, Brest cancer, iris, hypothyroid, and dermatology data sets. Data mining tool WEKA is used to conduct all the tests. Finally, we perform an experiment on disease data set by using some classifiers and predict the result.

CHAPTER 1

INTRODUCTION AND BACKGROUND

The lesson presents the background and introduction of IoT based health care expert systems in terms of scientific research. Moreover, we discuss the objectives of this thesis and our main contribution, structure and the environment used for this work.

1.1 Background

Internet of Things technologies can be changing the world in terms of life, drive, impart, and so on as we probably now it. The IoT can be express in two ways: one is Internet state as networks of networks the billions of users connect to it through some standard protocols. Various devices i.e. cell phones, personal computers, and business industries are connected via the internet. The second term is related to that these devices are intelligent [1]. The ability of the network device is to sense and gather data from the world and share them across the internet or become a smart object. With the invention of IoT based remote digital clinical system. The transaction of the medical record is done on a daily basis [2].

One of the most attractive and effective focusing areas of IoT based application is health and medical care because healthcare is important for all human beings. An IoT healthcare system has potential interconnected the various device to preform healthcare activities i.e. monitoring, diagnosing, and remote surgery, remote health monitoring, and chronic illness through the internet [3]. Treatment and medication are provided at home is another advantage. Moreover, several medical, imaging devices and body sensors can consider as a smart device help to establish the core IOT based health care intelligent system. The data in EHR used for the prediction that helps to give medical treatment properly at an early stage which gives a chance to make better human health [4].

IoT deals with the challenge faced by the health care department by its smart application and services. It has the ability to connect machine to machine (M2M), the device to machine (D2M), physician to machines, physician to devices, patient to machines, patient to the device, etc. By interconnecting the intelligent device such as a human, smart device, machine, objects and automated system to make sure a healthcare system is effective.

Actually, healthcare is more important for people. And this makes easy to examine for the doctor to examine the record of a patient. It can be observed that the workload of the doctor is rise due to rush and lake of resources in hospitals. It is important to provide better healthcare services to needy people. Hence it is important that data is properly classified in the various domains helps correctly diagnose and prescribe although it's really an important task. Diagnose and treatment process is based on patient symptoms, physiological condition and biographic include age, gender, medical history, and so on. As the world becomes an automated healthcare system also become more advanced and automated by the use of Electronic health record (EHR) [5] which was publicized in united states at 2010 [4] in which record of electronic healthcare present and need to improve healthcare sectors. Big Data is store EHR should be classified that make easy to use and secure from unauthorized access and not be used in an illegal way. The IT field playing a major role in the healthcare field by providing health care management services.

As some disease permanently monitoring. the proposed system allows real time health to examine. The purpose of short term and long-term analysis and diagnosis. The system is prepared with smart devices, different sensors are used to save life features such as reporting an emergency health situation directly to medical personnel and automated calling for an ambulance. The study focuses on collected data from a different sensor or takes the standard data sets. The then pre-process the data in some manner such as remove missing values, noise, outliers, and integrate of data store in EHR (Electronic health record).

Now a day, a heterogeneous amount of data captures automatically from patients and store in EHR (electronic health record). The healthcare system includes demographic systems, laboratories, radiology systems, etc., Data of these systems increase rapidly day by day and save in its raw form. Noisy data, missing values, outliers, and unexpected information are needed to handled properly. Data mining tool used in manage disease, data prevention, diagnosis and medicine prescription which reduce the risk of a medication error, safe human life because improper monitoring and medication error are a major health issue and may cause lead to death.

The intelligent or expert system design by using some automated technique i.e. machine learning and IoT devices included the sensor. Data is stored in EHR connected to e-health users such as the doctor, patient and third party. This goal is to classify medical data, stored in electronic health care (EHR). Used in various ways include continue health monitoring of the patient, save the patient lives in an emergency, emergency alert and report is generated, diagnose the chronic disease at an early stage and more doctor can diagnose and decide, view medication side effect, errors and prescribe accordingly. The main target is the classification and prediction of sensitive data and provides the proper and complete information to medical professionals.

1.2 Problem Statement

With the enhancement of technology, it required lots of effort to keep things update. Same situation in the medical and healthcare sector. Thus, it needs to merge with technology. IoT plays a vital role in the medical field. IoT devices embedded with healthcare sector which perform some daily tasks to reduce the workload of physicians and healthcare staff. The issues arise when the data is collected from smart sensors device must properly classify, aggregate and analysis because without maintenance data handling is very hard. As a classification problem to collect the data for a particular disease based on different attributed. At the stage of data preprocessing such type of error i.e. missing information, noise between transmission of data and integration of data, remove data duplication is done and keep all the information save in EHR (Electronic health record). EHR arrange in terms of different features. After the classification data can be prediction is based on trained data and prediction error identification is based on target values.

Therefore, in this project we will propose a framework where we can manage absent and strange data. Although, we propose a generalized framework that will work with any hospital structure. We will focus on the some very common disease. We need to manufacture solution that will make diseases prediction simpler than any time in recent memory. Some of people are dying for the absence of early notification due to the lack of the tracking the performance of the illness.

We observe that there are a few limitations like many proposed systems doesn't manage absent and unusual data. Therefore, we discovered issues and difficulties by reading various papers and by understanding their proposed model. From the issue of illnesses expectation and investigation, we firmly believe that we should contribute around there. There are a great many individuals over the world experiencing distinctive ailments just for absence of early identification system. People are occupied with their regular daily existence, so they don't get enough time to go for general checkup. In this way, we attempt to present a more quick and simpler approach to counsel with the doctor.

we will start with our proposed framework. Then, we will introduce a framework with manage missing data. From this framework, we can apply it on any healthcare sector. Additionally, we will replace any unusual values if occur in the data.

1.3 Research Question and Hypothesis

The query is: What's the suitable structure applied to classified healthcare datasets?

What types of attributes are taken? In which feature data in present? What sensors are used to detect the information and is determine the sensors should be working properly?

We will answer this question raised during this project. What is the minimum requirement applying to the system in the hospital?

Finally, we are going to answer these questions in detail in the following sections of this project.

1.4 Research Objectives

The objectives for this project will mainly focus on two essential objects.

First one, to design a framework that should be suitable to apply it on the infrastructure in most of the hospitals.

Second, to analyze the dataset of the diseases, investigate the results like missing values, noise between transmission and make sure there are no errors in gathering the data.

1.5 Scope of the study

The study will focus on building a healthcare framework and analyzing the collected data form the diseases.

Although, we are going to gather data and mining them to get the result that supposed to be done with the framework scope.

1.6 Research methodology

Our proposed method is based on automatic classification and prediction of data gathered from different sensors by using some machine learning algorithms or tools WEKA. We take the structure data set or we generate our data set from smart sensors. We design some structure and framework especially for the healthcare sector to make people live better.

Step 1: Collect data for various IOT sensors, data may be structured or unstructured form.

Step 2: Preprocessing is done to identify that proper information is available or there are some missing values. More transaction is done properly or there some noise and outlier error. In additional is integrate the data in various feature and save in EHR (Electronic health record).

Step3: Feature detection means that in which form data is present i.e. in images, number, and alphanumeric form etc., and remove the unwanted data.

Step 4: After the data collection, data preprocessing and feature detection/reduction is done finally the data is classified and send for prediction.

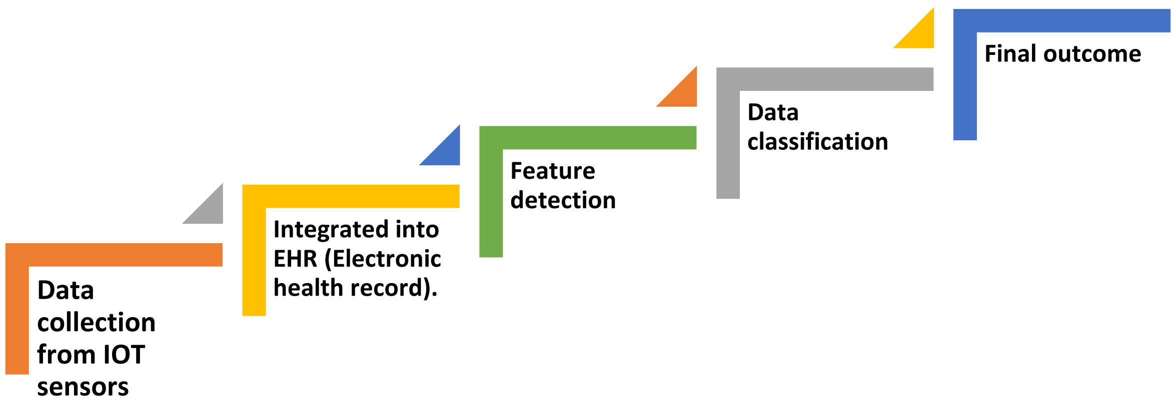


Figure 1: Proposed method of IoT Frame work

The main goal is for establishing a system that automatically classified data and also gathers information about patient health. And send accurate information. Based on that information Medication/ and the prescription is given. And alert the physician if an emergency occurs. The treatment is started by examining such objects like age, gender, medical history and stage of the disease. Improper monitoring and medication error are a major health issue and may cause lead to death. The benefit of technology helps to reduce medication error.

1.7 Used Environment

WEKA tool is used for data processing and visualization.

1.7.1 Data mining tool (wake)

The University of Waikato established the data mining tool named as WEKA (Waikato-Environment-for-Knowledge-Analysis). The 3.8 version of the WEKA is used. Practice indicates that there is no any other ML scheme is applicable to solve all problems of DM. The idealize phenomenon of worldwide learners is that actual datasets differ, and to gain correct information in the preference of the training algorithm should meet with the requirement of domain structure. That’s why it is an experimental art [6]. The WEKA tool is a collection of ML procedures and provides a tool for data preprocessing. It's a flexible tool in which you can apply an existing method for new data sets. It gives us general support for the entire procedure of experimental DM which includes data preparing, statistically evaluating learning systems, and data visualizing and the Learning result. Moreover, it includes a various variety of learning methods and preprocessing tools. This modern toolkit retrieves through a familiar interface. That's why users can easily do a comparison between different algorithms and classify individuals that are the best fitting for the real-time problem.

1.8 Organization of the project.

This section goal is to give a briefly overview of the following chapters in this project.

Chapter 2: covers all the studies that has been in this subject. We cover some of the topics that are clusters into four categories: Frameworks, Healthcare systems, Challenges, and advantages of internet of things, and Communications technology, and also about different medical datasets.

Chapter 3: In this chapter we go over the problem system and analyze the result of the diseases. Although, present the proposal framework, including the details regarding the infrastructure.

Chapter 4: This Chapter presents the detail about datasets and for datamining algorithms.

Chapter 5 & 6: Will cover the result and the discussion for the diseases. We will show some of the tables, and some figures refer to the results of the diseases.

Chapter 7: Conclusion for this project, lessons learned, and finally our recommendation to apply the proposal framework.

CHAPTER 2

LITERATURE REVIEW

This chapter cover all resent research work done about the topic “IOT based health care expert system”. However, this chapter is divided into sub-parts to cover all major areas relates to the topic in a large amount of information. Thus, the sub-parts are Framework which describes all resent frameworks that fit in the healthcare system. The healthcare system explains the availability of systems healthcare sectors, medication error, and missing values. Lastly, describes the drawback and benefits of implanting the Internet of things in medical departments.

2.1 Frameworks

In [7] the article is about introducing a new framework should be applied for vehicular networking application (VNA). This article is focusing on the tools that shall be used for applying a framework or business model for internet of thing to be use in a vehicle. Therefore, the author contribution in this article restricted in the following points:

1. A novel PPS business model and service architecture in vehicular networking are proposed, and this scheme is applied to VMS.
2. The inter-cloud computing architecture supporting telematics applications and inter-cloud service model including semantic modeling are designed.
3. The proposed VMS event processing flow can allow the management and cooperation among diverse components, ensure the communications to users in case of emergency, and enhance the detection of anomalous event.
4. The performance of our proposed PPS scheme against prevalent schemes is evaluated on the basis of probabilistic methods.” [7]

The another interduce four layers architecture and application domains, as following:

a) Object sensing and information gathering.
b) Information delivering.
c) Information processing.
d) Application and smart services.

Although, the author mentions in his article the platform production services (PPS), the platform production services have a scheme where it is has three types are listed as following:

1. **Digital platform**: The digital platform possessing some preliminary functions such as navigation and entertainment is a complicated embedded system installed in vehicle.
2. **Production**: After designing the digital platform, equipment suppliers usually set up some necessary applications
3. **Services**: Through implementation concept PPS business model, the users will achieve more emerging services.

The author, brings a critical issue that will happen while applying the business module, which is stick with one single cloud system. The author aware from the crises that may happen with this provided. Therefore, the service will not be able to be reached. However, the author suggests having another or let us say back up provide in case the primary provide has an issue. This action will save a lot of money and cost for the services provider.

Eventually, Because of the requirement for enhancing comfort, effectiveness, and security in VNA, this article considered the practicality of PPS plan of action by IoT detecting innovation and between distributed computing. In the wake of proposing an inventive PPS plan of action in vehicular organizing, we connected this plan to VMS. Likewise, we outlined the between cloud engineering supporting telematics applications, between cloud benefit show including semantic demonstrating examination and occasion preparing stream for dealing with the dependable messages. For the proposed plot, the exhibitions, for example, effectiveness and hazard factor, are assessed by embracing the probabilistic hypothesis.

In [8] the article talks about the methodology for internet of things. The author mention that the elder people don't have enough resource to be used to help them to keep themselves survived. The author brings us to the point of problem statement of the internet of things with the methodologies for the healthcare systems. the author addresses the problem for the methodologies by expanding the two part the design process which are problem formulation, and product development. The problem formulation stage consists of three steps: need definition, requirement formulation and feasibility assessment. case study of building a hart tracking in a person is consider in this article. However, modeling is one of the topics of this article where the author uses some mathematic to be used for the accelerometer that will be used for the case study. Therefore, this article is about getting an IOT device to be used by a patient and try to track the hart performance while daily activity and get the result by using the internet of things devices.

In [9] the purpose is to encourage future business inquire about by proposing a novel approach to consolidate plans of action as a reasonable instrument with fates examine methods. However, A future point of view is received to foreknowledge plans of action of the Web of Things (IoT) empowered social insurance division by utilizing plans of action as a fates business inquire about device. In doing as such, business show is combined with a standout amongst the most conspicuous foreknowledge procedure, Causal Layered Analysis CLA). Subjective examination gives further comprehension of the marvel through the layers of CLA; litany, social causes, worldview, and myth. As a result of this article, it is hard to foresee the far future for an innovation arranged area like human services. This paper presents three situations for short-, medium-and long-haul future. In view of these situations we likewise present an arrangement of plan of action components for various future time spans. This paper demonstrates a way to consolidate plans of action with CLA, a prescience procedure; keeping in mind the end goal to apply plans of action in prospects business inquire about. Other than offering early outcomes for prospects business consider, this investigation proposes a theoretical space to work with singular plans of action for administrative partners.

In [10] talks about the disability people. They are over a billion over the world. It is about 15% of the total of the people in the earth. Disabled people have a lack of support that can let them live normally. Therefore, they need their family or friends to help them to do some of their daily tasks. Internet of things is the best solution for this category of people. In this people the author is going to propose an architecture of the inter net of things in order to help the lovely disabled people. We will be going through different scenarios.

The author proposes the architecture for the internet of things for disabled people in three layers as the following:

Perception layer: its main function is to identify objects and gather information (devices). The components of this layer are:

Visually impaired.

RFID-based assistive devices.

Hearing impaired.

Assistive devices and sensors.

RFID-based devices.

Physically impaired.

Body sensors

Body sensors and RFID technology

Network layer: it consists of a converged network made up of wired/wireless privately owned networks.

Application layer: it is a set of intelligent solutions that apply the Internet of things technology.

The author gives an example of the application scenarios such as:

At store:

Visual weakness people have a hard time doing shopping, because they have a people with the paths in the store, as well as finding the products. Therefore, blind navigation system helps them to their way to the store. The store is divided into cells containing shelfs. RFID tags are distributed in the floor of the store, this can lead the visual weakness people to their desired cell to get the wanted product. However, the RFID is connecting with the user by Bluetooth to keep tracking the user where the steps are going.

At school:

Internet of things can be used in the school to keep tracking the toddler where they are hiding. Moreover, the author prefers to use the RFID with their toys, since the toddler are addicted with their toys. Therefore, this might be a good lead to find the toddler in case we lost him.

Domestic environment:

The author keeps giving brilliant ideas for the internet of things. The author talks about the smart home, where people can use the internet of things in their refrigerator, microwave, Air condition. All these machines can be easier to use with the internet of things. Do not forget, we are still using the proposed architecture for the internet of things in the beginning, and that could be apply for one of the given examples.

In this [11] paper plots an Internet of Things (IoT)- based community-oriented structure which gives an establishment for digital physical communications and coordinated efforts for innovative fabricating spaces. A general system for community-oriented fabricating is proposed trailed by a talk of such an IoT-based structure for the space of miniaturized scale gadgets gathering. The outline of this synergistic system is talked about with regards to distributed computing and additionally the rising Next Internet which is the concentration of late activities in the USA, EU, and different nations. The information/data trade between the different programming and physical parts is demonstrated utilizing the building Enterprise Modeling Language (eEML), which gives an organized establishment to planning and building up this IoT-based community-oriented system. The key digital physical parts and modules are portrayed took after by a talk of the usage of this structure.

Table 1: Related work done on framework

Author	**Journal**	**Year**	**Title**	**Methods**	**Result**	**Comp…**
Chen, Y. Y., Lu, J. C., & Jan, J. K.	Journal of medical systems	2012	A secure EHR system based on hybrid clouds [5]	Implement both the private cloud in a hospital and the public cloud	No	Yes
J. Wan, C. Zou, K. Zhou, R. Lu, and D. Li,	Applied Sciences	2017	IoT sensing framework with inter-cloud computing capability in vehicular networking. [7]	The Internet of Things (IoT)-based information system for indoor and outdoor use. Use the decision trees algorithm.	No	No
D. Dziak, B. Jachimczyk, and W. J. Kulesza	Appl. Sci	2017	IOT-Based Information System for Healthcare Application: Design Methodology Approach [8]	machine learning techniques, the decision trees II, KNN I and KNN II	classification validity of 94.8%, 96.0% And 95.8%, respectively.	yes
J. F. Gomes and S. Moqaddemerad,	J. Bus. Models,	2016	Futures Business Models for an IoT Enabled Healthcare Sector: A Causal Layered Analysis Perspective [9]	Causal Layered Analysis CLA applies with in three situations for short-, medium-and long-haul.	No.	No.
M. C. Domingo	J. Netw. Comput. Appl	2012	An overview of the Internet of Things for people with disabilities. [10]	Proposes the architecture for the internet of things for disabled people in three layers Perception, Network and Application.	No	No
Y. Lu and J. Cecil	Int. J. Adv. Manuf. Technol	2016	An Internet of Things (IoT)-based collaborative framework for advanced manufacturing. [11]	Internet of Things (IoT)- based community-oriented structure, Enterprise Modeling Language (eEML),	No	No
K. Suresh and M. R. Babu	Journal of Biomimetics, Biomaterials and Biomedical Engineering	2016	Emerging Biomedical Health Care System by Using Internet of Things,"	The paper depicts the relationship and effort of the patients and doctors in a figure, where it has three elements; doctor, patients, and MADC Medical Digital Assistance Center.	No	No.

S. M. R. Islam, D. Kwak, M. H. Kabir, M. Hossain, and K. Kwak	IEEE Access	2015	The Internet of Things for Health Care: A Comprehensive Survey. [13]	The IOT healthcare network (IOThNet) considers topology, architecture, and platform. IoT based health care Service include WDA, CH, m-IOT, CHI, and many more	No	yes

2.2 Healthcare System

In [12] the paper explains the necessities and significance of the Bio restorative human services framework developing with Internet of Things. Are basic to consider with innovation necessities. Demonstrated reenactment condition to be diverse gadgets introducing. In this paper what is the barrenness of bio healing medicinal services in different nations. Particularly in this situation taken thought and centering by the Indian voting public. Talked about what are the incompetent biomedical field adding with advanced innovations to Simplest approach to lessen persistent holding up time. Prognotive Computing is the extraordinarily required survival reference for Mankind. With the current imaginative capacity what's all the more, Big Data Analytics never has all the earmarks of being considerably nearer to fulfilling. If we examine this period being the Processing Era, for what reason not a specific enrolling methodology in the field of Human administrations and Solutions. With honest to goodness financing and sponsorship of associations and Government concerns Computational Prognosis will outline the destiny of the World. Generally finished setup is full fill the connection enhances more emotional way. The main benefits from this paper to make more time for the doctors by letting the internet of things devices keep tracking the patients. Therefore, the job has be done by the doctor is only to read these reports and give the patients the convenient prescription, rather than visiting the doctor three times a week. This paper depicts the relationship and effort of the patients and doctors in a figure, where it has three elements; doctor, patients, and MADC Medical Digital Assistance Center. patients have the big portion where he or she has to do the big part of this participation, and doctor and MADC have the same medium portion of effort.

In [13] authors writer tries to cover areas linked to IoT technology in healthcare department. IoT healthcare network (IOThNet) considers architecture, platform, topology, and service include WDA (Wearable Device Access), AAL (Ambient Assisted Living), SMA (Semantic Medical Access), CH (Community Healthcare), m-IoT (Internet of m-Health-things), CHI (Children-Health-Information), ECP (Embedded Context Prediction), IEH (Indirect Emergency Healthcare), EGC (Embedded-Gateway-Configuration) and ADR (Adverse-Drug-Reaction). An application such as Temperature, Oxygen Saturation, Electrocardiogram, and Blood Pressure checking, Management of medication on time and wheelchair as needed, Scanning the glucose level faster response on pending or close up details by using smartphone. This paper also describes the model to decrease the security risk and determine how innovation i.e. ambient intelligence, big data, leveraged in the medical background. And discuss how e-health and IOT are facilitating societies in supportable development. The authors also talk about the policies in some countries around the world in deep. However, this paper has a figure contains a brief about them as follows:

Country/region	eHealth strategy (year published)
Australia	National eHealth strategy (2008)
Australia	State eHealth strategy-Queensland (2006)
Denmark	National IT strategy 2003-2007 for the Danish healthcare service (2003)
The European Commission, DG for Information Society and Media, ICT for Health	European countries on their journey towards national eHealth infrastructures, evidence on progress and recommendations for cooperation actions (2011)
The European commission, DG for Information Society and Media, ICT for Health	eHealth priorities and strategies in European countries (2007)
The European Commission	Repository of eHealth strategies and priorities for EU member states (N/A)
Finland	eHealth roadmap – Finland (2007)
Kenya	National eHealth strategy (2011)
Mauritius	National eHealth Strategy: He@lth2015, seamless continuity of care (2010)
Saudi Arabia	National eHealth strategy (2011)
Scotland	National eHealth strategy (2011)
Sweden	National strategy for eHealth (2006)
Switzerland	Swiss eHealth strategy (2007)
The U.S.	Federal health IT strategic plan (2011)

Figure 2: E-health strategy [13]

2.3 Medication Error

The [14] Paper published in 2010 by U.S. Pharmacopeia (USP) to promote safe medication used in a physical environment. Describe that five major types of medication errors including prescription (39%), transcription (12%), monitoring (38%) and dispensing (11%). According to the report from US 9.5%, people die due to medical errors. Incorrect or improper monitoring cause to medication error and patient health become the risk of missing dose, wrong medication. Because of medication error death rate is increasing day by day [15].

The [16] the author describes the importance of medication adherence in elderly people. The issues are associated with the factors include medication, patient, healthcare system, socioeconomic and healthcare providers. so, it is important to keep in mind these issues while treating the patient.

The [17] review paper present's about medication error in 36 health care sectors. Result shows, that in 36 health sectors 19% of error in medication doses. 43% medication giving on wrong time, omission 30%, 17 % is wrong medication and unauthorized drug was 4%. It results that 7% of error were potentially very harmful.

Author	**Journal**	**Year**	**Title**	**Methods**	**Result**
M. Grissinger	*Pharm. Ther*	2012	Physical Environments That Promote	Pharmacopeia (USP) to promote safe	Prescription (39%), transcription (12%), monitoring

			Safe Medication Use [14]	medication used in physical environment	(38%) and dispensing (11%)
J. Hopkings Medicine	*John Jopkins Med.-News Publ*	2016	Study Suggests Medical Errors Now Third Leading Cause of Death in the U.S. [15]	Death ratio increase because of medication error case in hart and cancer diseases.	US 9.5%, people die due to the medical error
F. Yap, T. Thirumoorthy, and Y. H. Kwan	*J. Clin. Gerontol. Geriatr*	2016	Medication adherence in the elderly [16]	Medication error can effect elderly people, it's important to prescribe pervert medication to elderly people.	--
K. N. Barker, E. A. Flynn, G. A. Pepper, D. W. Bates, and R. L. Mikeal,	*Arch. Intern. Med.*	2002	Medication Errors Observed in 36 Health Care Facilities [17]	present's about medication error in 36 health care sectors	19% error in doses. 43% wrong time, omission 30%, 17 % wrong medication and unauthorized drug was 4%.

Table 2: Medications Error Results

2.4 Data Analysis

In [18], the authors analyze cloud based solution for EHR's (electronic health records) systems. Different solutions and scenarios are analyzed to reach an optimized solution for large hospitals or a network of hospitals.

In [19], the authors propose an advance emergency system for hospitals. Automation system design by utilizing mobile and cloud computing services. Their goal is to create a cloud-based solution that allows for establishing cross-linking connections. These algorithms can assist in the growth of such fields like medicine, enable personalization of treatment and help reduce the cost of healthcare. Big data mining and analytics could serve to pave the way for preventive medicine and lifesaving intervention.

In [20], a suitable cardiac-arrhythmia management and an analytic solution are proposed. It facilitates early recognition of serious cardiac conditions i.e. asystole. In [21], the authors show the capability of big data mining, analysis and decision making assisting process. Their proposed system is able to process measurements from millions of body area network sensors and perform real time response in case of emergencies.

Table 3: Present the Review of Data Analysis

Author	**Journal**	**Year**	**Title**	**Methods**	**Result**	**Comp…**

Fernández-Cardeñosa, G., de la Torre-Díez, I., López-Coronado, M., & Rodrigues, J. J	Journal of medical systems	2012	Analysis of cloud-based solutions on EHRs systems in different scenarios [18]	Cloud-based solution for a Primary care centers network	No	No
Poulymenopoulou, M., Malamateniou, F., & Vassilacopoulos, G.	Journal of medical systems	2012	Emergency healthcare process automation using mobile computing and cloud services [19]	create a cloud-based solution that allows cross-linking institutional healthcare systems	No	No
Puri, C., Ukil, A., Bandyopadhyay, S., Singh, R., Pal, A., & Mandana, K	In Proceedings of the First Workshop on IOT-enabled Healthcare and Wellness Technologies and Systems ACM	2016	ICarMa: Inexpensive Cardiac Arrhythmia Management--An IOT Healthcare Analytics Solution. [20]	The iCarMa; an inexpensive cardiac arrhythmia management and analytics solution is proposed. It facilitates the early detection of fatal cardiac conditions like asystole	No	Yes
Rathore, M. M., Ahmad, A., Paul, A., Wan, J., & Zhang, D.	Journal of medical systems	2016	Real-time Medical Emergency Response System: Exploiting IOT and Big Data for Public Health. Journal of medical systems [21]	The system design big data mining, analysis and decision-making assisting process	No	Yes

2.5 Communications technology for Internet of Thing (IoT)

IoT can move the world to automation[1]. It is the interconnection of various devises and objects having ubiquitous stations that entertain with the unique physical identity such as smart devices, sensors, industrial systems and so on. The information exchange can be done by means of communication tags and location information. IoT servers can handle all short and long data communication. There are separate protocols like COAP, MQTT, etc., that IOT supports and the separate platform i.e. Raspberry Pi, Arduino, etc. supports IoT devices. It adopts a suitable security mechanism to secure from unauthorized access in real-time or online monitoring.

IoT can play a vital role in many areas include marketing, transportation, education, etc., Medical field is also the same impact gives the benefit or remote electronic surgery (e-surgery), electronic diagnosis (e-diagnosis), etc. In [22] the author describes a wireless device for reminding medication perception to patient time to time and diagnose of medication with android applications using an android watch.

The author discovers the use of a wearable device for medication alerts, reminders and notification by designing an android application for android watches in order to remind the patient time and dose of medicine. Missing a dosage of a certain medicine is a problem that even happens in the hospital with inpatients as indicated [23].

This [24] study proposed the IoT based medication adherence system designed for diabetes mellitus patients. It's the chronical illness serious health disease grows with the age of patients. The system develops by using RFID for determines the type of treatments, sensors, and notification send to mobile. The duty is to ensure that the patient can take the correct medicine on time. The advanced feature is that in case wrong medical the system direct contact with the doctor to emergency control on the situation.

In [13], a review paper is presented that provides a summary and comparison of available IoT constructed healthcare systems. Mostly these systems exist for elderly patients, especially those suffering from dementia. In [25], the authors utilize IoT in healthcare by proposing a Smart water bottle for the elderly. The bottle functions to remind the elderly of their daily intake of water.

H. Eskelinen designs an efficient electronic system by conducting two surveys. The system design in four stages that are mechanical, electronic and manufacturing design [26]. The survey is used to list all the requirements of electronic software components.

M. J. Hannafin and F. Wang say that the design-based study would be "grounded, iterative, integrative, interactive, pragmatic, contextual and flexible" [27]. Research is based upon the theory of following principles of this research. The principles are moral support research with design in the outset from, the goal set is for developed the initial scheme theoretically, conduct the survey that presents actual setting of real-world, cooperate faithfully with the team members, try to implement the proposed design systematically, explore data automatically in respective to quick and immediate response, upgrade design document frequently, and verify the generic capabilities of the system.

P. Yammiyavar and A. Saini establishment of the medical health system according to user point of view [28]. The system software design by using an object-oriented approach and the system model is built according to the specific user requirement. Application coded in the User-driven method where different types of end-user and investor prompt about the several constraint and request.

The author proposed a data mining approach that has several design model such as unsuccessful targets, the predictable environment use under certain condition, the condition of the predictable product. [29]. The Internet of Thing (IoT) technique is used for built the method. The statistical approach used for accessing the reliability of hardware by giving inexact condition to determine complete reliability of the system.

Developing technologies generate novel prospects and also make the strong screening of things or people, similar, inside and outside surroundings, due to the importance of various technical and manufacturing applications, a form which one and only most significant domain is human health. Thus, the conducted experiment expresses the project scheme, although their efficiency that the design of the medical care information system project is much interesting in the building by taking emerging developing IoT technologies. The MEMS (Micro Electro Mechanical Systems) used from developing this medical application, that allows integrating various devices such as sensors, actuators, smartphones, etc. [30] [31].

The better choice is to used wireless devices for monitoring and screening purpose[32], The WSN's (Wireless-Sensor-Network) is the most necessary part of the Internet of Things and give the advantage to build smart application including healthcare operate wirelessly [33] .

In this research, the author defines the recognition of the wireless heartbeat and blood oxygen saturation model for monitoring patient health in a short time. [34] The proposed methodology allows automatically to examine the patient's health at hospitals, home, office, in a limited time. Wan-Young Chung, Sing-Hui Toh, and Seung-Chul Lee implement the mobile phone u-healthcare system through various physical signs to identify the real-time capability of the design system. This system design used to capture ECG (Electrocardiography) and BP (Blood pressure) signals from sensors implanted in patients and data spread through WSN by using cellular phones. The smartphone is used to perform all activities of the u-healthcare system. Diagnose the patient's health anywhere in any place with this wireless system. [35].

The WBAN (wireless-body-area network) is a suitable example of IoT based healthcare model. The writer uses wireless technologies for data transfer about the BP(blood pressure), pulse rate, oxygen saturation, heart rate, and body temperature to reduce the error in diagnosing and analyze the correct accurate information [36]. Reza M. and Johan W., apply the Bluetooth and mobile phone technology for checking the health factors of patients [37].

In IoT based healthcare approaches, permanent-monitoring of localization of patient health is the most significant challenge. That’s why the applications are depended upon varies issues technologies and methods. Various applications adopt the RSS (Received Signal Strength) methodology [38]. Maxim, S., uses an indoor localization approach by using RSS dimensions from numerous well-known Wi-Fi accessing points and guaranteeing for accurate localization inside the buildings. This method facilitates its intelligent system for permanent monitoring from home.

The writer of [39] uses Wireless LAN infrastructure with six accessing points. The existing system depends upon tiny Wi-Fi tags and RSS measurements. The expert system captures things such as bed, chair, wheelchairs and many more, with the accuracy roundabout 2m, in the size of 63m x 46m. Wei-Hong C. propose RFID (radio frequency identification) based dynamic indoor monitoring of the patient. The main goal is to enhance the accuracy of healthcare applications. RFID (radio frequency identification), RSS dimensions is used for monitoring and reference tags. [40].

Filippo P. et al. propose a range-based approach for indoor localization application whose activity is receiving signal strength of Bluetooth beacon chunks transmit through anchor node and receive by BLE (Bluetooth-low-energy) devise. The methodology is describing the combination of the stigmergy and RSS measurements approach of Bluetooth low energy. [41] This technique reduces the error of localization up to 75% in1.8m from 6m.

The research paper also deals with reducer the error of indoor localization machinimas used in the healthcare sector for permanent patient monitoring. The algorithm is based on Bluetooth, RSS and supports square vector device which determines the accuracy at the four-floor building. [42].

The research through light upon patient tracking, monitoring, and localization environment. They take RSS of the ZeeBee version. The target is to define an algorithm that makes an 80% accurate tracking system. The several methods and algorithms used for enhancement of accuracy.

In [43], The author uses a hybrid artificial neural network intelligent algorithm to maximize the accuracy of the tracking or localization system. It enables us to capture 70% of the room area.

Table 4: Shows the comparison of all related work on Health care IOT

Author	**Journal**	**Year**	**Title**	**Methods**	**Result**	**Comp…**
Shrivastava, A		2015	Exploring the smartwatch as a tool for medical adherence. [22]	Diagnose of medication on time with smart watch	No	No
Baqir, W., Jones, K., Horsley, W., Barrett, S., Fisher, D., Copeland, R., ... & Stephenson, R.	International Journal of Pharmacy Practice	2015	Reducing unacceptable missed doses: pharmacy assistant-supported medicine administration. [23]	The system medication alerts, reminders and notification by designing an android application for android watches in order to remind the patient time and dose of medicine	Patients with UOD* (96%) Patients with critical UOD* (40%) Total doses due 5927 Number of UOD (170%) Number of critical UOD (69%)	Yes
Islam, S.R., Kwak, D., Kabir, M.H., Hossain, M. and Kwak, K.S.,	IEEE Access	2015	The internet of things for health care: a comprehensive survey. [13]	an intelligent collaborative security model to minimize security risk; discusses how different innovations such as big data, ambient intelligence, and wearables can be leveraged in a health care context	No	Yes
Lee, N. E., Lee, T. H., Seo, D. H., & Kim, S. Y. (2015)	International Journal of Bio-Science and Bio-Technology	2015	A Smart Water Bottle for New Seniors: Internet of Things (IOT) and Health Care Services.[25]	IOT in healthcare by proposing a Smart water bottle for the elderly. The bottle functions to remind the elderly of their daily intake of water.	No	No
Eskelinen, H.	IEEE Aerosp	2001	Improving the productivity of complex electronic systems design by utilizing applied design methodologies. [26]	System design, electronics design, mechanical design and design for manufacturing.	50% of accuracy result by using CAD (computer-aided design)-application	Yes
Wang, F.; Hannafin, M.J.	Educ. Technol. Res. Dev	2005	Design-based research and technology-	design-based research should be “pragmatic, grounded, interactive,		Yes

			enhanced learning environments. [27]	iterative, flexible, integrative, and contextual		
Saini, A.; Yammiyavar, P. Weak eyesight therapy	In Proceedings of the 2013 International Conference on Human Computer Interactions (ICHCI), Chennai, India	2013	Weak eyesight therapy: A case study in designing an application for m-health systems. [28]	The user as the focal point of the design of m-health system		No
Ahmad, M	In Proceedings of the 2015 IEEE 65th Electronic Components and Technology Conference (ECTC), San Diego, CA, USA	2015	Designing for the Internet of Things: A paradigm shift in reliability. [29]	The method is applicable to estimate the target's lifetime in the Internet of Things (IOT).	2465 cycles (assuming a 2% failure rate, and a 90% lower confidence bound characteristic life with a beta of 4.57)	Yes
Coetzee, L.; Eksteen, J	In Proceedings of the 2011 IST-Africa Conference Proceedings, Gaborone, Botswana	May 2011	The Internet of Things—Promise for the future? An introduction. [30]	The robust monitoring of persons or things, alike, in indoor and outdoor environments, becomes of interest to many scientific and industrial applications, where one of the most important is the healthcare domain.	No	No
Ruffieux, D.; Contaldo, M.; Enz, C.	In Proceedings of the 2011 IEEE International Symposium on Circuits and Systems (ISCAS), Rio de Janeiro, Brazil	May 2011	MEMS-based all-digital frequency synthesis for ultralow-power radio for WBAN and WSN applications. [31]	Micro-Electro-Mechanical Systems MEMS which enable the integration of various devices like actuators, sensor nodes or mobiles	Noise of - 136.6 dBc/Hz at 1 MHz offset frequency, for a total synthesizer current consumption of 7.52 mA under 1.6 V supply.	Yes
Mainetti, L.; Patrono, L.; Vilei, A.	In Proceedings of the SoftCOM 2011, 19th International Conference on Software, Telecommunications and Computer Networks, Split, Croatia	September 2011	Evolution of wireless sensor networks towards the Internet of Things [32]	Sketches a framework able to harmonize legacy and new installations, allowing migrating to an all-IP environment at a later stage.	No	No

Khalil, N.; Abid, M.R.; Benhaddou, D.; Gerndt, M.	In Proceedings of the 2014 IEEE Ninth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), Singapore	April 2014	Wireless sensors networks for Internet of Things. [33]	Sensor Networks (WSNs), which constitute the substantial part of IOT.	No	No
Rotariu, C.; Manta, V.	In Proceedings of the 2012 Federated Conference on Computer Science and Information Systems (FedCSIS), Wroclaw, Poland	Septe mber 2012	Wireless system for remote monitoring of oxygen saturation and heart rate. [34]	System will allow the automatic remote monitoring in hospitals, at home, at work, in real time, of persons with chronic illness, of elderly people, and of those having high medical risk.	SpO2 in the range of 40 to100%, HR was kept constant at 60bpm.	Yes
Chung, W.-Y.; Lee, S.-C.; Toh, S.-H.	In Proceedings of the 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vancouver, BC, Canada	Augu st 2008	WSN based mobile u-healthcare system with ECG, blood pressure measurement function. [35]	embed Electrocardiography (ECG) and blood pressure sensors into a cellular phone	No	No
Tan, S.-L.; Garcia-Guzman, J.; Villa-Lopez, F. A	In Proceedings of the 2012 IEEE International Conference on Consumer Electronics (ICCE-Berlin), Berlin, Germany	Septe mber 2012	A wireless body area network for pervasive health monitoring within smart environments. [36]	Bluetooth technology and a smartphone for monitoring the patient’s health parameters. Using PPG Single molding algorithms.	Result is based on Hart rate, blood pressure, oxygen, and skin temperature	Yes
Wannenburg, J.; Malekian, R	IEEE Sens. J	2015,	Body sensor network for mobile health monitoring, a diagnosis and anticipating system. [37]			
Shchekotov, M.	In Proceedings of the 2015 17th Conference of Open Innovations Association	April 2015	Indoor localization methods based on Wi-Fi late ration and signal	In IOT healthcare applications, one of the most frequently-monitored issues are the localization of patient or equipment.		Yes

	(FRUCT), Yaroslavl, Russia		strength data collection. [38]	Numerous approaches are based on Received Signal Strength (RSS)		
Youn, J.H.; Ali, H.; Sharif, H.; Deogun, J.; Uher, J.; Hinrichs, S.H	IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), White Plains, NY	October 2007	WLAN-based real-time asset tracking system in healthcare environments. [39]	a real-time asset tracking system deployed within a hospital clinic setting, which tracked a set of mobile assets using small Wi- Fi tags	accuracy of about 2 m in the hospital clinic environment of 63 m x 46 m size	No
Chen, W.H.; Chang, H.H.; Lin, T.H.; Chen, L.K.; Hwang, S.J.; Yen, D.H.J.; Yuan, H.S.; Chu, W.C.	In Proceedings of the 2009 2nd International Conference on Biomedical Engineering and Informatics, Tianjin, China	2017	Dynamic indoor localization based on active RFID for healthcare applications: A shape constraint approach. [40]	Architecture includes RFID RSS measurements of the reference and monitored tags to estimate the cost function	0.736m in average error that achieved an approximate amount of 32%	no
Palumbo, F.; Barsocchi, P.; Chessa, S.; Augusto, J.C.	In Proceedings of the 2015 12th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Karlsruhe, Germany	August 2015	A stigmergic approach to indoor localization using Bluetooth Low Energy beacons. [41]	stigmergy approach combined with RSS measurements of Bluetooth Low Energy (BLE)	The average localization error of a patient or asset in a 5 m x 10 m healthcare environment is about 0.74	yes
Wyffels, J.; Brabanter, J.D.; Crombez, P.; Verhoeve, P.; Nauwelaers, B.; Strycker, L.D.	. IEEE J. Biomed. Health Information	2014	Distributed, signal strength-based indoor localization algorithm for use in healthcare environments. [42]	a healthcare dedicated indoor localization algorithm based on BLE RSS measurements and squares-support vector machine, resulting at the four room-zone level localization accuracy	localization error of less than 1.8 m in 75% cases in a 6 m	Yes
Jachimczyk, B.; Dziak, D.; Kulesza, W.J.	In Proceedings of the 2014 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)	May 2014	RFID-hybrid scene analysis-neural network system for 3D indoor positioning optimal system arrangement approach is	Radio Frequency Identification (RFID) fingerprints method and the artificial neural network.	accuracy of about 70 cm within a room-sized environment	Yes

			indoor localization [43]			

2.6 Challenges and advantage of using Internet-of-Things in healthcare

In [44] paper describe that implementation of IOT in healthcare sectors the services are changing in health care sectors. Which help physician to diagnose patient health in right moments. Moreover, the writer presents the architecture for internet of things which especially build only for hospitals and divided in to six following parts.

1. The technologies rate increase by up to 117$ billion by 2020. There are following technologies like Identification, sensing, communication, location, visualization and many more.
2. The author figures out the general architecture by using the technologies. Using sensors, servers, database system etc.,
3. The services include BP monitoring, temperature checking, glucose level monitoring for diabetes patient, hate rate sensor and so on.
4. The main challenges in as the no. of departments increase the devices are so slow. The challenges include scalability, interoperability, safety, security and mobility.
5. The IOT analytical are Electronic health record, analysis data from sensor, biomedical signal and analysis data from genomic, clinical data mining and social media etc.,
6. Machine learning algorithms used to collect data from and make the system to work in an effective and in efficient manner. Or in case of any emergency sent alert notification to doctor or automatically arrangement of ambulance if doctor is busy at that time.

2.7 Medical Data Analysis

This section covers the latest related work done on different medical data sets such as Brest cancer, iris, hypothyroid, dermatology and diabetes.

Mostly, females can die because of Brest cancer. With the Data mining hybrid model, there is a chance to detect cancer at an early stage. By applying the classification techniques on the patient record. Research says that 600,000 woman deaths occur due to this cancer [54]. In 37.3% of Brest cancer in Egypt. 30% in Arab countries and 42% in European countries [55].

Skin Cancer is also a major health problem in every year 5.4 million non-melanoma skin cancer cause in the US. And in American, there is 20% [56]. The Data mining classification tool makes easy for the healthcare sector to identify cancer.

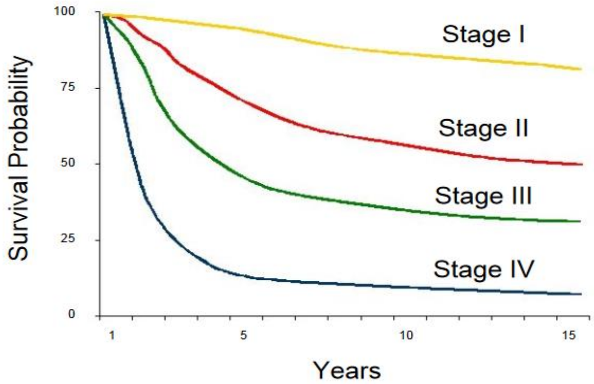


Figure 3: shows the dataset stages of skin cancer

Diabetes disease is the most serious health care challenge. International Diabetes Federation predicts that 285 million diabetic patients all over the world. And in the future predictable that 380 million patients increase within 20 years. According to WHO the 442 million peoples suffer from diabetes and 50% of medical organization cannot provide the treatment to diabetic patient [57]. A classifier scheme used for disease detection with the best cost and enhanced performance in an early stage.

Disease	**World Wide Ratio of disease occurrence**
Brest cancer	37.3% in Egypt, 30% in Arabs and 42% in Europeans
Skin Cancer	20% in the US. And in American
Diabetes	442 million

Table 5: World Wide Ratio of disease occurrence

Hyperthyroid or an overexcited thyroid cause swelling of the thyroid. The significance of thyroid illnesses must not be taken too lightly as thyroid gale and myxedema coma leads to death major number of cases [58]. Diagnose of the thyroid is the main classification problem. Appropriate clarification of the thyroid records, also medical analysis, is a vital issue in the diagnosis of thyroid disease. The intelligent technique is applied to analyze the result of hypothyroid disease.

A [45] researcher proposed the hybrid smart system consists of few algorithms such as Min-Max neural network and decision tree. These efficient models used to classify data of the patient to make an accurate decision. The experiment conducts on three benchmark clinical datasets i.e. Pima-Indians-Diabetes, Breast-Cancer Wisconsin UCI Origin of ML (machine learning) are often used performance. To figure out the prediction giving via a hybrid system.

As the worldwide issues in the data processing is noise. Outliers, irrelevant, missing values, inconsistency and inaccuracy to justify the issues the feature selection is the main issue. This researcher used hybrid intelligent techniques chaos and rough-set theory and optimal Binary Grey Wolf algorithm to resolve the feature selection issue [46]. The whole outcome specifies that the proposed method conveys superior performance, lesser miscalculation, the greater speed with less implementation time.

Health analytical support schemes compact through the enormous no. of infection depths and a comparatively lesser amount of the patient data. The measurements critically disturb the effects of the diagnosing system. Furthermore, these attributes require more system time and space to diagnosis patient health. The [47] design approach fuzzy genetic algorithm is based neural network (FGANN) addressed such problem. The fuzzy logic, neural network and genetic algorithm used for identifying the results of health illness in less system space and time.

The different type of hybrid intelligent data mining algorithms such as genetic algorithm, support vector mechanism, artificial neural network with single-layer and multi-layer preceptor, fuzzy logic Bayesian, decision tree, min-max algorithm, K-nearest neighbors and many more are used for data processing and prediction, remove the error such as missing values, outlier and noise data. The resents research apply these latest hybrid techniques on medical data sets to diagnose patient health accurately and predict the correct complete information for feature classification. Each algorithm has its own classifier methods. The main aim to analyze the disease at the early stages and save human life.

Table 6: Shows the comparison of all related work on Medical Data Sets

Author	**Journal**	**Year**	**Title**	**Methods**	**Result**	**Com p…**
M. Seera and C. P. Lim	Expert Systems with Applications	2019	A hybrid intelligent system for medical data classification [45]	Fuzzy Min–Max neural network, the Classification and Regression Tree, and the Random Forest model.	The LSSVM with fuzzy the sensitivity and specificity results, i.e., 95.00% and 93.33%, and FMM-CART-RF 96.54% sensitivity and 94.44% specificity.	yes
A. M. Anter, A. T. Azar, and K. M. Fouad	The International Conference on Advanced Machine Learning Technologies and Applications	2019	Intelligent Hybrid Approach for Feature Selection [46]	Rough Set Theory (RST), Chaos Theory and Binary Grey Wolf Optimization Algorithm (CBGWO)	Accuracy value obtained by CBGWOFS for two chaotic (Piecewise, Singer maps) (0.95).	yes
P. Aruna Kumari and G. Jaya Suma	Computational Intelligence and Big Data Analytics: Applications in Bioinformatics, Springer Singapore	2019	FGANN: A Hybrid Approach for Medical Diagnosing [47]	fuzzy approach and genetic algorithm (GA) and neural network (NN).	Accuracy is greater than 88% for diabetic and 98% for thyroid.	Yes
V. Chaurasia, S. Pal, and B. Tiwari	Journal of Algorithms & Computational Technology	2018	Prediction of benign and malignant breast cancer using data	Na¨ıve Bayes, RBF Network, J48 predict the dataset (683 breast cancer cases)	Naïve 97.36%, RBF 96.77% and J48 93.41% accuracy.	yes

			mining techniques [54]	with 10-fold cross-validation methods.		
E. Alickovic and A. Subasi	Springer International Publishing	2019	Normalized Neural Networks for Breast Cancer Classification [55]	The model developed by using Normalized Multi Layer Perceptron Neural Network to classify breast cancer with high accuracy.	The results achieved is very good (accuracy is 99.27%)	yes
Andre Esteva, Brett Kuprel, Roberto A. Novoa	A Nature research journal	2017	Dermatologist-level classification of skin cancer with deep neural networks [56]	Define classification of skin lesions using a single CNN, trained end-to-end from images directly, by using a dataset of 129,450 clinical images which consisting of 2,032 different diseases.	Three ways accuracy of CNN is 69.4%	yes
M. M. F. Islam, R. Ferdousi, S. Rahman, and H. Y. Bushra	Springer Singapore	2019	Likelihood Prediction of Diabetes at Early Stage Using Data Mining Techniques [57]	The dataset of 520 instances. We have analyzed the dataset with Naive Bayes Algorithm, Logistic Regression Algorithm, and Random Forest Algorithm and after applying tenfold Cross- Validation and Percentage Split evaluation techniques, Random forest has been found having best accuracy on this dataset.	Random Forest Algorithm using 10- fold cross-validation 97.4%. Percentage split technique classify 99% of the instances correctly	
S. Gaikwad and N. Pise	International Journal of Data Mining & Knowledge Management Process	2014	An Experimental Study on Hypothyroid Using Rotation Forest [58]	Decision learner, Best first decision tree and J48	96.63% classification accuracy has been gained.	no

CHAPTER 3

PROPOSED METHODOLOGY

3.1 Framework definition and types

A framework is a design planned to help and empower the reconciliation and interoperation of programming components. It might comprise of ideas, innovations, apparatuses, conventions, principles, control instruments, interfaces, and procedures expected to empower the quick, productive, and adaptable get together of system from parts in a down to setting. However, to be empowered to understand the definitions as every form and relies on the former definitions.

Software Reuse: refers to utilizing a formerly created unit, bundle, or module of programming more than once, either for the reason for which it was initially created or for another reason or in another specific situation.

Component: is key with regards to general programming reuse.

Composability and Composition: is the ability to choose and collect recreation segments in different blends into legitimate reproduction frameworks to fulfill client necessities.

Standards: is an agreed-upon tradition or prerequisite that characterizes or recommends some viewpoint or parts of a specialized system.

Resolution: is depicting a model, determination is characterized as "the level of detail which with this present reality is mimicked".

Library: is a gathering of components that are accessible for reuse.

3.1.1 Types of Frameworks.

3.1.1.1 Common Library.

It is relying upon an amassing or set of programming modules or segments that are composable or reusable through conformance to a standard interface or set of interfaces that empowers the modules to interoperate with substitute modules in the library or a subset of them.

3.1.1.2 Product Line Architecture.

Depends on the neatly arranged improvement of various related recreation items that, to the degree conceivable, and share basic programming parts.

3.1.1.3 Interoperability Protocol.

Depends on the run-time trade of reenactment information or administrations, commonly utilizing a conveyed recreation interoperability convention, for example, ALSP, DIS, HLA, or TENA.

3.1.1.4 Object Model.

Depends on a standard for segment particulars. Put in mind that the standard is for detail, not usage, of the segment.

3.1.1.5 Formal.

Relies on a formal scientific documentation to characterize the segments, arrangements of models, and the interfaces between them.

3.1.1.6 Integrative Environment.

Is a software development and execution condition are utilized to associate segments which may have been composed with no earlier purpose to interoperate?

3.2.1 Framework types summary

Framework type	Components	Composition mechanism
Common library	Models implemented as software modules	Component interfaces defined by components or framework; components linked into common executable; data exchanged via method calls
Product line architecture	Software modules	Component interfaces defined by framework; components linked into common executable; data exchanged via method calls
Interoperability protocol	Independent executable	Components execute independently as separate processes; data exchanged via network messages
Object model	Conceptual model	None; connection depends on mapping and implementing conceptual module within another framework
Formal	Formal model	Interpreter for formal models
Integrative environment	Model implemented as file, spreadsheet, or software module	Components "wrapped" with special interface software; components linked into common executable; data exchanged via method calls

Figure 4 Framework types summary [48]

3.2 Proposed Framework

In this section, we are going to propose the general framework for the hospitals. However, this general framework shall fit in most of these hospitals. We suggested this proposed "general framework" to have four stages. The working strategy or work process of this general framework is given underneath:

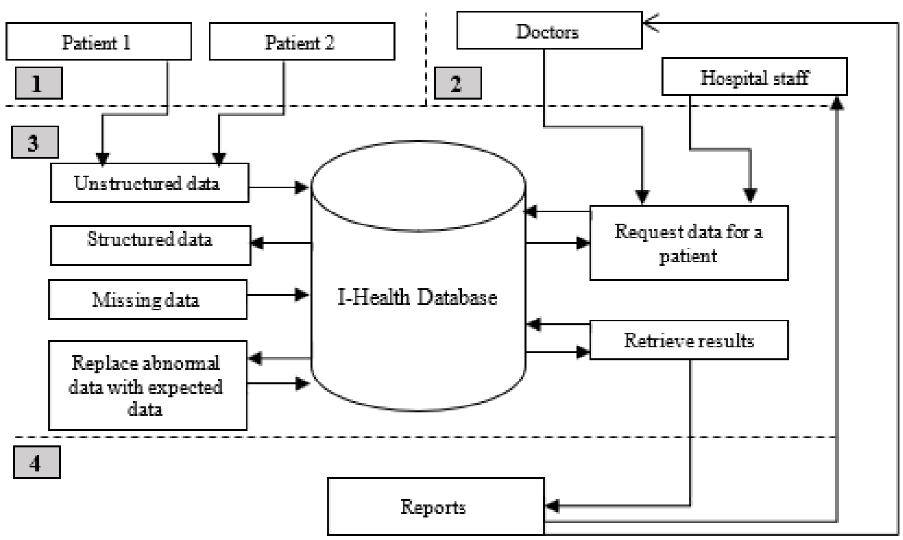


Figure 5: proposed [I-health] general framework

Stage1: Input by the devices stage: Through the wristband or smart watch or smartphone, patient can send properties estimation of illness like plasma glucose, pulse and so forth into a database by the web association (mobile application might work in this stage). Like the wristband, Medicare associated with the I-health databases and can send every one of the data in this I-health database.

Stage2: Input | output stage: In this stage it is consider as input and output. Doctors can input the patients' information into the dataset. Although, can review and verify this information.

Stage3: Processing stage: This stage is the essential of this proposed framework. Patients send data, it can be unstructured data. the Dataset can change over those data indexes into "structured data". After the course of action.

Stage4: Reports stage: in this section, it will allow the doctors, or the hospital staff to generate reports and the result of the patients. This stage will give the doctors a chance to take a decision for the patients. Therefore, this stage is depending on the quality of the work that done in the previous stages to insure the good decision.

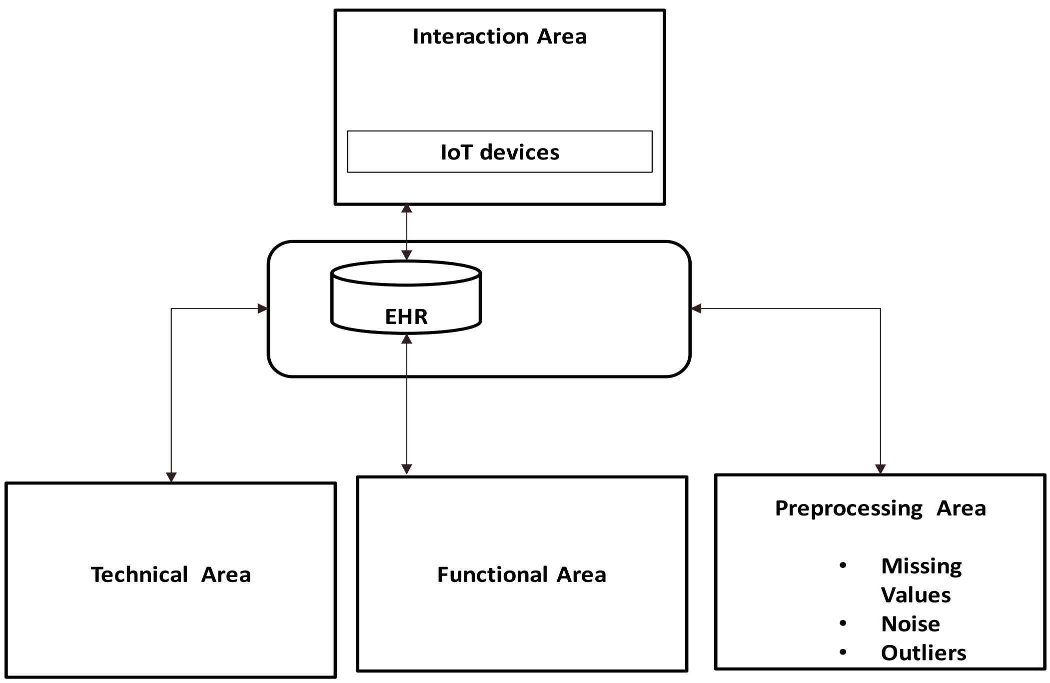


Figure 6: Proposed Framework

3.2.1 Interaction Area.

3.2.1.1 Human interaction.

The Human Interaction is an essential component of this framework. Along these lines, the Automated Medicine is outfitted with a touch screen interface that permit patients (clients) to associate with the giving data, for example, quick criticism on negative reactions from certain solution, adding another sensor, and speaking with the specialist. On the opposite side of the framework, the framework should contain an easy to use interface essentially intended for the specialists, which can be gotten to by patients and any authorized individual also. It should have an association interface open from any figuring devices for specialists to endorse medication, see measurements, reports, suggested choices, and analysis prescribed

3.2.1.2 IOT devices

Comprises of IOT devices with sensors that can gauge indispensable signs including however not restricted to pulse rate, blood pressure, body temperature, stress level. This influences IOT devices to plug-and-play effortlessly configurable by the patients. These devices send the readings to the EHR-I Health dataset, which thus sends these readings to the cloud, if necessary. The EHR-I Health dataset send the readings to the [processes] to be analyzed. Then, it shall be sent to "Automated Medicine" to send alerts to the patients to remind him/her of the planned and out of schedule medicine dosages. Even though, these IOT devices are outlined particularly for home utilize and with patients that require consistent therapeutic checking at home, the supposition is that patients will be at home constantly and the correspondence interface amongst devices and Automated Medicine is steady. Nevertheless, if the patients leave home for a brief period, for a specialist visit for instance, the IOT devices should be furnished with a memory unit to hold information recorded until the point when the patient returns home and the information is consequently exchanged to the Automated Medicine when correspondence connect is set up. Since we do not implement these devices, we will download some readings from the internet.

3.2.2 Electronic Health Database.

3.2.2.1. EHR- I-Health

Middleware is programming that lives between applications, administrations, and their fundamental appropriated design and stages. Middleware gives a few sorts of capacities to engineers, including giving more elevated amount programming reflections to help the advancement of utilizations and administrations; supporting end-to-end quality traits, for example, adaptability, determination, and security; and veiling issues, for example, system disappointment and heterogeneity of dialects, working frameworks, and systems.

The difficulties for middleware radiating from the appearance of the Internet of Things and distributed computing, new research is required, especially regarding the kinds of middleware expected to help key areas, for example, digital physical systems, shrewd urban communities, the savvy lattice, enormous information examination, advanced earth, etc.

The Internet of Things: IoT presents critical new challenges for middleware originating from the immense number of associated objects, the volume and abundance of the information delivered, the rich examples of correspondence required, the heterogeneity of conveying components, and new challenges as far as nature of-administration, protection, and security.

Cloud computing: middleware for cloud computing is more develop than middleware for IoT, yet imperative difficulties remain. One of the key challenges is heterogeneity, which is really expanding after some time.

Systems of systems: appropriated systems without bounds will never again be singular systems, however rather will be systems of systems and ultra-substantial scale systems, which themselves will be progressively created.

3.2.2.2. Cloud storage

Cloud service is the best for some of the companies, where they do not want to spend money and effort for dealing with this every day. Regardless, the security perspective, the company who only decide what is better for them. However, for this part; we are recommending the cloud storage to be a part of the framework, and the hospitals are taking this decision either to go with cloud or not.

3.2.3 Functional Area.

3.2.3.1 Automated Medicine.

Automated Medicine part comprises smart decision, after the readings are analyzed by the [processes] part. The Automated Medicine permitted parameters and permits out-of-plan measurements in light of medicinal readings to deal with low need crisis cases. For instance, if the "Automated Medicine" dispenses a fever prescription as indicated by some booked measurements. This solution should produce results inside one to two hours from measurements. Nonetheless, if a measure following two hours shows a high temperature at that point in view of the therapeutic parameters and if restoratively allowed, the prescription framework will administer an out-of-plan fever reducer and guarantee that the patient takes another dosage.

3.2.4 Processing Area -Processes.

In this section, we are going deep in the process's components. They are listed as the following:

3.2.4.1 Data Collection

Since we are not going to implement this project, and we cannot use the IOT devices. We have to get the data from the internet to test them. However, these data will be shown in the chapter 4.

3.2.4.2 Data Preprocessing

Preprocessing is a data mining technique that involves converting raw data into a clear format. Captured data is often unfinished, unreliable, and/or lacking in behaviors, and is likely to comprise many mistakes. Data preprocessing is a confirmed method of resolving such matters. Data preprocessing makes raw data is prepared for more processing. Data preprocessing goes through series of stages before finishing the transformation:

Data cleaning: data is going through in processes in this stage such: filling in the missing values and solving the irregularities data.

Data integration: Data with different impersonation putting together and solve the conflict with the data.

Data transformation: making all the data normalized and aggregated.

Data reduction: This step goal to present a reduced representation of the data in a data warehouse.

Data Discretization: Includes the decrease of many values of a constant attribute by dividing the range of attribute intervals.

3.2.4.2.1 Data Cleaning

Also, called data cleansing or scrubbing, offers with detecting and removing errors and inconsistencies from data to be able to enhance the nice of data. Data excellent troubles are present in unmarried data collections, along with files and databases, e.g., because of misspellings throughout facts entry, missing records, or different invalid records. While more than one data assets need to be incorporated, e.g., in data warehouses, federated database systems or global net-based information systems, the need for data cleaning will increase drastically. That is due to the fact the resources often contain redundant data in different representations. On the way to provide access to accurate and consistent data, consolidation of various data representations and elimination of reproduction information emerge as essential. Following is a figure explain the processes for cleaning the data:

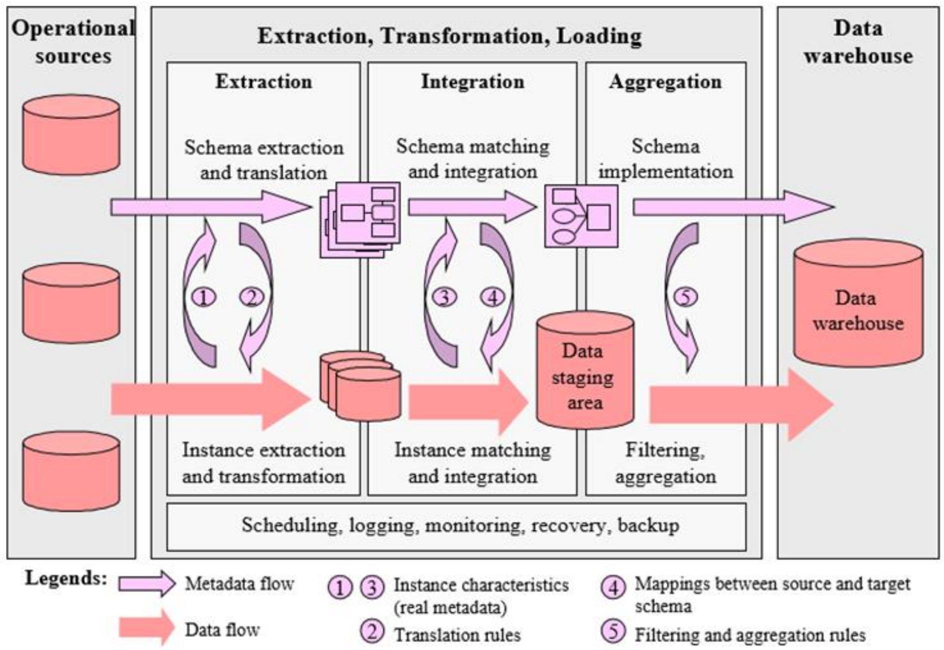


Figure 7: Cleaning Data

3.2.4.2.2 Data analysis

It can check with a ramification of precise procedures and methods. However, before applications can effectively use those procedures and methods, we agree with it is far crucial to see data analysis as part of a process. By using this, we suggest that data analysis entails desires; relationships; decision-making; and thoughts, similarly to running with the real data itself. Data analysis includes ways of running with information (data) to aid the paintings, goals and plans of your software or enterprise.

From this perspective, we present a data analysis process that consists of the subsequent key components:

- Purpose.
- Questions.
- Data Collection.
- Data Analysis Procedures and Methods.
- Interpretation/identification of Findings.
- Writing, Reporting, and Dissemination.
- Evaluation.

3.2.4.2.3 Missing Data

Missing data technique has divided into three types as follows:

[MCAR] Missing Completely at Random: there is no connection between the messiness of the data and any values, watched or missing. Those missing data focuses are an irregular subset of the data. There is nothing orderly going on that makes a few information more inclined to miss than others.

[MAR] Missing at Random: there is an efficient connection between the penchant of missing values and the watched data, yet not the missing data. Regardless of whether a perception is missing has nothing to do with the missing values, yet it has to do with the estimations of a person's watched factors. Along these lines, for instance, if men will probably reveal to you their weight than ladies, weight is MAR.

[MNAR] Missing Not at Random: there is a connection between the inclination of an incentive to miss and its value. This is where the general population with the least training are absent on instruction or the most debilitated individuals are well on the way to drop out of the study.

3.2.4.2.4 Normalization.

The technique of existing normalization methodologies is given below:

Min-Max: The technique which keeps relationship among unique data is called Min-Mix Normalization. Min-Max Normalization is a basic method where the procedure can particularly fit the data in a pre-characterized limit with a pre-characterized limit:

$$A' = \left(\frac{A - min\ value\ of\ A}{maxvalue\ of\ A - \min value\ of\ A}\right) * (D - C) + C$$

Z-score: in this method, assume we are having five columns in particular X Y, Z, U, and V with various factors or sections that are 'n' in each line. So, in each column above z score strategy can be connected to ascertain the standardized ones. On the off chance that assume some column having every one of the qualities are indistinguishable, so the standard deviation of that column is equivalent to zero then all esteems for that line are set to zero. Like that Min-Max standardization the z-score additionally gives the range of qualities near 0 and 1.

$$std(E) = \sqrt{\frac{1}{(n-1)}\sum_{i=1}^{n}(v_i - \bar{E})^2}$$

Decimal Scaling: The method which gives the range between - 1 and 1 is only Decimal Scaling.

$$v^i = \frac{v}{10^j}$$

3.2.4.3 Feature Selection and Extraction.

In machine learning as the dimensionality of the data rises, the measure of information required to give a dependable examination develops exponentially. A prevalent way to deal with this issue of high-dimensional datasets is to look for a projection of the information onto fewer factors (or highlights) which protects the data however much as could be expected. Microarray information is run of the mill of this sort of little specimen issue. Every datum point (test) can have up to 450,000 factors (quality tests) and handling an expansive number of information focuses includes high computational cost. At the point when the dimensionality of a dataset develops fundamentally there is an expanding trouble in demonstrating the outcome measurably huge because of the sparsity of the significant information in the dataset being referred to. Extensive datasets with the alleged "expansive, little " issue (where is the quantity of highlights and is the quantity of tests) have a tendency to be inclined to overfitting. An overfitted model can mix up little changes for vital fluctuation in the information which can prompt order blunders. This trouble can likewise increment because of loud highlights. Commotion in a dataset is characterized as "the blunder in the difference of a deliberate variable" which can come about because of mistakes in measurements or normal variety. Machine learning calculations have a tendency to be influenced by uproarious information. Commotion ought to be diminished however much as could reasonably be expected with a specific end goal to maintain a strategic distance from superfluous intricacy in the induced models and enhance the proficiency of the calculation. Regular commotion can be partitioned into two sorts.

Attribute Noise: caused by mistakes in the trait values (wrongly measured factors, missing values).

Class Noise: caused by tests that are marked to have a place in more than one class or potentially misclassifications [49].

3.2.4.4 Classification.

We use three main types of the classification algorithm, decision trees (J48 classifier), Bayesian (Naïve Bayesian classifier) and Artificial neural network (MLP classifier) are evaluated in this study. “Classification is a data mining calculation that makes a well-ordered guide for how to decide the yield of another data

<u>Decision tree(J48 classifier)</u>

The tree it makes is precisely that, a tree whereby every hub in the tree speaks to a spot where a choice must be made considering the information, and you move to the following hub and the following until the point that you achieve a leaf that lets you know the anticipated yield. Sounds confounding, however it's extremely very direct” [50].

```
    [ Will You read This book?]
          /          \
        Yes           No
        /              \
[Will You Understand It?]  [Won't Learn It]
     /           \
   Yes            No
   /               \
 [Will Learn It] [Won't Learn It]
```

Table 7**:** Classification tree [50]

Decision tree is supervised learning algorithm i.e. classes or categories of the data instances are known, and concept based on if-then rules. The target function can be Boolean or discrete valued. It can be represented by as following:

- Each node corresponds to an attribute
- Each branch corresponds to an attribute value
- Each leaf node assigns a classification

Decision trees represent a disjunction of conjunctions of constraints on the attribute values of instances. Each path from the tree *root* to a *leaf* corresponds to a conjunction of attribute tests (one rule for classification). The tree itself corresponds to a disjunction of these conjunctions (set of rules for classification).

First of all, we *select* the best attribute to be tested at the root of the tree. For making this selection each attribute is evaluated using a statistical test to determine how well it alone classifies. The selection process is then repeated until associated with each descendant node to select the best attribute to test at that point in the tree. For this we have to calculate the entropy and information gain of each attribute. Calculation steps are as following

1. **Identify the options**
2. Add possible outcomes
3. Add Associate cost, outcome probabilities and financial results
4. Calculate the expected value and net gain

In order to define information, gain precisely, we begin by defining entropy. Entropy is a measure commonly used in information theory, called. Entropy characterizes the impurity of an arbitrary collection.

$$H(x) = -\sum_{j=1}^{m} p_j \ log_2 \ \ p_j$$

This formula is called Entropy H. High Entropy means that the examples have more equal probability of occurrence (and therefore not easily predictable). Low Entropy means easy predictability.

Information Gain is the expected reduction in entropy caused by partitioning the examples according to an attribute's value.

$$Info\ Gain(n) = Entropy(x) - ([weighted\ average]) * entropy\ (chidern\ of\ feature\ )$$

$$Info\ Gain(Y|X) = H(Y) - H(Y|X)$$

For transmitting Y, how much bits would be saved if both side of the line knew X

A decision-tree model is the same as a structure of a tree where the attributes are representing as an internal node. The outcome shows the test of a single branch that leads to a leaf node which characterized as class allocation. The first node of the tree is the root node. This classifier is built-in the top-down method, recursive technique, and divide-and-conquer approach. First, the class label is defining according to the set of training tuples, the tree establishes into the subclass minor and training set divided reclusively. However, not all outlets are seen in a decision tree. The tree built from the trained class and recursively divided into its branches. However, not all outlets are seen in a decision tree. Algorithm J48 is used in our experiments [51]. Tree pruning attempts to classify and remove internal nodes that reflect outliers and noise. The target is to of improving accuracy and reduce errors.

Artificial neural network (MLP classifier):

Artificial neural network is a simulation of biological nervous systems, such as the brain, to process information. The primary element of that system is the process element called neurons, which work in unison to solve specific problems.

MLP is a frequently used supervised learning classifier based on feed-forward pass network and back-propagation network, this technique is used for pattern recognition [52] used Artificial neural network. MLP is based on three layers input, hidden and an output layer that takes out meaningful information during training and plot update weights values to the elements of the input layer. In forward pass, method weights assign between input units and hidden layers' nodes and then from hidden layer node to output layers. The comparison is done between target output and output value to get the error. BP method is used to reduce errors and adjust weights accordingly. Throughout the training multidimensional space is occupied for the activation of hidden components.

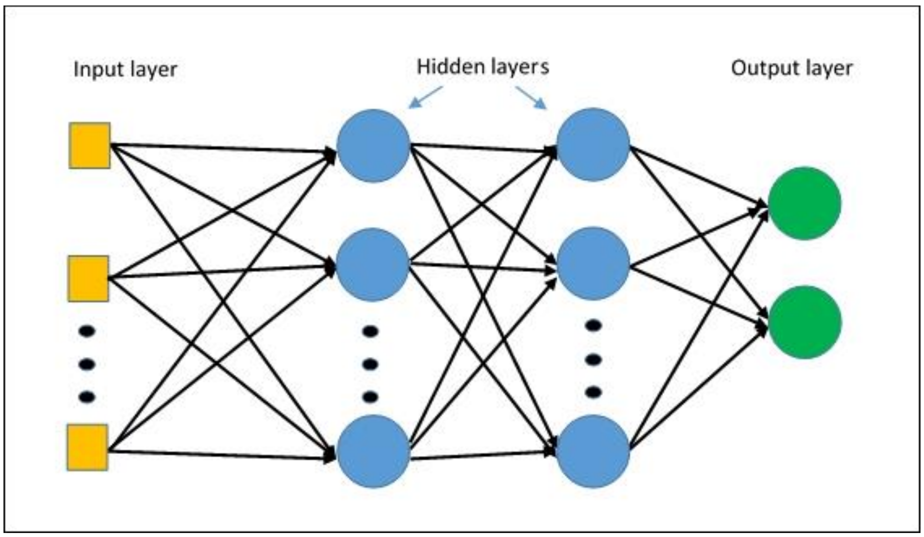


Figure 8: The Inputs and the Output of the ANN

In this section, we will illustrate how the ANN works. The ANN with n inputs, k hidden and m outputs is considered. The weight between input i and hidden j and between hidden i and output j are denoted as $w_{ij}^{(1)}$, $w_{ij}^{(2)}$respectively. The inputs in each layer are extended with 1 constant component with value equal to 1.The weight between the constant 1 and hidden j and between constant 1 and output j are denoted as $w_{n+1,j}^{(1)}$, $w_{K+1,j}^{(2)}$respectively. Figure below illustrates how the ANN structure is and how it works.

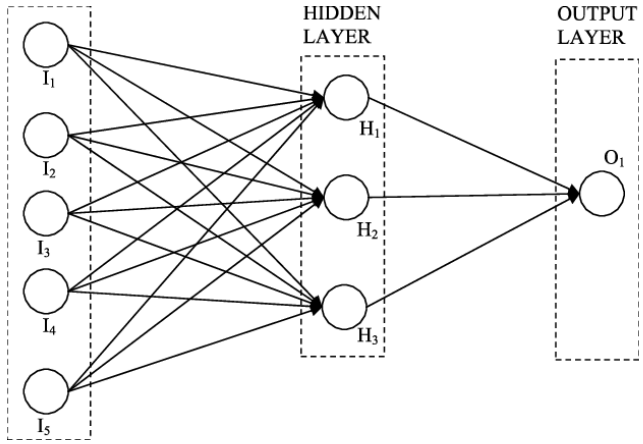


Figure 9 Artificial Neural Network Structure

The n-dimensional input vector $o = (o1, \ldots, on)$ is transformed to $\hat{o} = (o1, \ldots, on, 1)$

Let $(n+1) \times k$ matrix with component $w_{ij}^{(1)}$ and $(K+1) \times m$ matrix with component $w_{ij}^{(2)}$ at the i-th row and the j-th column and lets denoted each of them as $\overline{W_1}$, $\overline{W_2}$ respectively. Below formula is the excitation net_j of the j-th hidden unit.

$$\mathrm{net}_j = \sum_{i=1}^{n+1} w_{ij}^{(1)} \hat{o}_i$$

The activation function is a sigmoid and the output $o_j^{(1)}$ is given by

$$o_j^{(1)} = s\left(\sum_{i=1}^{n+1} w_{ij}^{(1)} \hat{o}_i\right)$$

The output from the hidden unit is denoted as $o^{(1)}$ and given by

$$o^{(1)} = s(\hat{o}\overline{W_1})$$

The output from the network is the m-dimensional vector denoted as $o^{(2)}$ and given by

$$o^{(2)} = s\left(\hat{o}^{(1)}\overline{W_2}\right)$$

Mentioned formulas can be used for generating any ANN no matter of how numbers of the nodes/ neurons are there.

Bayesian (Naïve Bayesian classifier):

The Bayesian algorithm is one of the DM techniques. We use key Bayesian techniques namely naive Bayes used for classification that are executed in WEKA software [53]. It is the statistical classifiers depend upon Bayes' theorem which calculates the probability of a single attribute belongs to a certain class.

Bayesian learning algorithms are among the most practical approaches to certain types of learning problems. These results are comparable to the performance of other classifiers, such as decision tree and neural networks in many cases.

Bayes Theorem

Let *X* be a data sample, e.g. red and round fruit

Let *H* be some hypothesis, such as that *X* belongs to a specified class *C* (e.g. X is an apple)

For classification problems, we want to determine $P(H|X)$, the probability that the hypothesis *H* holds given the observed data sample *X*

The probability *P(H)* is called the prior probability of *H*, i.e the probability that any given data sample is an apple, regardless of how the data sample looks

The probability $P(H|X)$, is called posterior probability. It is based on more information, then the prior probability *P(H)* which is independent of *X*

calculating the posterior probability as following:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}$$

$P(X|H)$, is the posterior probability of *X* given *H.* And *P(X)* is the prior probability of *X*

Bayes Theorem: Proof

Similarly:

$$P(X|P) = \frac{P(H \wedge X)}{P(H)}$$

Since we have:

$$P(H \wedge X) = P(H|X)P(X)$$

And also

$$P(H \wedge X) = P(X|P)P(H)$$

Therefore

$$P(H|X)P(X) = P(X|HX)P(H)$$

And hence

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}$$

Naïve (Simple) Bayesian Classification

A probabilistic NB (naïve-Bayes) classifier based is feature independent model. It is conditional class independence classifiers that mean the effect of a given class attribute value is independent of another class value.

Studies comparing classification algorithms have found that the simple Bayesian classifier is comparable in performance with decision tree and neural network classifiers. It works as follows:

Each data sample is represented by an n-dimensional feature vector, X = (x_1, x_2, …, x_n), depicting n measurements made on the sample from n attributes, respectively A_1, A_2, … A_n

Suppose that there is m classes C_1, C_2, … C_m. Given an unknown data sample, X (i.e. having no class label), the classifier will predict that X belongs to the class having the highest posterior probability given X

Thus if

$P(C_i|X) > P(C_j|X)$ for $1 \leq j \leq m$, $j \neq i$then X is assigned to C_i

This is called Bayes decision rule. We have

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)}$$

As P(X) is constant for all classes, only $P(X|C_i)$ $P(C_i)$ needs to be calculated

The class prior probabilities may be estimated by

$$P(C_i) = s_i / s$$

where s_i is the number of training samples of class C_i & s is the total number of training samples

If class prior probabilities are equal (or not known and thus assumed to be equal) then we need to calculate only $P(X|C_i)$

In order to reduce computation, the naïve assumption of *class conditional independence* is made

This presumes that the values of the attributes are conditionally independent of one another, given the class label of the sample (we assume that there are no dependence relationships among the attributes).

Thus, we assume that $P(X|C_i) = \prod_{k=1}^{n} P(X_k|C_i)$

The probabilities $P(X_1|C_i), P(X_2|C_i), \ldots\ldots, P(X_n|C_i)$ can be estimated from the training samples, where an attribute A_k, which can take on the values x_{1k}, x_{2k}, …

3.2.4.4 Technical Area

In this area all users shall be able to connect to the internet in order to get this service for monitoring and reporting the medical devices.

3.3 Summary

In this chapter we have talked about the proposed framework. The proposed framework has full details about its domains and components. Each one of these domains or components have been explained fully. We have explained most of popular types of frameworks that are using by companies according to the source. Then, we moved to the proposed framework to explain each area in it. Starting from the interaction area, Electronic Health Database, functional area, and finally the processing area. However, each one of these areas have some components that are linked coherently for best performance for the IOT devices that will be connected between patients and doctors. This chapter is the great guide to whom wants to get more details about what is in the proposed framework.

CHAPTER 4
ANALYSIS AND DISCUSSION

4.1 Introduction

In this chapter, we have gathered some information related to some patients in the hospitals. Since, we are not applying a system method we have gathered the information from the internet. We have focus on five diseases which are breast-cancer, hypothyroid, and Pima_diabetes, iris, skin cancer. However, we are applying this data in Weka program to mining them with some classifiers. The description classifiers are mentioned in the chapter 3 for each one. However, the classifiers are as the following: Bayesian (naïve- Bayesian), MLP (multiple layer perceptron), and decision-tree (J48).

Use training set: the classifier is assessed on how well it predicts the class of the examples it was prepared on.

Cross-validation: the classifier is assessed by cross-approval, utilizing the quantity of folds that are entered in the Folds content field. In this option we are going to test 20 folds, 50-fold and 80 folds.

4.2 Performance of a classification Measures

Mostly testing begins with a confusion matrix, which demonstrations incorrect and correct classification from respective class. The correct classification is TN (True-Negative) rate and TP (true-positive) rate. And for incorrect classification predication is done by FN (False Negative) rate and FP (False positive) rate [6]. The total achievement rate is figured out through the below mention formula:

$$\frac{TP + TN}{TP + TN + FP + FN}$$

For error

$$\frac{FP + FN}{TP + TN + FN + FP}$$

MCC (Matthews correlation coefficient)

$$\frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Accuracy: To calculate the accuracy of a classifier value determined in the percentage of a testing set. The classifier accuracy is identifying by the instance of the several classes.

Receiver Operating Characteristic (ROC): ROC is the representation of the graphical notations given by the TP (True positive) rate as a function of the FP (false positive) rate, it's a better visualizing tool to identify the performance of the classifier and helps in selecting an appropriate starting point. The AUC (area under the curve) of Receiver Operating Characteristic is frequently used as a statistic for ideal comparison. The result is accurate only If under curve area is longer.

Kappa: The complete correct classified instances reveal a simple calculation of the classifier, is same as calculating through the under-curve area under a ROC. As a classifier depend on random selected value which often categorize some illustrations correctly. The kappa statistic mechanism uses to control those illustrations that might be by chance classified correctly. Likewise, it calculates the correctness of each classifier via F-measure that represents the vocal mean among accuracy. Kappa is used to measure an exactitude of a specific assessing case. It normal differentiate among the consistency data validity.

F-Measure

When recall and precision both are balanced the high results produce by F-measure produce. When proportion is the Precision.

$$\frac{TP}{FP + TP}$$

When proportion is the Recall

$$\frac{TP}{FN + TP}$$

F-Mean derivative

$$\frac{2 * Recall * Precision}{Recall + precision}$$

Table is for the performance of classification

The following table is for the performance of classification for the results that are added in next section. The purpose of this section to understand these terms and know the accuracy of the result.

Term	Description
True positives (TP)	We predicted yes. They do have the disease.
False positives (FP)	We predicted yes, but they don't have the disease
precision	When we predict it yes, how often is it correct?
Recall	When it's yes, how often does it predict yes?
F-Measure	It is a derived adequacy estimation. The resultant value is translated as a weighted normal of the precision and recall.
MCC	Is fundamentally a connection coefficient between the genuine and predicted series
ROC Area	Gives an estimate of the most ideal execution of a classifier
PRC Area	has a one-to-one association with a consistent ROC

4.3 Medical Datasets

There are the following datasets used to analyze the results. All the datasets are collected from GitHub and Kaggle websites.

4.3.1 Brest Cancer:

Data set 1: breast-cancer-master

We take Breast cancer Wiscon (original) dataset which is created by Dr. WIlliam H. Wolberg (physician) University of Wisconsin Hospitals Madison, Wisconsin, USA. With the following 10 attributes are: age, mefalsepause, tumour-size, inv-falsedes, falsede-caps, deg-malig, breast, breast-quad, irradiate class. With 278 no. of rows.

Data Set Characteristics:	Multivariate	**Number of Instances:**	699	**Area:**	Life
Attribute Characteristics:	Integer	**Number of Attributes:**	10	**Date Donated**	1992-07-15
Associated Tasks:	Classification	**Missing Values?**	Yes	**Number of Web Hits:**	522123

Table 8: Breast cancer Wiscon

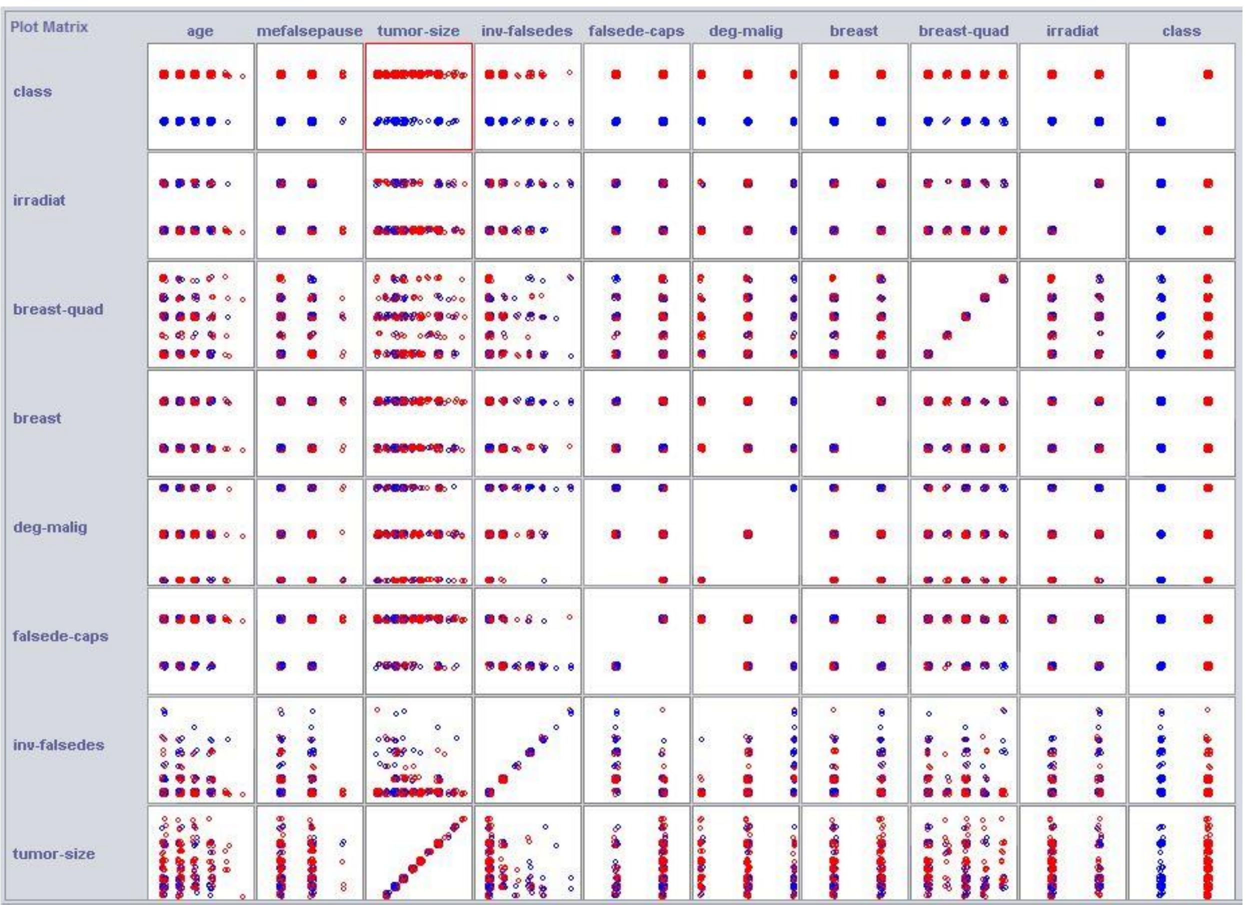


Figure 10: Breast cancer Wiscon (original) dataset

Data Set 2: breast-cancer-prediction-dataset:

Worldwide, breast cancer is the most common type of cancer in women and the second highest in terms of mortality rates. Diagnosis of breast cancer is performed when an abnormal lump is found (from self-examination or x-ray) or a tiny speck of calcium is seen (on an x-ray). After a suspicious lump is found, the doctor will conduct a diagnosis to determine whether it is cancerous and, if so, whether it has spread to other parts of the body.

Dataset for AI for Social Good: Women Coders' Bootcamp organized by Artificial Intelligence for Development in collaboration with UNDP Nepal. This breast cancer dataset was obtained from the University of Wisconsin Hospitals, Madison from Dr. William H. Wolberg. With 6 attributes mean_radius, mean_texture, mean_perimeter, mean_area, mean_smoothness, and diagnosis (false-recurrence-events and recurrence-events).

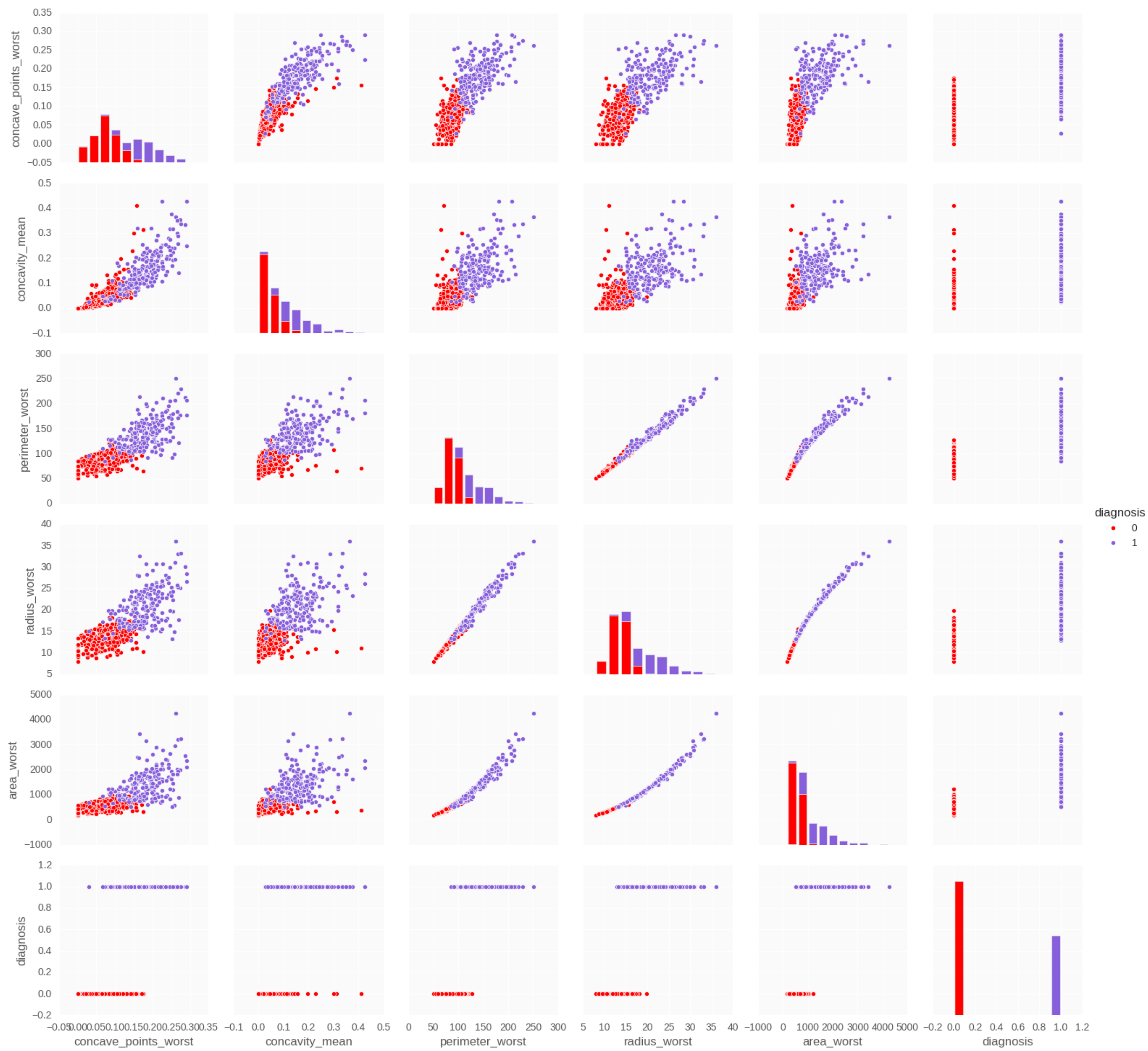


Figure 11: The breast-cancer-prediction-dataset attributes

4.3.2 Dermatology

Dataset 1: Dermatology Data Set:

This database contains 34 attributes, 33 of which are linear valued and one of them is nominal.

Data Set Characteristics:	Multivariate	Number of Instances:	366	Area:	Life
Attribute Characteristics:	Categorical, Integer	Number of Attributes:	33	Date Donated	1998-01-01
Associated Tasks:	Classification	Missing Values?	Yes	Number of Web Hits:	197062

Table 9: Dermatology Data Set Detail

Attribute Information:

Clinical Attributes: It takes values 0, 1, 2, 3, unless otherwise indicated. The attributes are erythema, scaling, definite borders, itching, koebner phenomenon, polygonal papules, follicular papules, oral mucosal involvement, knee and elbow involvement, scalp involvement, family history, (0 or 1) and Age (linear)

Histopathological Attributes: It also take values 0, 1, 2, 3. The attributes are melanin incontinence, eosinophils in the infiltrate, PNL infiltrate, fibrosis of the papillary dermis, exocytosis, acanthosis, hyperkeratosis, parakeratosis, clubbing of the rete ridges, elongation of the rete ridges, thinning of the supra papillary epidermis, spongiform pustule, munro microabcess, focal hyper granulosis, disappearance of the granular layer, vacuolization and damage of basal layer, spongiosis, saw-tooth appearance of rates, follicular horn plug, perifollicular parakeratosis, inflammatory mononuclear infiltrate and band-like infiltrate.

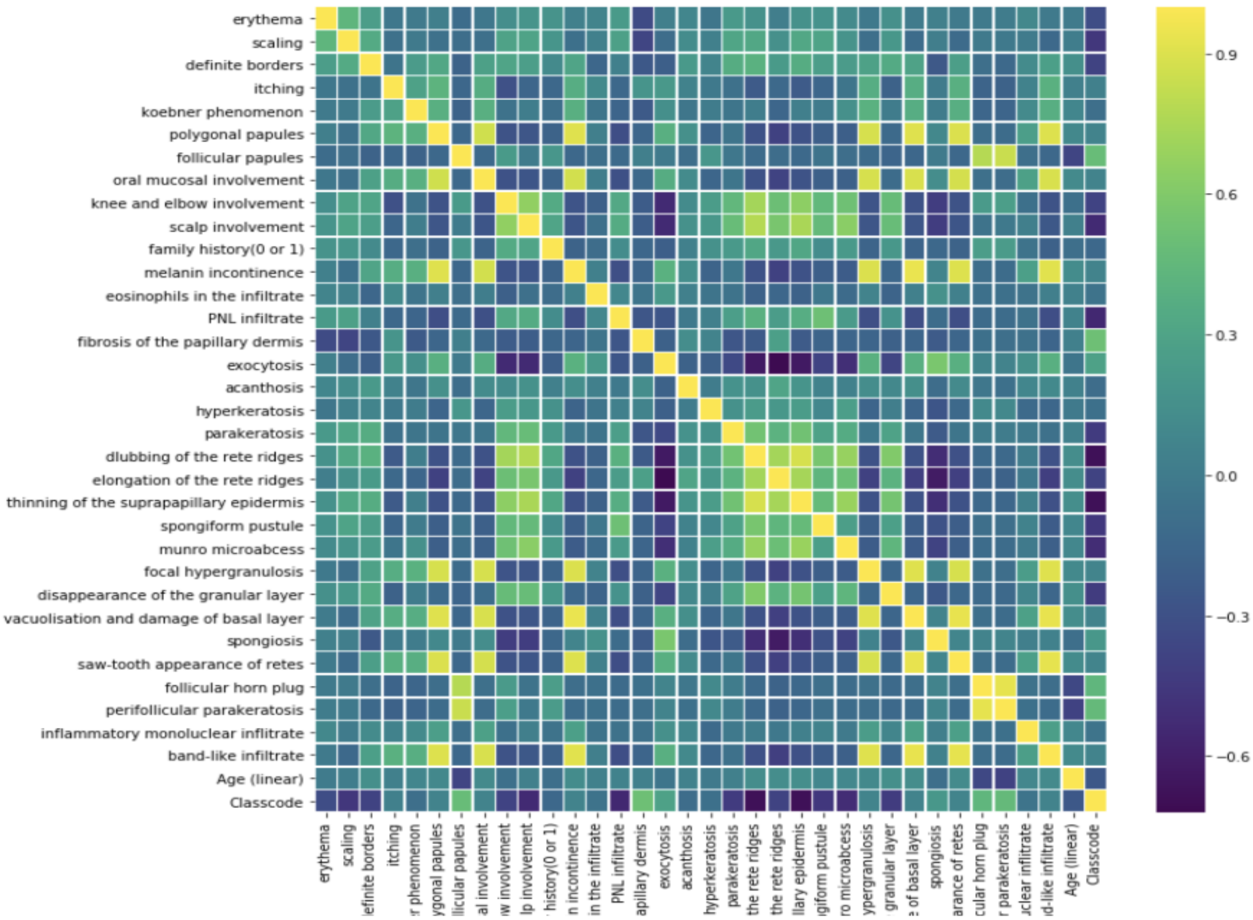


Figure 12: Dermatology Data Set Plots

Dataset 2: Dermograpic disease control dataset

We took demograpic disease control dataset from the National Health and Nutrition Examination Survey (NHANES) is a program of studies designed to assess the health and nutritional status of adults and children in the United States. NCHS is part of the Centers for Disease Control and Prevention (CDC) and has the responsibility for producing vital and health statistics for the Nation. The 47 attributes are used for conducting this experiment.

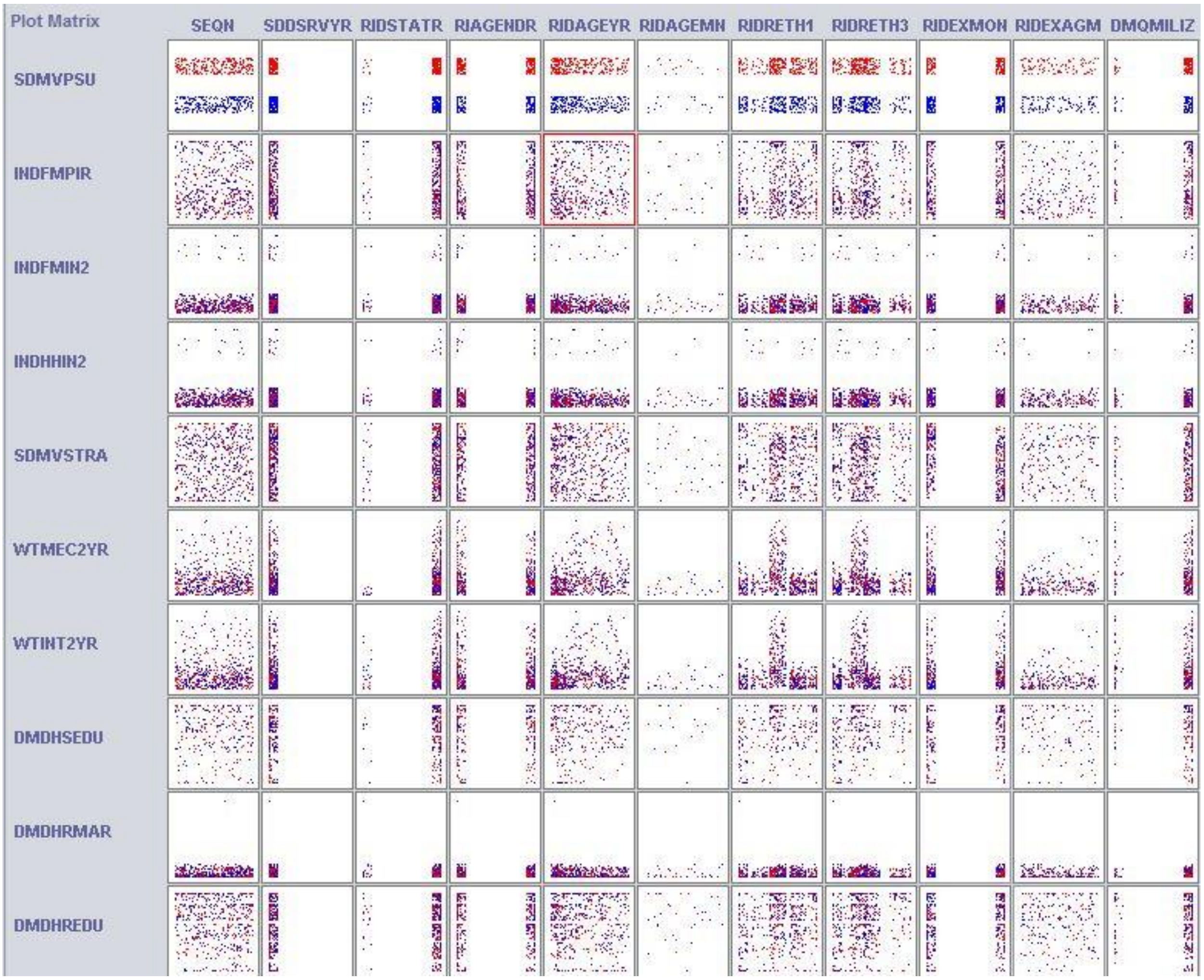


Figure 13: Dermograpic disease control dataset

4.3.3 Diabetes

The Data mining testing algorithm is applied to the PIAM Indian diabetic database to predict the accuracy in data classification.

This dataset is originally from the National Institute of Diabetes and Digestive and Kidney Diseases. The objective is to predict based on diagnostic measurements whether a patient has diabetes. Several constraints were placed on the selection of these instances from a larger database. In particular, all patients here are females at least 21 years old of Pima Indian heritage.

- Pregnancies: Number of times pregnant
- Glucose: Plasma glucose concentration 2 hours in an oral glucose tolerance test
- Blood Pressure: Diastolic blood pressure (mm Hg)
- Skin Thickness: Triceps skin fold thickness (mm)
- Insulin: 2-Hour serum insulin (mu U/ml)
- BMI: Body mass index (weight in kg/ (height in m) ^2)
- Diabetes Pedigree Function: Diabetes pedigree function
- Age: Age (years)

- Outcome: Class variable (0 or 1)

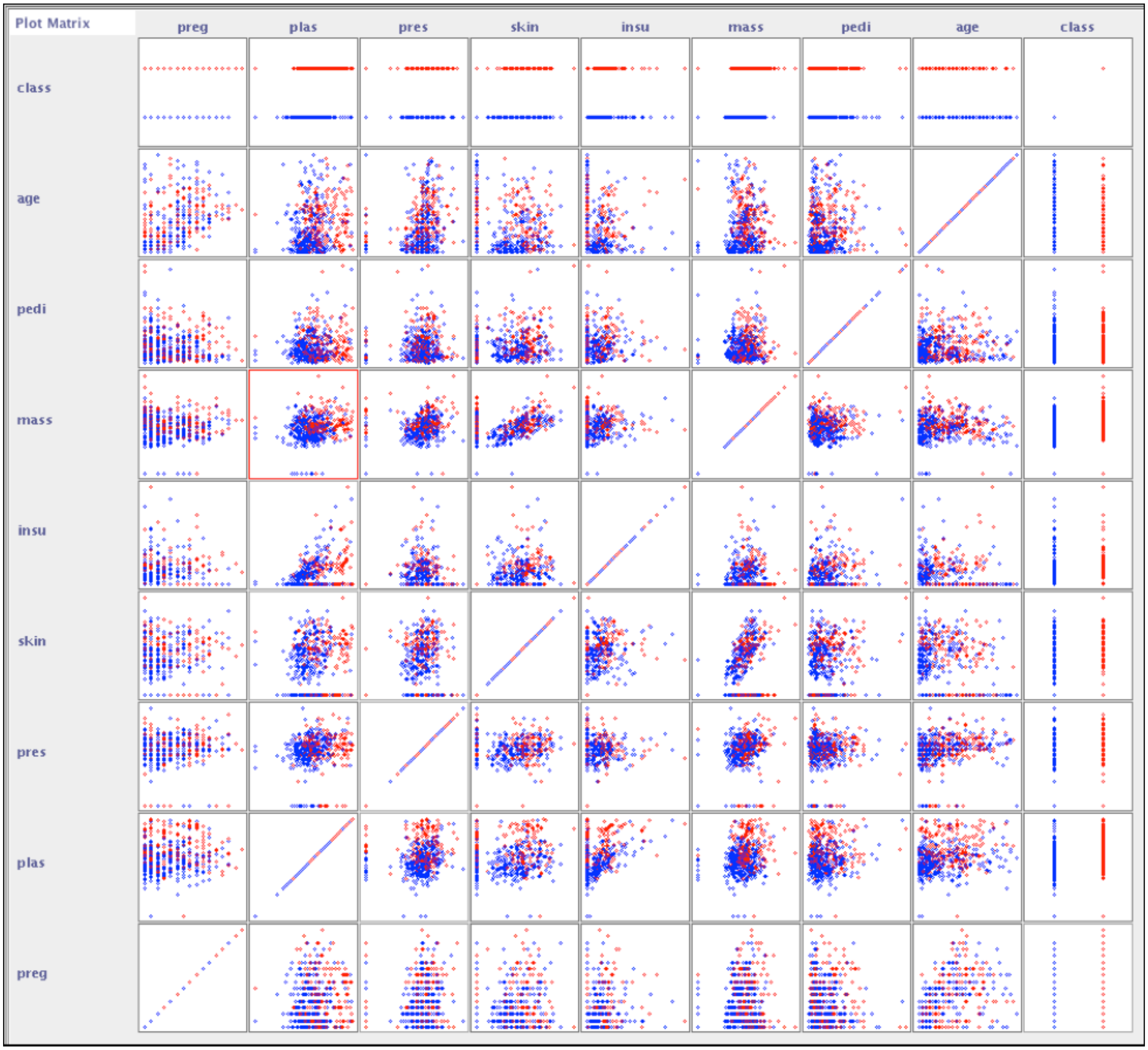


Figure 14: Diabetes the Data Set the blue color shows abnormal and red present normal

4.3.4 Iris

Iris it has following attribute Sepal Width and Length, Peta Length and Width. Irvine, CA University of California, School of Information and Computer Science. Data set containing values for six values into 3 classes (normal, disk hernia or spondilolysthesis) or 2 classes (normal or abnormal). Each patient is represented in the data set by six biomechanical attributes derived from the shape and orientation of the pelvis and lumbar spine (in this order): pelvic incidence, pelvic tilt, lumbar lordosis angle, sacral slope, pelvic radius and grade of spondylolisthesis. The following convention is used for the class labels: DH (Disk Hernia), Spondylolisthesis (SL), Normal (NO) and Abnormal (AB).

Data Set Characteristics:	Multivariate	**Number of Instances:**	310	**Area:**	N/A
Attribute Characteristics:	Real	**Number of Attributes:**	6	**Date Donated**	2011-08-09
Associated Tasks:	Classification	**Missing Values?**	N/A	**Number of Web Hits:**	155772

Table 10: Iris dataset

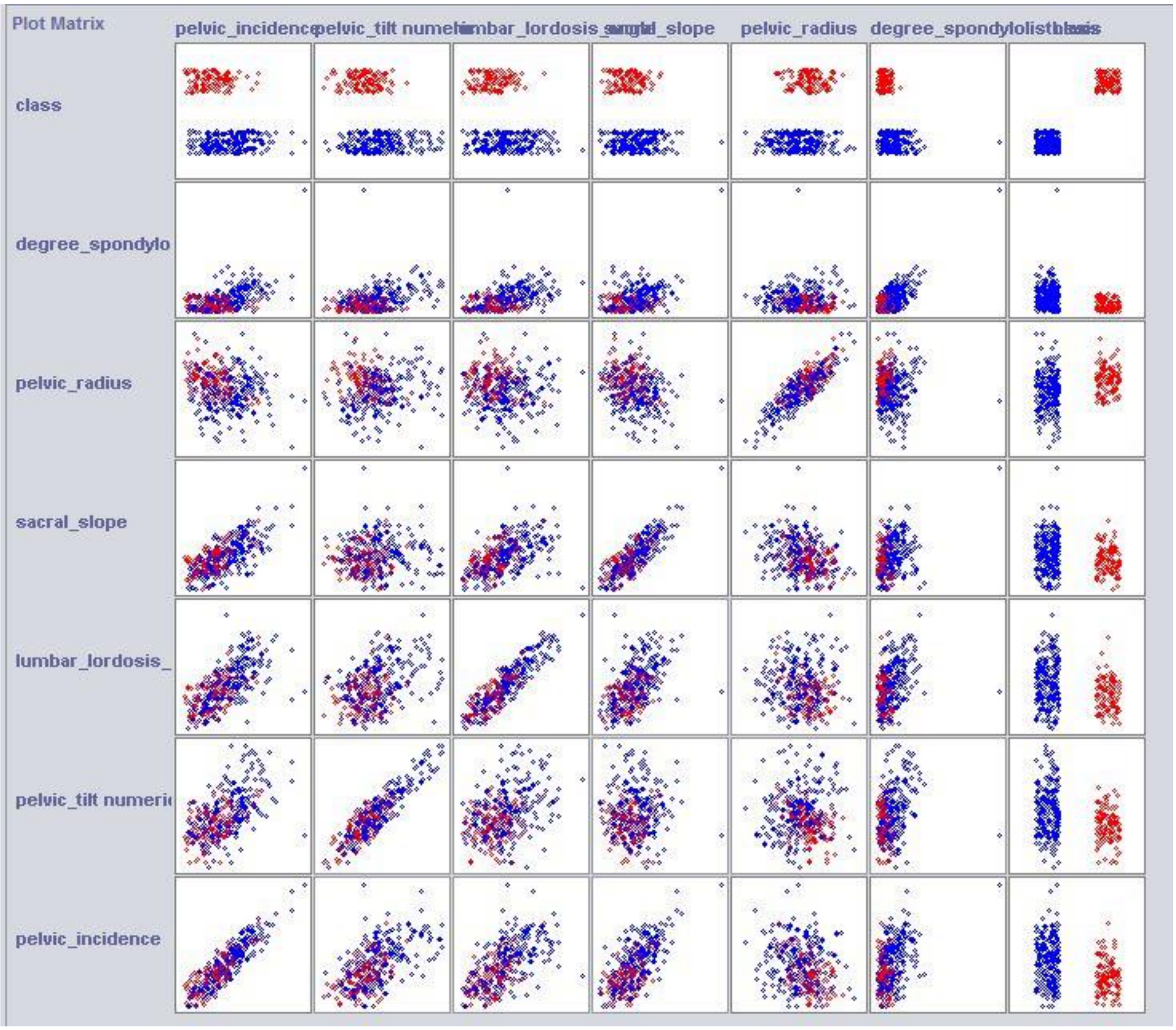


Table 11: Iris Dataset

4.3.5 Hypothyroid

we have two hypothyroid datasets with around 3772 instances for the Hypothyroid.

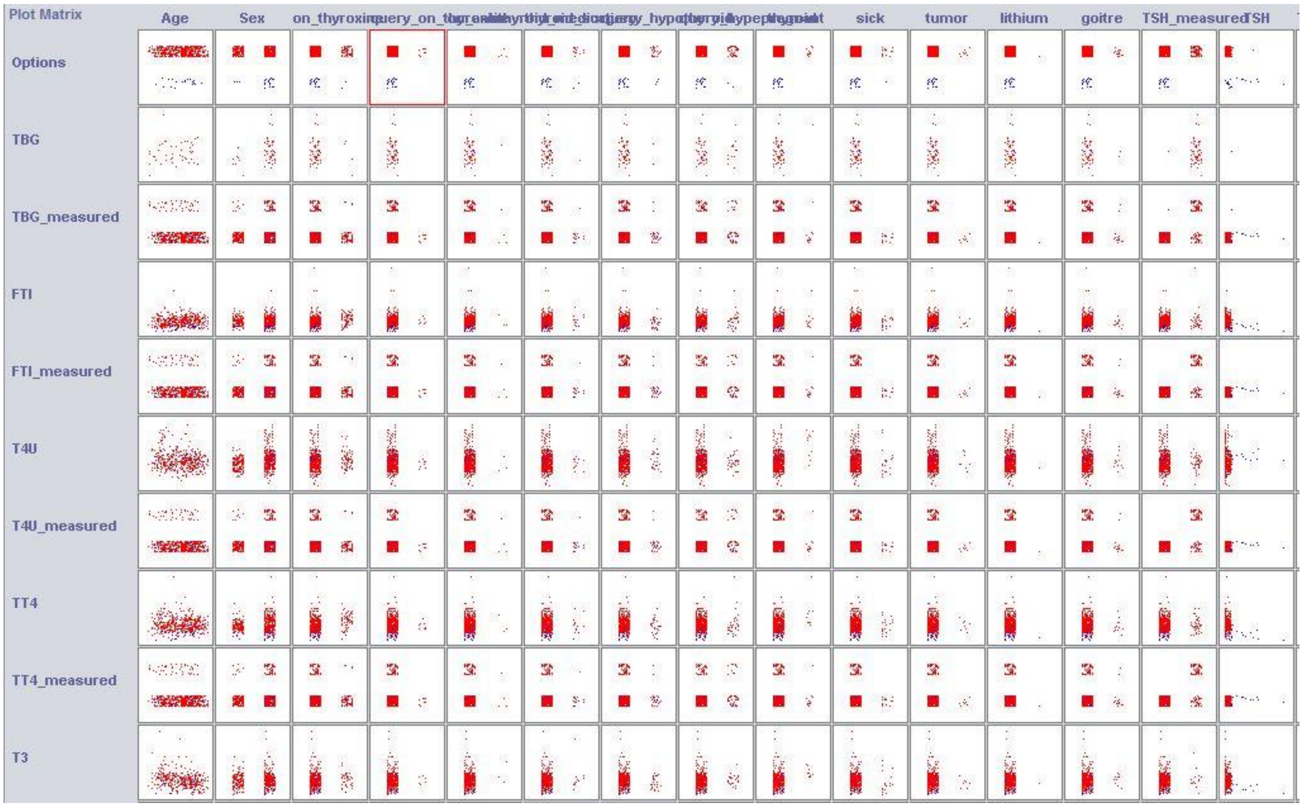


Table 12 : Dataset – 1 hypothyroid

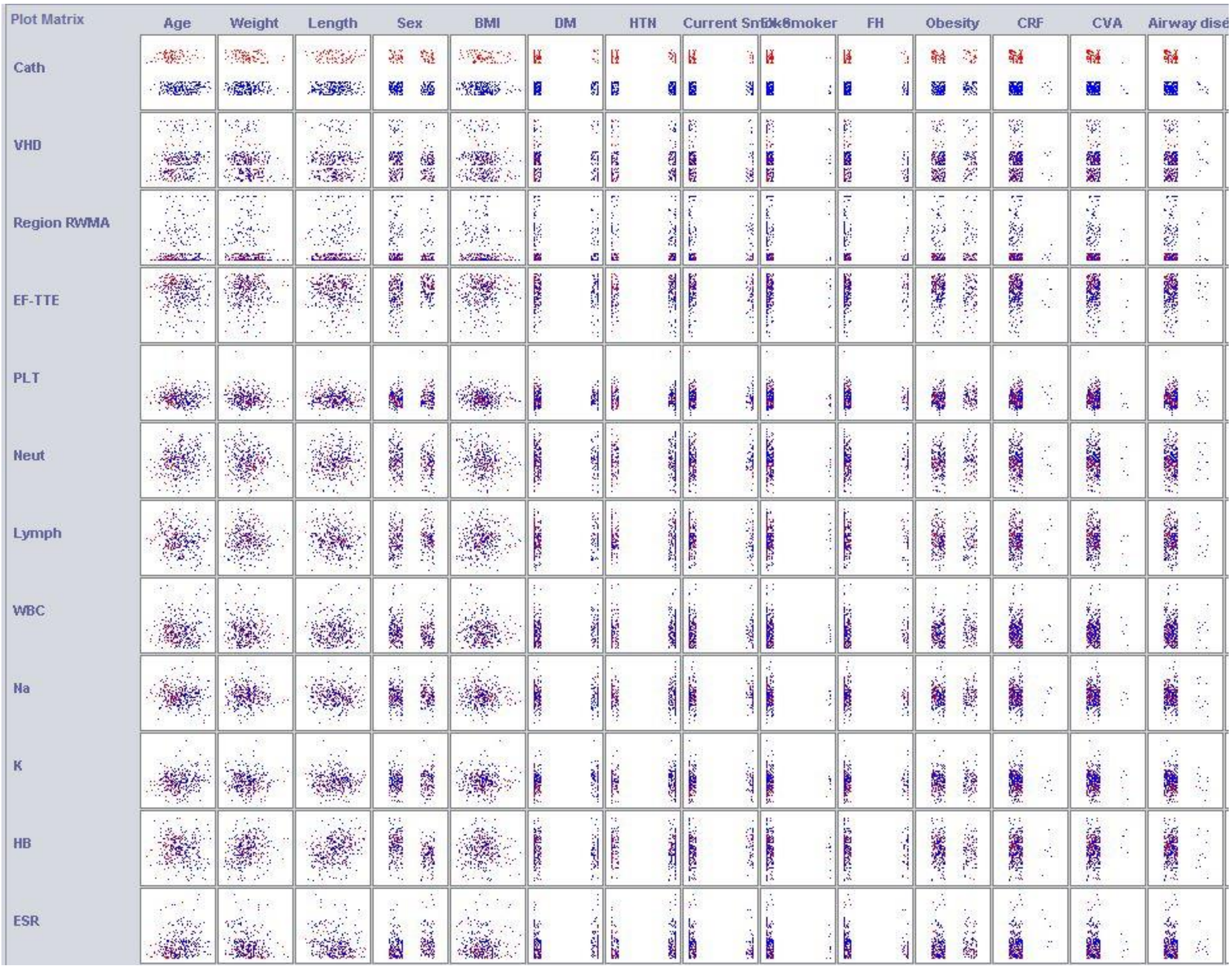


Table 13: Dataset – 2 Hypo-thyroid-Prediction

4.4 Summery

The various analytic technique and algorithms used to increase data quality. In the healthcare sector huge amounts of data such as patient information, disease information, treatment data and many more store in electronic health care records. The collection of various diseases datasets and apply data mining technique for data classification, predict sensitive data and provides the proper complete information for removing the barrier from health care record identify the accuracy between dataset by applying different classifiers J48 (decision-tree), NB (Naïve-Bayes) and MLP (Multi-Layer-Perception). Confusion and accuracy matrix based on 20, 50, 80-fold cross-validation method. The experiment conducts on diabetes, Brest cancer, iris, hypothyroid, and dermatology data sets. Data mining tool WEKA is used to conduct all the experiments. The results display that the fusion of Naïve-Bayes, MLP and J48 used in classification and describe which classifier approach is suitable for each dataset.

CHAPTER 5

RESULTS BASED ON CLASSIFIERS

The experiment performs on the WEKA data mining tool by taking five medical diseases and 2 data sets of each disease. Applying J48, MLP and NB classifier upon for 80 folds, 50 folds, 20 folds' cross-validation test option and evaluate the accuracy, kappa, MAN, Root absolute error, root relative squared error, the root-relative absolute error of the following medical data sets.

5.1 Results based on Classifiers upon data sets presents in the form of weka

1. <u>Data Set: Brest Cancer</u>

Table shows Accuracy, RAE and RRSE: -

Classifier	Data Ratio	Accuracy	RAE	RRSE
	80	75.52%	88.87%	95.26%
J48	50	**75.87%**	88.31%	94.47%
	20	75.17%	**89.11%**	96.24%
	80	72.38%	78.51%	99.84%
Bayesian	50	72.03%	78.35%	99.65%
	20	71.68%	78.68%	99.92%
	80	68.53%	79.67%	115.60%
MLP	50	67.13%	81.27%	**116.92%**
	20	68.88%	75.81%	113.20%

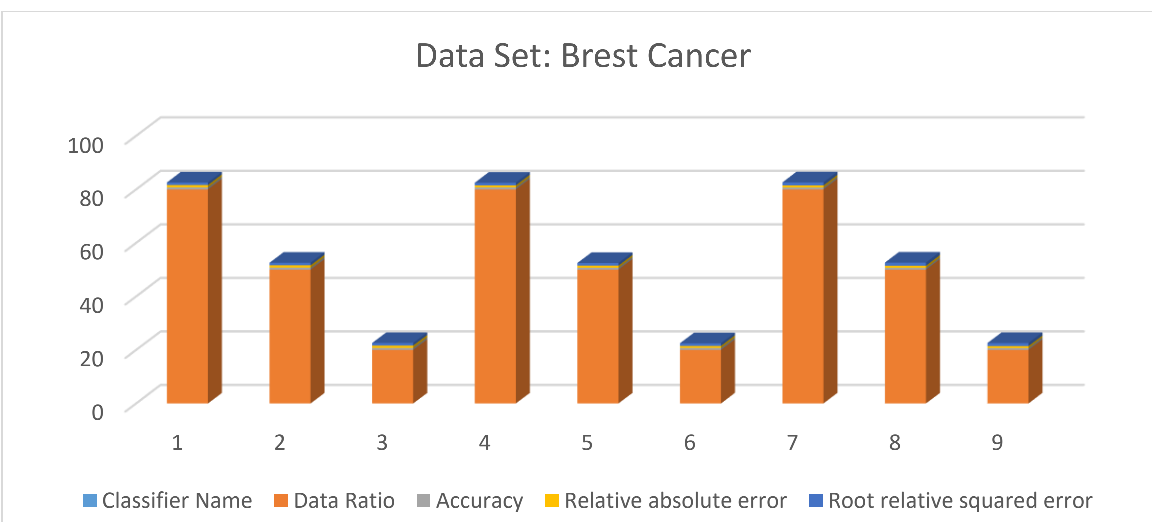


Table Show kappa, MAE and RMSE: -

Classifier Name	Data Ratio	Kappa	MAE	RMSE
J48	80	0.2826	0.3719	0.4355
	50	0.2899	0.3695	0.4319
	20	0.2754	**0.3728**	0.4399
Bayesian	80	**0.2932**	0.3285	0.4564
	50	0.2919	0.3278	0.4556
	20	0.2805	0.3292	0.4567
MLP	80	0.2363	0.3334	0.5285
	50	0.1798	0.3400	**0.5345**
	20	0.2369	0.3172	0.5174

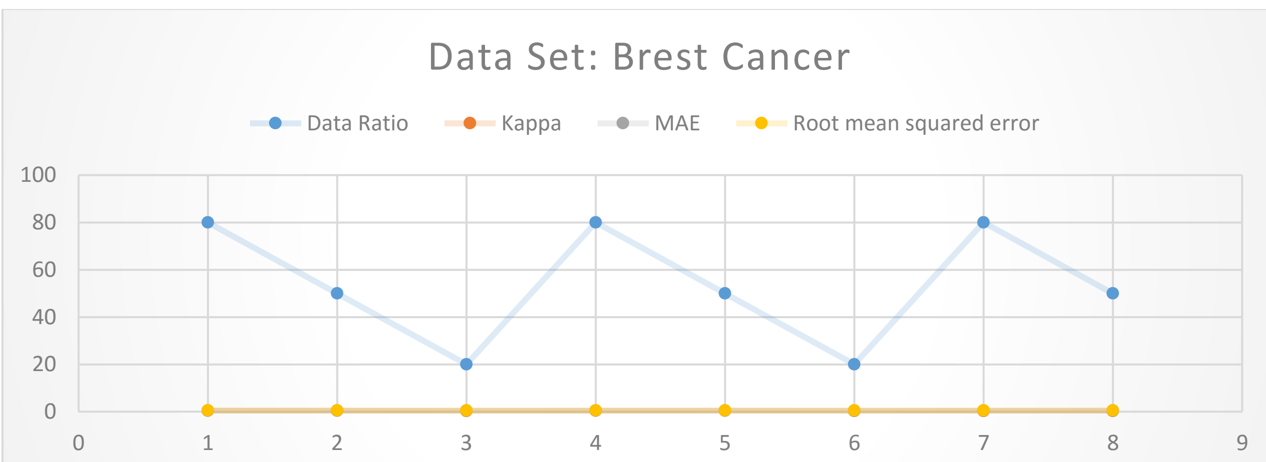
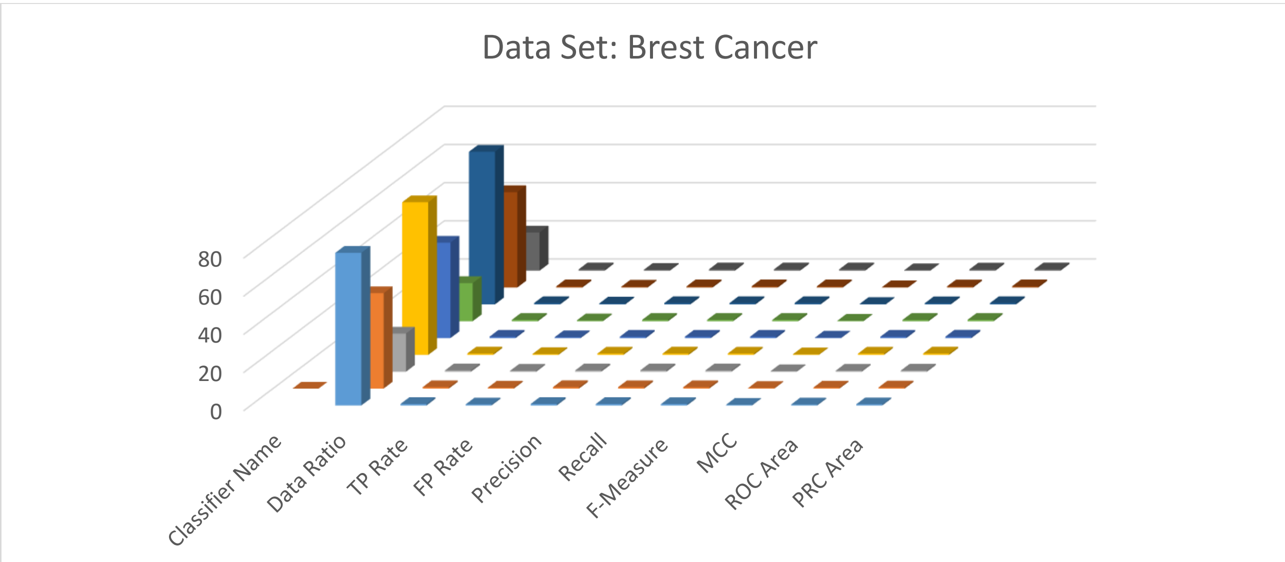


Table present TP and FP Rate, Precision, Recall, F-measure, MCC, ROC and PRC

Classifier Name	Data Ratio	TP	FP	Precision	Recall	F-measure	MCC	ROC	PRC
J48	80	0.755	0.524	0.752	0.755	0.713	0.339	0.498	0.587
	50	**0.759**	0.523	**0.760**	**0.759**	**0.716**	**0.352**	0.541	0.614
	20	0.752	**0.526**	0.745	0.752	0.710	0.327	0.551	0.624
	80	0.724	0.450	0.708	0.724	0.713	0.298	0.698	**0.740**

Bayesian	50	0.720	0.444	0.707	0.720	0.711	0.295	0.696	0.736
	20	0.717	0.452	0.702	0.717	0.707	0.284	**0.699**	**0.740**
MLP	80	0.685	0.452	0.681	0.685	0.683	0.236	0.627	0.691
	50	0.671	0.499	0.658	0.671	0.664	0.181	0.630	0.691
	20	0.689	0.457	0.682	0.689	0.685	0.237	0.646	0.698



2. Data Set: Diabetes

Table shows Accuracy, RAE and RRSE: -

Classifier Name	Data Ratio	Accuracy	RAE	RRSE
J48	80	75.39%	65.79%	88.94%
	50	75.78%	66.40%	90.22%
	20	75.13%	**68.37%**	**92.44%**
Bayesian	80	75.65%	64.03%	86.01%
	50	**76.04%**	64.18%	86.27%
	20	**76.04%**	64.84%	86.95%
MLP	80	74.87%	65.05%	89.21%
	50	75.13%	65.18%	88.33%

	20	74.87%	65.05%	89.21%

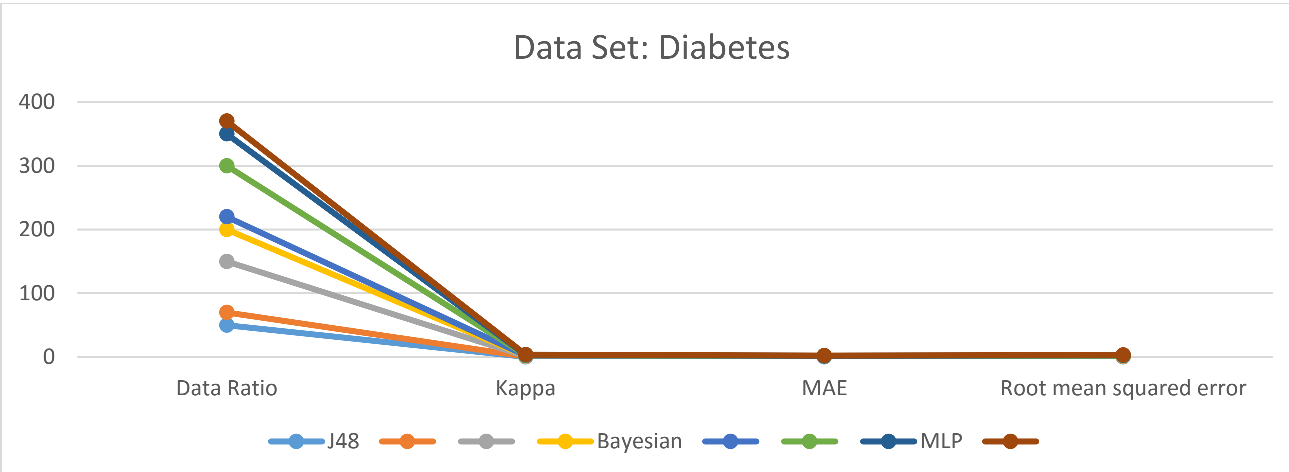


Table Show kappa, MAE and RMSE: -

Classifier Name	Data Ratio	Kappa	MAE	RMSE
J48	80	0.4474	0.2990	0.4239
	50	0.4585	0.3018	0.4301
	20	0.4366	**0.3108**	**0.4406**
Bayesian	80	0.4599	0.2910	0.4100
	50	0.4681	0.2917	0.4112
	20	**0.4700**	0.2947	0.4144
MLP	80	0.4484	0.2957	0.4252
	50	0.4455	0.2963	0.4210
	20	0.4484	0.2957	0.4252

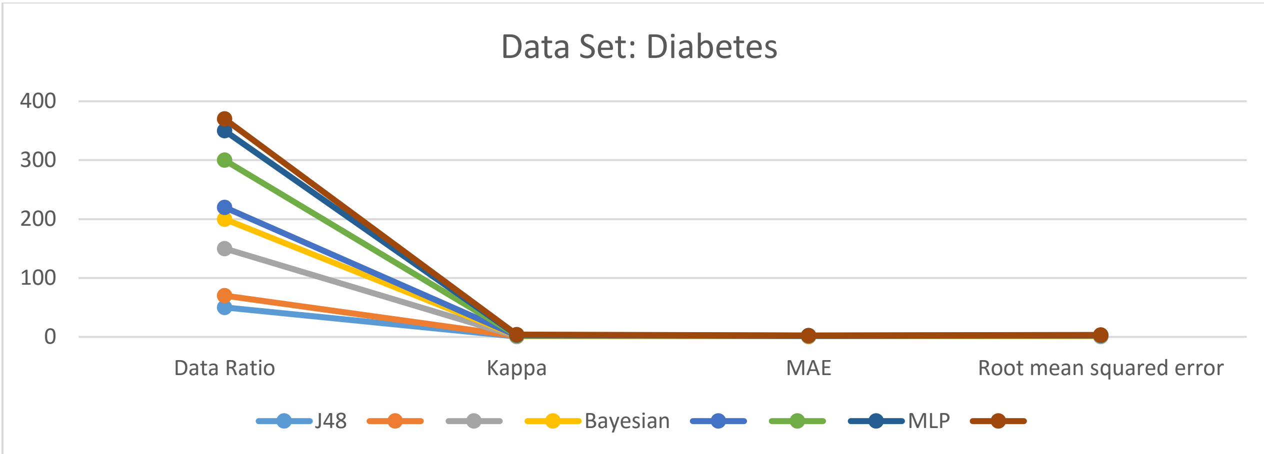
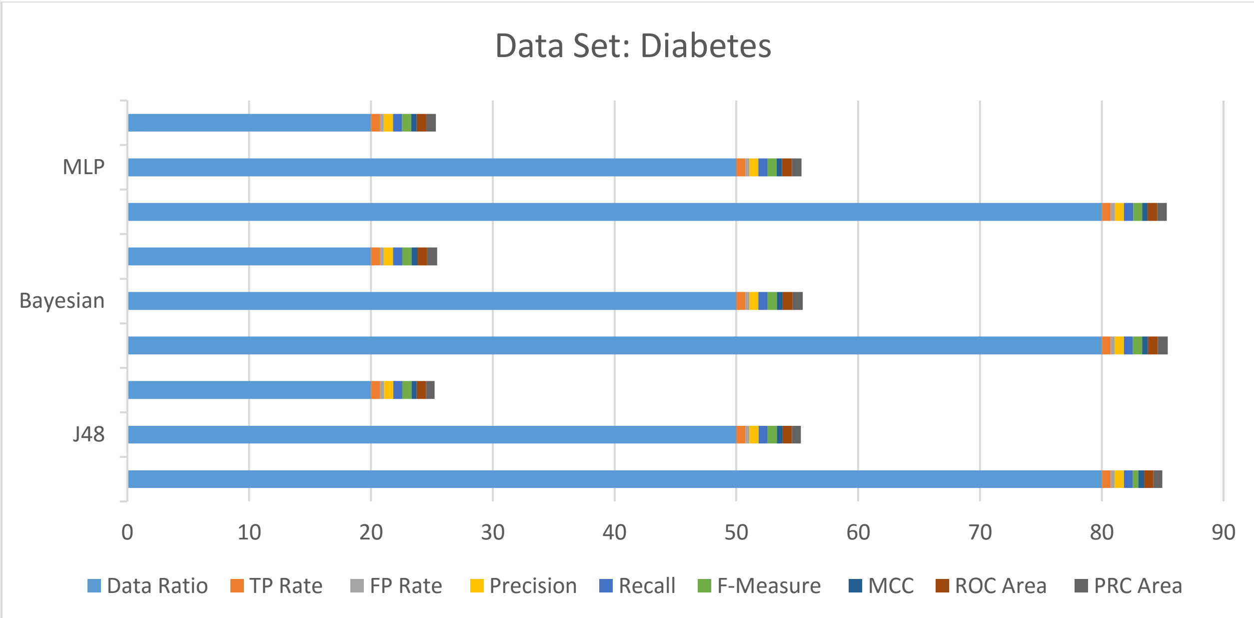


Table present TP and FP Rate, Precision, Recall, F-measure, MCC, ROC and PRC

Classifier Name	Data Ratio	TP	FP	Precision	Recall	F-measure	MCC	ROC	PRC
J48	80	0.754	0.315	0.750	0.751	0.448	0.448	0.767	0.750
	50	0.758	0.306	0.755	0.758	0.756	0.459	0.769	0.747
	20	0.751	**0.327**	0.746	0.751	0.747	0.439	0.748	0.721
Bayesian	80	0.757	0.300	0.755	0.757	0.756	0.460	**0.819**	**0.825**
	50	**0.760**	0.296	**0.759**	**0.760**	0.759	0.468	0.817	0.823
	20	**0.760**	0.293	**0.759**	**0.760**	**0.760**	**0.470**	0.811	0.816
MLP	80	0.751	0.325	0.746	0.751	0.747	0.439	0.795	0.782
	50	0.751	0.312	0.748	0.751	0.750	0.446	0.798	0.795
	20	0.749	0.299	0.749	0.749	0.749	0.448	0.793	0.781



3. Data Set: iris.2D

Table shows Accuracy, RAE and RRSE: -

Classifier Name	Data Ratio	Accuracy	RAE	RRSE
	80	95.33%	8.730%	36.15%
J48	50	**96.00%**	7.990%	33.59%
	20	**96.00%**	7.950%	33.62%
	80	94.67%	9.460%	**37.06%**
Bayesian	50	94.67%	9.230%	36.22%
	20	94.67%	8.750%	34.69%
	80	**96.00%**	11.66%	32.37%
MLP	50	95.33%	**12.00%**	32.83%
	20	95.33%	11.75%	32.86%

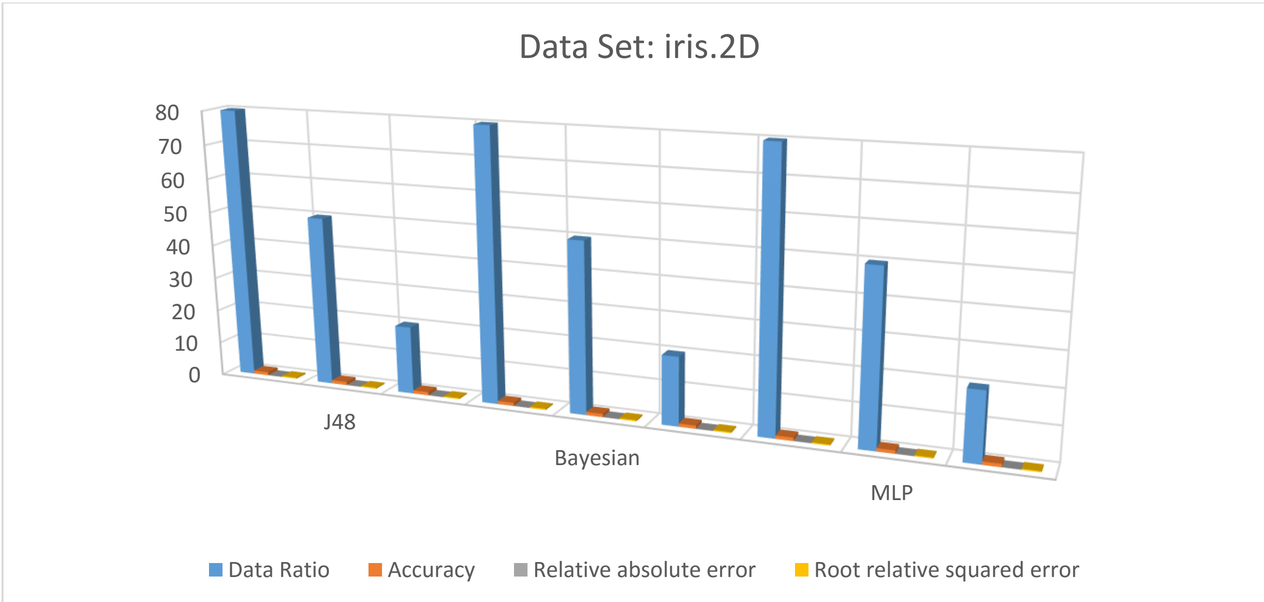


Table Show kappa, MAE and RMSE: -

Classifier Name	Data Ratio	Kappa	MAE	RMSE
J48	80	0.9300	0.0389	0.1710
	50	**0.9400**	0.0355	0.1584
	20	**0.9400**	0.0354	0.1586
Bayesian	80	0.8900	0.0422	**0.1753**
	50	0.9200	0.1707	0.1707
	20	0.9200	0.0389	0.1637
MLP	80	**0.9400**	0.0520	0.1532
	50	0.9300	**0.0533**	0.1548
	20	0.9300	0.0523	0.1551

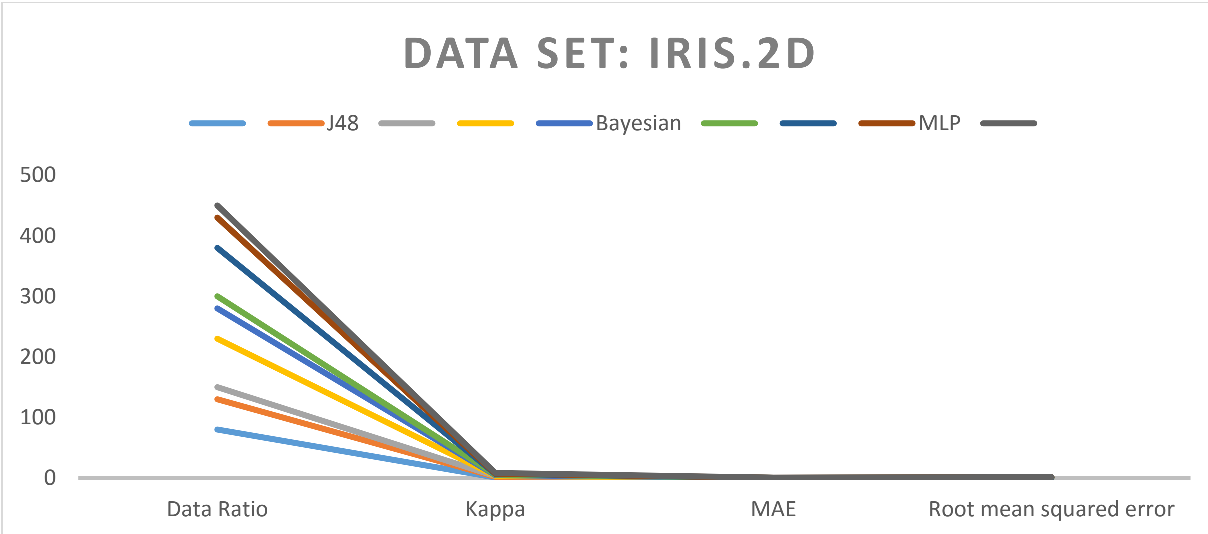
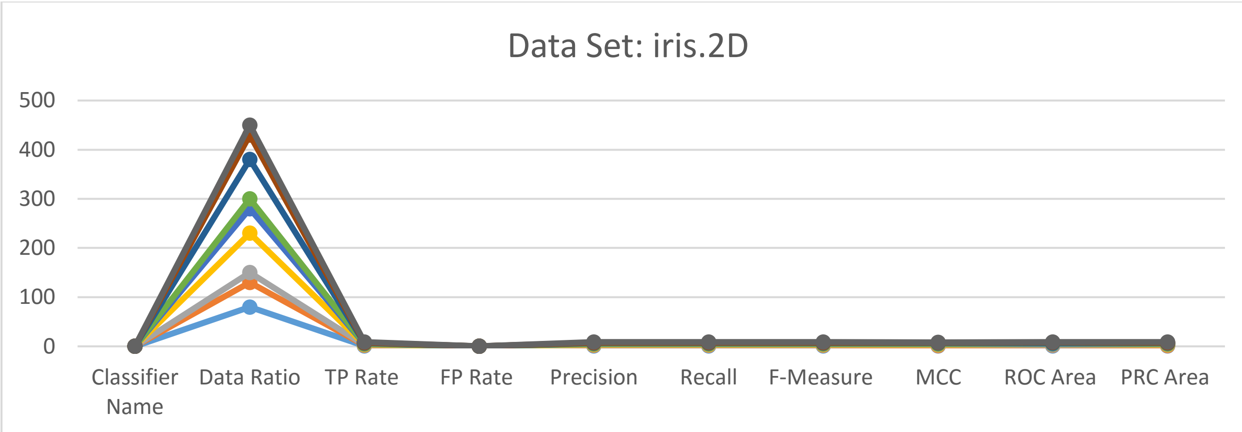


Table present TP and FP Rate, Precision, Recall, F-measure, MCC, ROC and PRC

Classifier Name	Data Ratio	TP	FP	Precision	Recall	F-score	MCC	ROC	PRC
J48	80	0.953	0.023	0.954	0.953	0.954	0.930	0.956	0.926
	50	**0.960**	0.020	**0.960**	**0.960**	**0.960**	**0.940**	0.966	0.941
	20	**0.960**	0.020	**0.960**	**0.960**	**0.960**	**0.940**	0.966	0.919
Bayesian	80	0.927	0.037	0.927	0.927	0.927	0.890	0.978	0.941
	50	0.947	**0.027**	0.947	0.947	0.947	0.920	0.978	0.950
	20	0.947	**0.027**	0.947	0.947	0.947	0.920	0.981	0.948
MLP	80	**0.960**	0.020	**0.960**	**0.960**	**0.960**	**0.940**	0.994	0.989
	50	0.953	0.023	0.953	0.953	0.953	0.930	**0.995**	**0.991**
	20	0.953	0.023	0.953	0.953	0.953	0.930	**0.995**	**0.991**



4. Data Set: hypothyroid

Table shows Accuracy, RAE and RRSE: -

Classifier Name	Data Ratio	Accuracy	RAE	RRSE
	80	99.55%	4.240%	22.49%
J48	50	99.55%	4.230%	22.48%
	20	**99.58%**	4.180%	21.71%
	80	98.44%	15.67%	40.67%
Bayesian	50	98.49%	15.23%	39.92%
	20	98.54%	15.29%	40.00%
	80	94.94%	45.22%	76.36%
MLP	50	94.51%	**47.17%**	**77.05%**
	20	96.17%	5.110%	26.15%

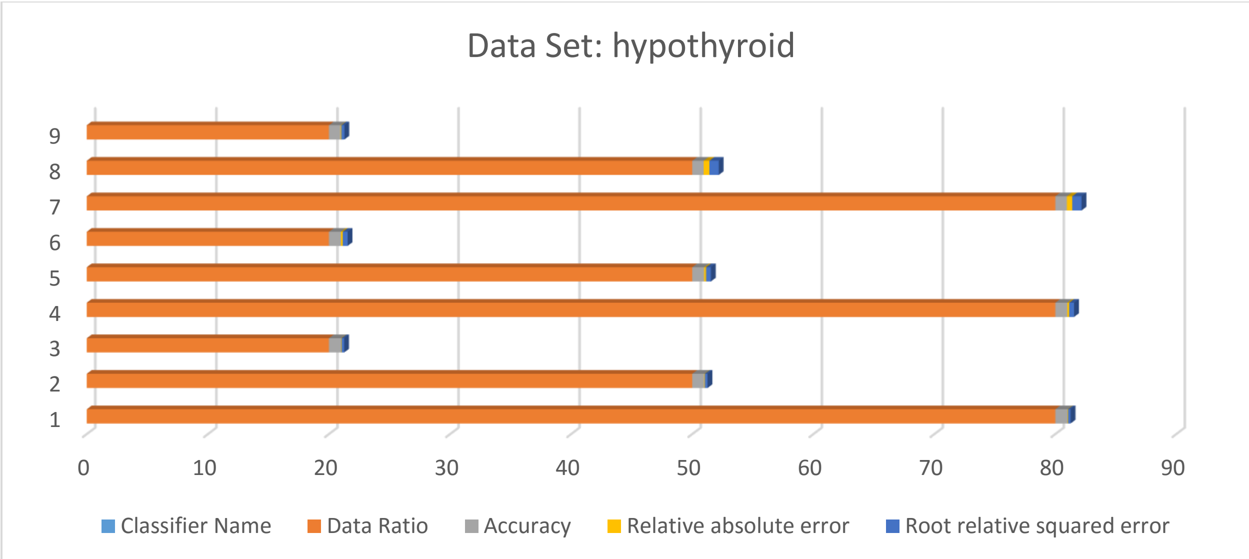


Table Show kappa, MAE and RMSE: -

Classifier Name	Data Ratio	Kappa	MAE	RMSE
	80	0.9689	0.0031	0.0428
J48	50	0.9689	0.0031	0.0428
	20	**0.9707**	0.0030	0.0414
	80	0.8906	0.0114	0.0774
Bayesian	50	0.8947	0.0111	0.0760
	20	0.8980	0.0111	0.0762
	80	0.5853	0.0329	0.1454
MLP	50	0.5636	**0.0344**	**0.1467**
	20	0.9521	0.0136	0.0955

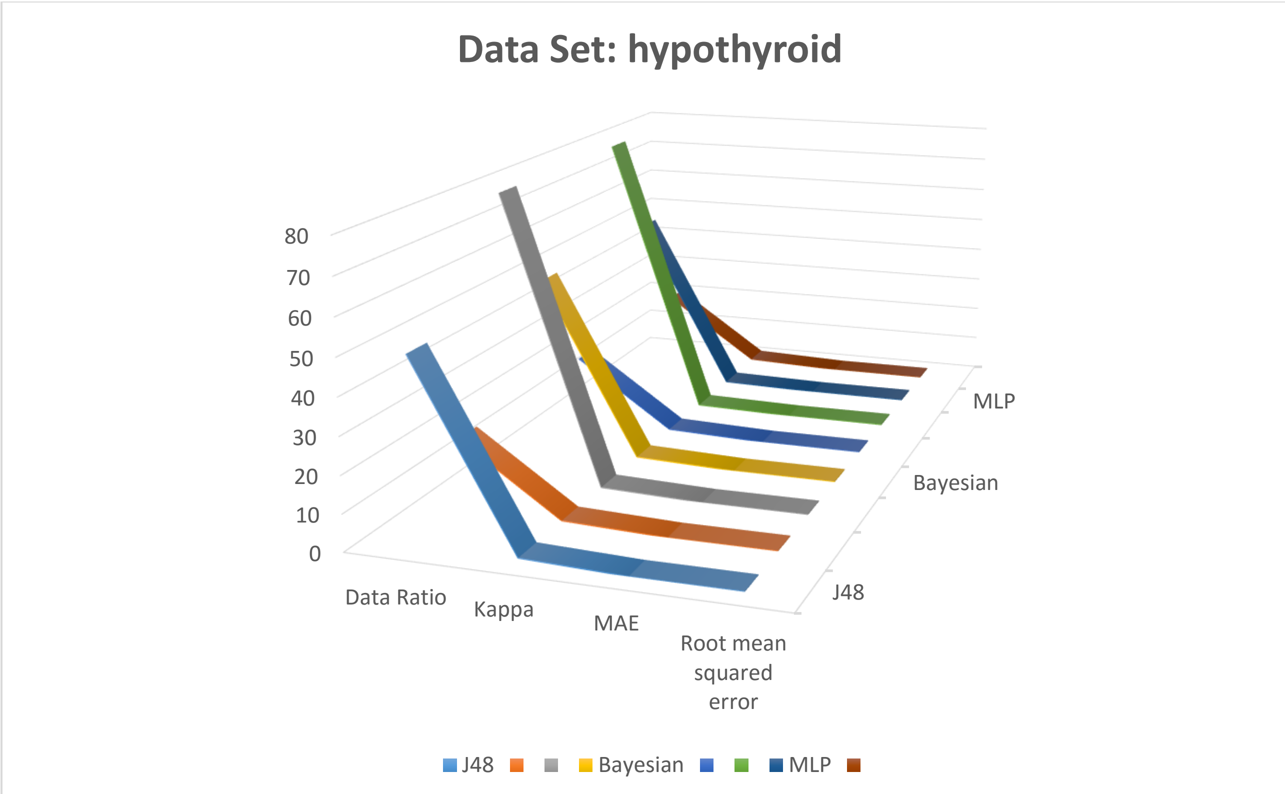
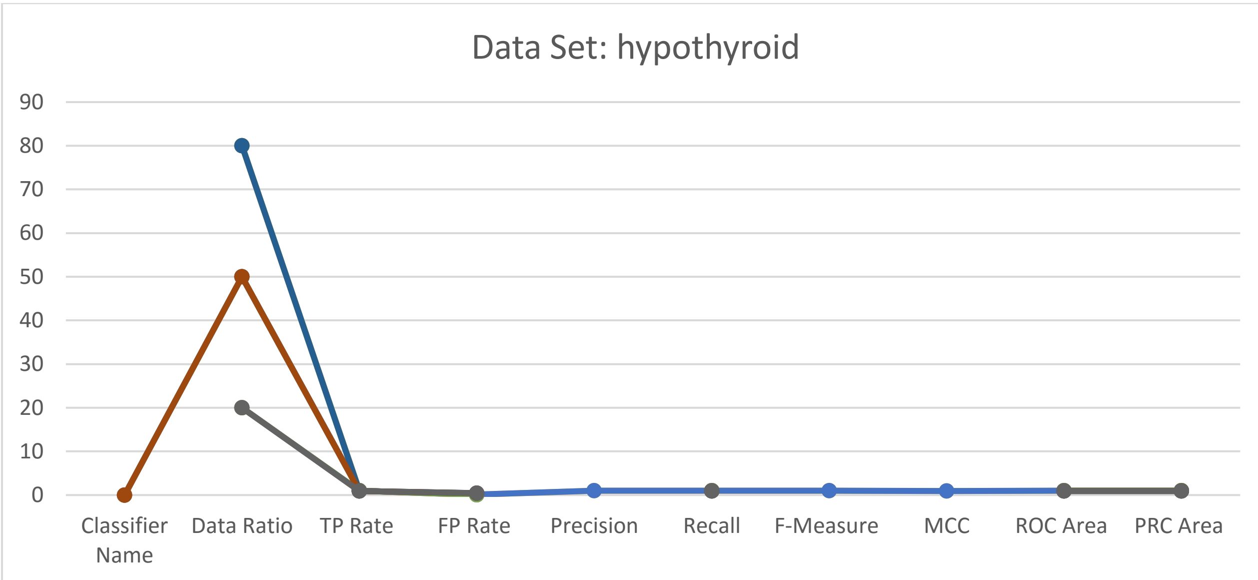


Table present TP and FP Rate, Precision, Recall, F-measure, MCC, ROC and PRC

Classifier Name	Data Ratio	TPR	FPR	Precision	Recall	F-measure	MCC	ROC	PRC
J48	80	0.995	0.019	0.000	0.995	0.000	0.000	0.993	0.995
	50	0.995	0.019	0.000	0.995	0.000	0.000	0.993	0.995
	20	**0.996**	0.019	0.000	**0.996**	0.000	0.000	0.993	**0.996**
Bayesian	80	0.984	0.102	0.984	0.984	0.984	0.895	0.996	**0.996**
	50	0.985	0.095	**0.985**	0.985	**0.985**	**0.899**	0.996	**0.996**
	20	0.985	0.095	0.000	0.985	0.000	0.000	**0.997**	**0.996**
MLP	80	0.949	0.435	0.000	0.949	0.000	0.000	0.909	0.956
	50	0.945	**0.445**	0.000	0.945	0.000	0.000	0.914	0.957
	20	0.949	**0.445**	0.000	0.949	0.000	0.000	0.919	0.959



5. Data Set: dermatology

Table shows Accuracy, RAE and RRSE: -

Classifier Name	Data Ratio	Accuracy	RAE	RRSE
J48	80	93.44%	10.13%	38.70%
	50	93.72%	9.860%	37.37%
	20	93.99%	9.640%	36.65%
Bayesian	80	**98.09%**	2.980%	17.48%
	50	97.81%	3.050%	17.72%
	20	97.81%	3.070%	17.83%
MLP	80	94.94%	**44.70%**	**75.88%**
	50	95.90%	5.460%	27.40%
	20	95.90%	5.240%	26.84%

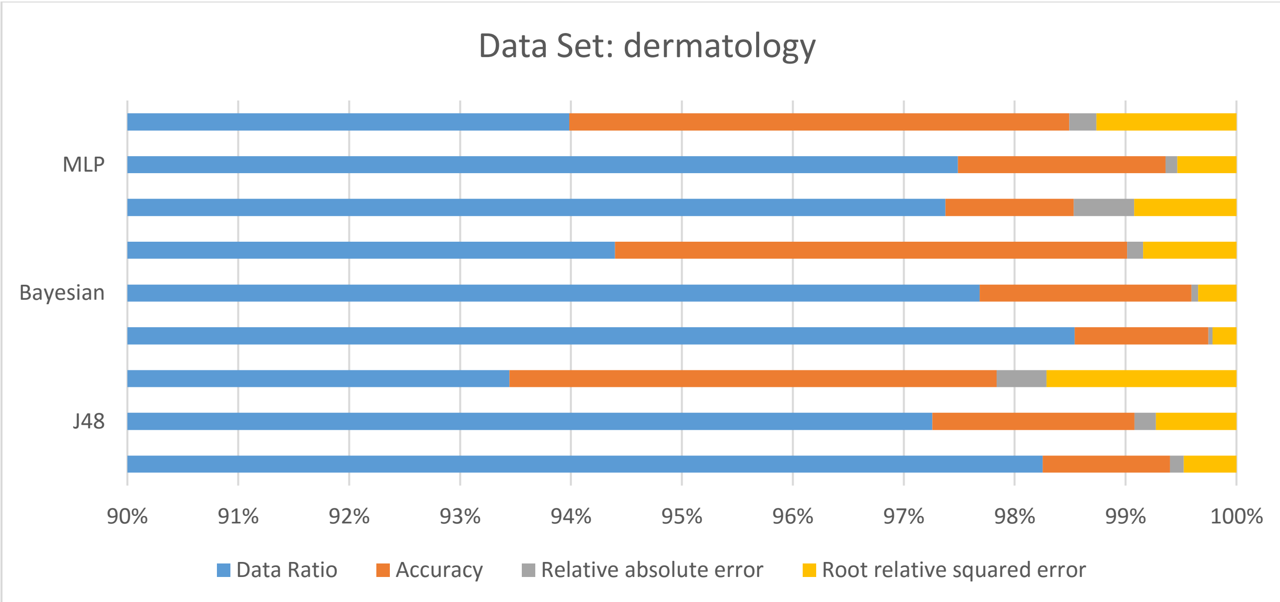


Table Show kappa, MAE and RMSE: -

Classifier Name	Data Ratio	Kappa	MAE	RMSE
J48	80	0.9180	0.0270	0.1413
	50	0.9214	0.0263	0.1364
	20	0.9247	0.0257	0.1337
Bayesian	80	**0.9761**	0.0079	0.0638
	50	0.9727	0.0081	0.0647
	20	0.9727	0.0082	0.0651
MLP	80	0.5821	**0.0326**	**0.1445**
	50	0.9487	0.0145	0.1000
	20	0.9487	0.0140	0.0980

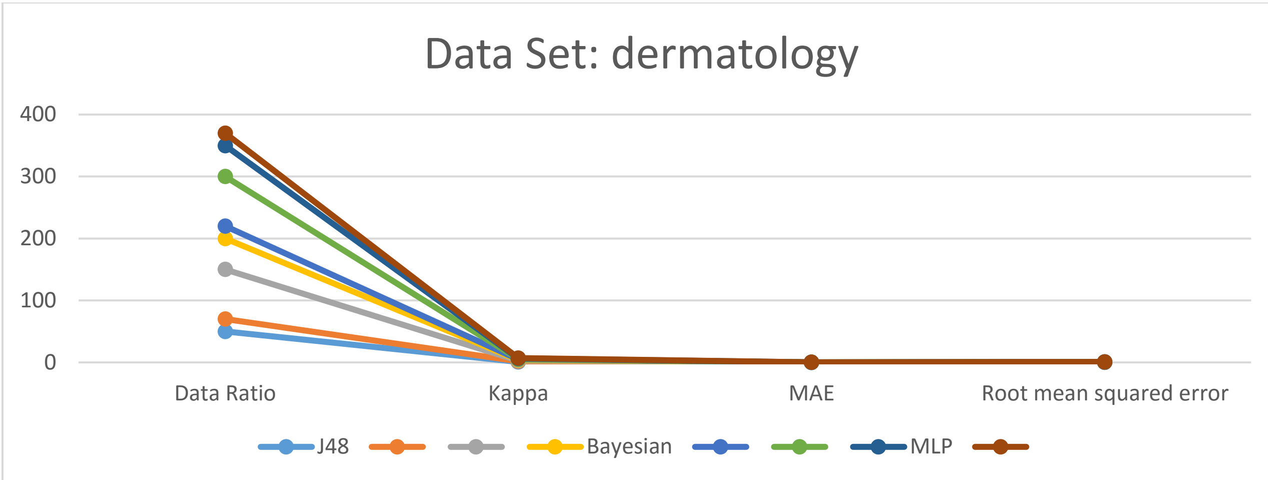
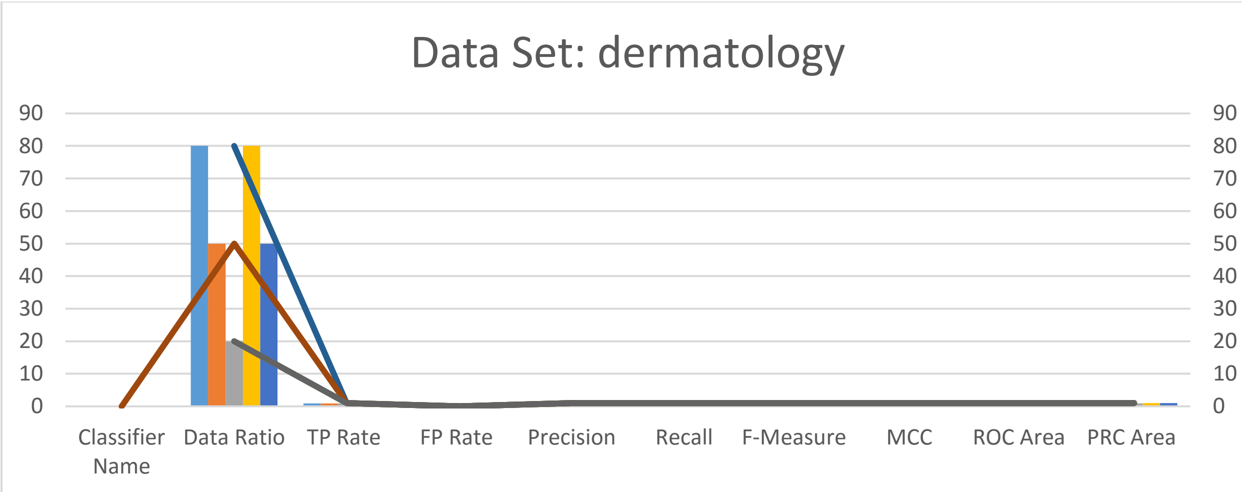


Table present TP and FP Rate, Precision, Recall, F-measure, MCC, ROC and PRC

Classifier Name	**Data Ratio**	**TPR**	FPR	Precision	**Recall**	**F-measure**	**MCC**	**ROC**	**PRC**
J48	80	0.934	**0.016**	0.934	0.934	0.934	0.918	0.974	0.927
	50	0.937	0.015	0.937	0.937	0.937	0.922	0.974	0.924
	20	0.940	0.015	0.940	0.940	0.940	0.925	0.979	0.935
Bayesian	80	**0.981**	0.003	**0.982**	**0.981**	**0.981**	**0.978**	**1.000**	**0.998**
	50	0.978	0.004	0.980	0.978	0.978	0.975	**1.000**	**0.998**
	20	0.978	0.004	0.980	0.978	0.978	0.975	**1.000**	**0.998**
MLP	80	0.962	0.008	0.962	0.962	0.962	0.954	0.998	0.993
	50	0.959	0.008	0.959	0.959	0.959	0.951	0.998	0.990
	20	0.959	0.008	0.959	0.959	0.959	0.951	0.998	0.991



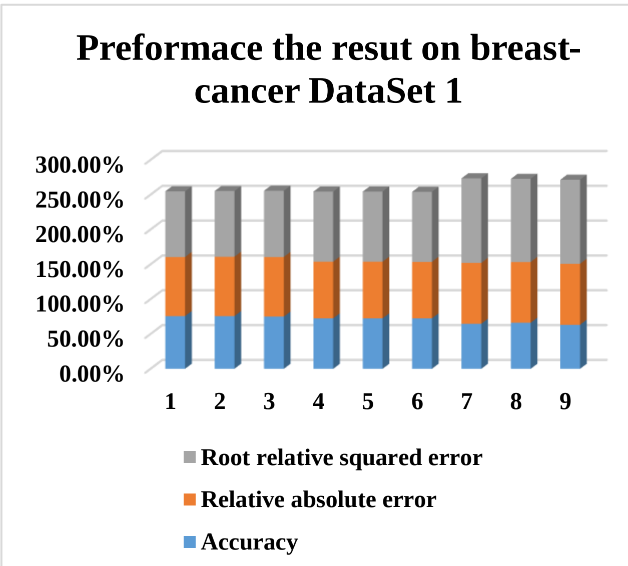
5.2 Classifier Performance of available datasets on Kaggle and GitHub

1. Brest Cancer

Table shows Accuracy, RAE and RRSE: -

Classifier Name	Data Set Name	Data Ratio	Accuracy	RAE	RRSE
J48	Data Set 1	80	**75.74%**	84.88%	94.21%
		50	**75.74%**	85.19%	94.35%
		20	75.00%	85.62%	95.09%
Bayesian	Data Set 1	80	72.43%	81.35%	100.68%
		50	72.43%	81.35%	100.78%
		20	72.43%	81.17%	100.52%
MLP	Data Set 1	80	64.71%	87.47%	**121.35%**
		50	66.18%	87.19%	119.40%
		20	63.24%	**87.58%**	120.62%
J48	Data Set 2	80	92.27%	23.79%	55.58%
		50	91.56%	**25.28%**	57.78%
		20	91.04%	24.98%	58.18%
Bayesian	Data Set 2	80	90.33%	22.22%	58.27%
		50	90.33%	22.24%	**58.32%**

		20	90.33%	22.23%	58.29%
MLP	Data Set 2	80	93.50%	19.66%	46.26%
		50	93.15%	20.32%	47.40%
		20	**93.85%**	20.28%	47.36%



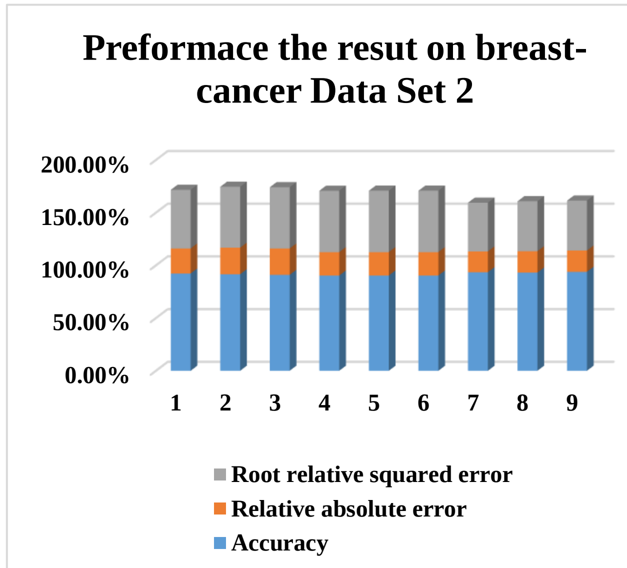
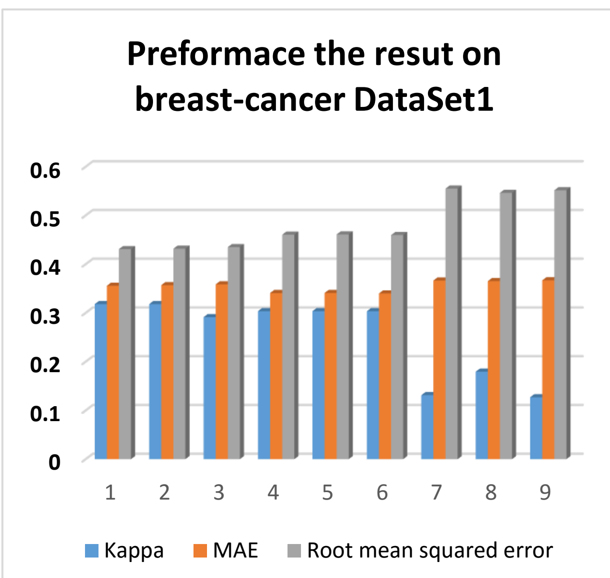


Table Show kappa, MAE and RMSE: -

Classifier Name	Data Set Name	Data Ratio	Kappa	MAE	RMSE
J48	Data Set 1	80	**0.3180**	0.3556	0.4309
		50	**0.3180**	0.3569	0.4317
		20	0.2914	0.3586	0.4349
Bayesian	Data Set 1	80	0.3036	0.3408	0.4605
		50	0.3036	0.3409	0.4611
		20	0.3036	0.3400	0.4597
MLP	Data Set 1	80	0.1314	0.3664	**0.5550**
		50	0.1796	0.3653	0.5463
		20	0.1271	**0.3669**	0.5516

J48	Data Set 2	80	0.8320	0.1113	0.2688
		50	0.8171	**0.1182**	0.2794
		20	0.8066	0.1168	0.2813
Bayesian	Data Set 2	80	0.7869	0.1039	0.2818
		50	0.7869	0.1040	**0.2820**
		20	0.7869	0.1040	**0.2820**
MLP	Data Set 2	80	0.8597	0.0920	0.2237
		50	0.8524	0.0950	0.2292
		20	**0.8673**	0.0948	0.2290



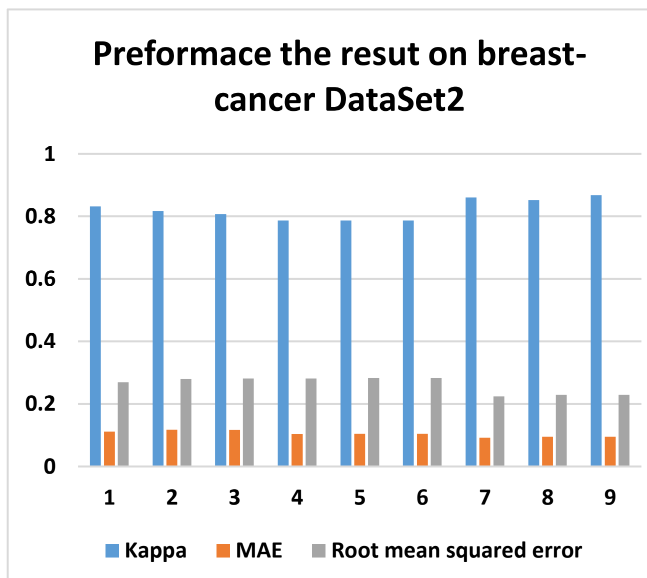
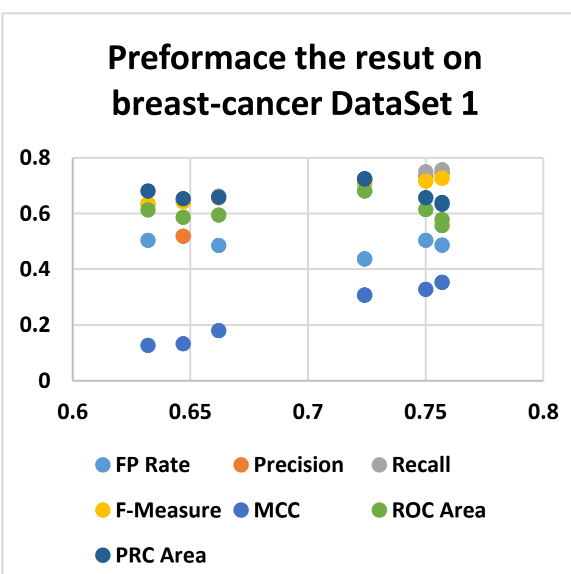
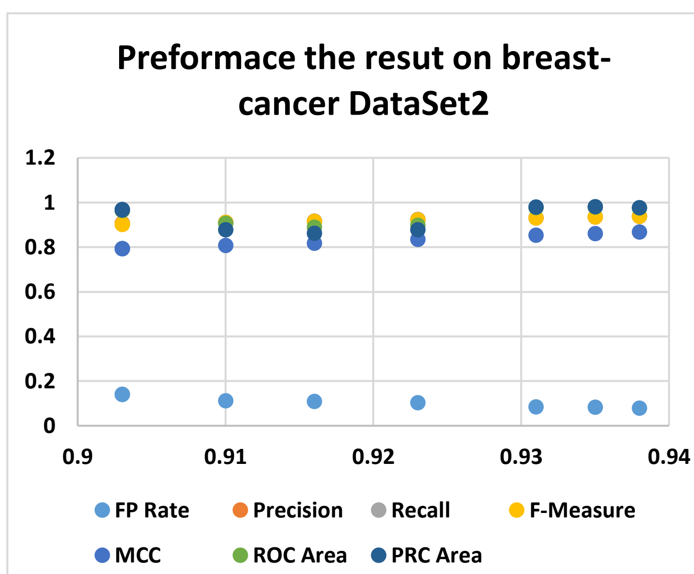


Table present TP and FP Rate, Precision, Recall, F-measure, MCC, ROC and PRC

Classifier Name	Data Sets	Data Ratio	TPR	FPR	Precision area	Recall plot	F-measure	MCC	ROC plot	PRC plot
J48	Data Set 1	80	**0.757**	0.487	**0.746**	**0.757**	**0.727**	**0.353**	0.557	0.630
		50	**0.757**	0.487	**0.746**	**0.757**	**0.727**	**0.353**	0.579	0.640

		20	0.750	0.504	0.736	0.750	0.716	0.328	0.614	0.660
Bayesian	Data Set 1	80	0.724	0.437	0.711	0.724	0.715	0.307	0.681	**0.730**
		50	0.724	0.437	0.711	0.724	0.715	0.307	**0.682**	**0.730**
		20	0.724	0.437	0.711	0.724	0.715	0.307	**0.682**	**0.730**
MLP	Data Set 1	80	0.647	**0.519**	0.519	0.647	0.642	0.132	0.587	0.650
		50	0.662	0.485	0.657	0.662	0.659	0.180	0.595	0.660
		20	0.632	0.504	0.635	0.632	0.634	0.127	0.613	0.680
J48	Data Set 2	80	0.923	0.103	0.923	0.923	0.922	0.834	0.898	0.880
		50	0.916	0.109	0.916	0.916	0.915	0.818	0.889	0.860
		20	0.910	0.111	0.910	0.910	0.910	0.807	0.906	0.880
Bayesian	Data Set 2	80	0.903	**0.140**	0.906	0.903	0.902	0.793	0.964	0.970
		50	0.903	**0.140**	0.906	0.903	0.902	0.793	0.964	0.970
		20	0.903	**0.140**	0.906	0.903	0.902	0.793	0.965	0.970
MLP	Data Set 2	80	0.935	0.083	0.935	0.935	0.935	0.860	**0.979**	**0.980**
		50	0.931	0.085	0.931	0.931	0.931	0.853	0.977	**0.980**
		20	**0.938**	0.079	**0.938**	**0.938**	**0.938**	**0.868**	0.975	**0.980**





2. Diabetes

Table shows Accuracy, RAE and RRSE: -

Classifier Name	Data Ratio	Accuracy	RAE	RRSE
	80	75.39%	65.79%	88.94%
J48	50	75.78%	66.40%	90.22%
	20	75.13%	**68.37%**	**92.44%**
	80	75.65%	64.03%	86.01%
Bayesian	50	**76.04%**	64.18%	86.27%
	20	**76.04%**	64.84%	86.95%
	80	75.39%	64.69%	86.94%
MLP	50	74.48%	65.60%	88.06%
	20	75.91%	64.57%	87.75%

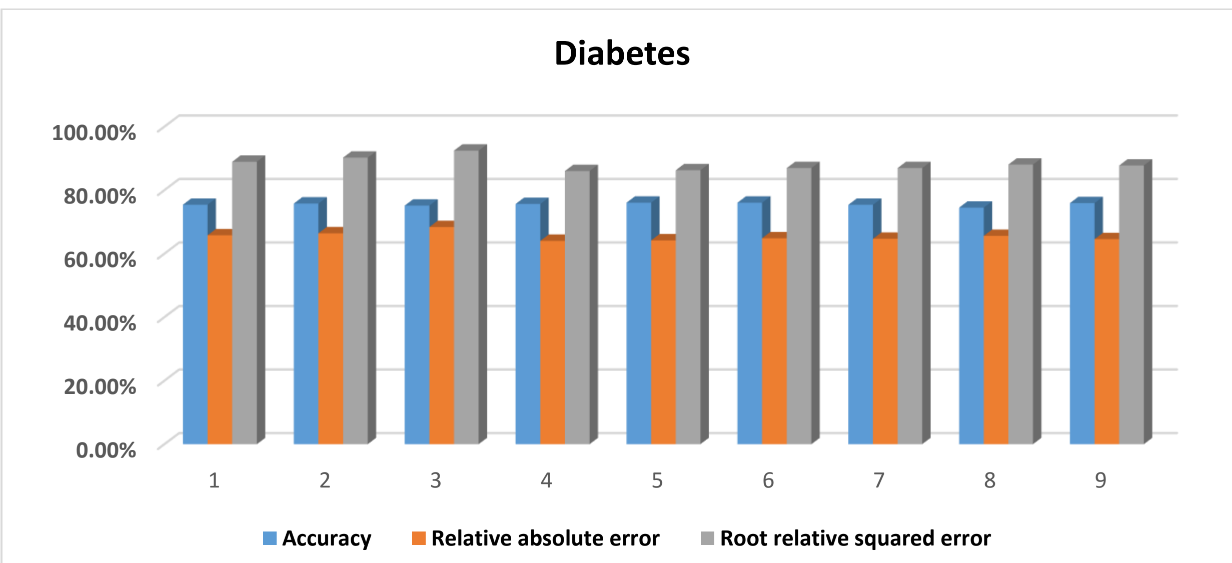


Table Show kappa, MAE and RMSE: -

Classifier Name	Data Ratio	Kappa	MAE	RMSE

J48	80	0.4474	0.2990	0.4239
	50	0.4585	0.3018	0.4301
	20	0.4366	**0.3108**	**0.4406**
Bayesian	80	0.4599	0.2910	0.4100
	50	0.4681	0.2917	0.4112
	20	**0.4700**	0.2947	0.4144
MLP	80	0.4484	0.2940	0.4144
	50	0.4254	0.2982	0.4197
	20	0.4629	0.2935	0.4182

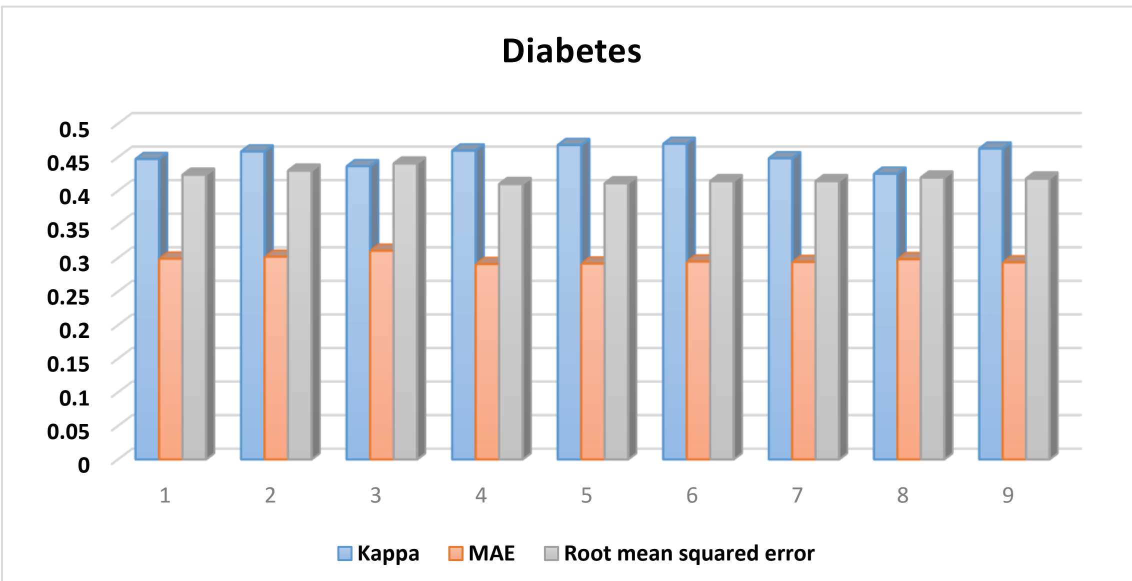
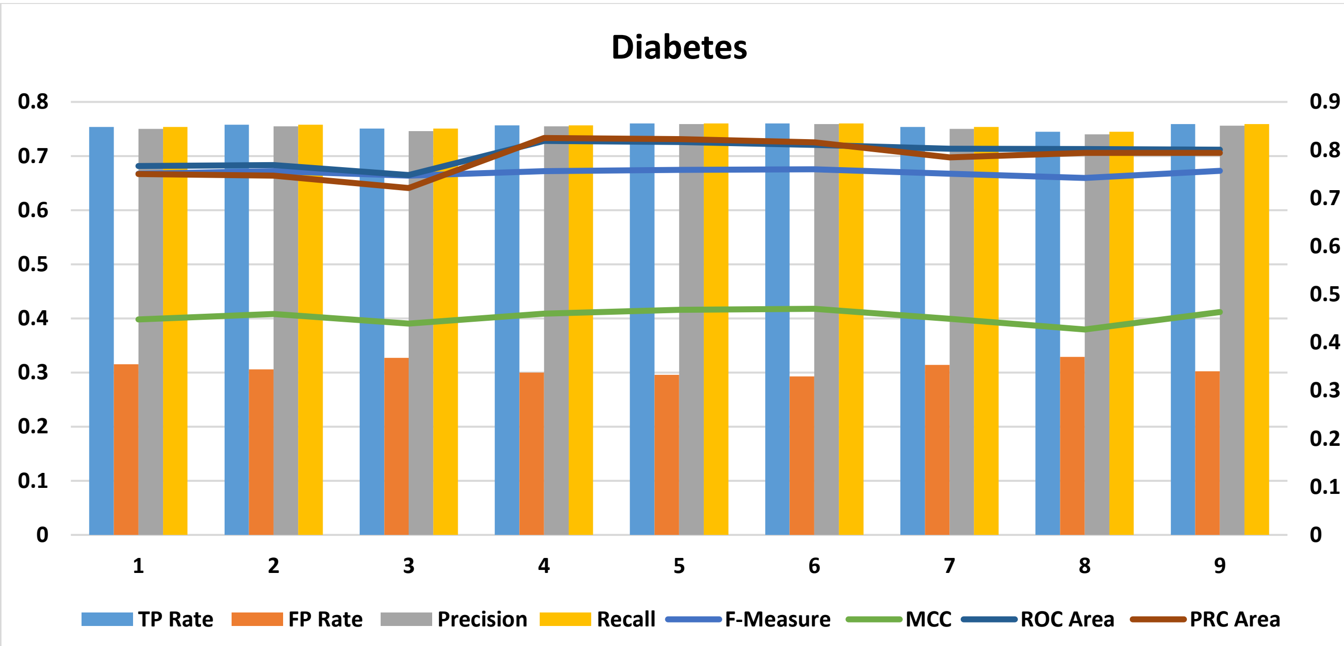


Table present TP and FP Rate, Precision, Recall, F-measure, MCC, ROC and PRC

Classifier Name	Data Ratio	TPR	FPR	Precision area	Recall area	F-Score	MCC	ROC plot	PRC plot
J48	80	0.754	0.315	0.750	0.754	0.751	0.448	0.767	0.750
	50	0.758	0.306	0.755	0.758	0.756	0.459	0.769	0.747
	20	0.751	0.327	0.746	0.751	0.747	0.439	0.748	0.721
	80	0.757	0.300	0.755	0.757	0.756	0.460	**0.819**	**0.825**

Bayesian	50	**0.760**	0.296	**0.759**	**0.760**	0.759	0.468	0.817	0.823
	20	**0.760**	0.293	**0.759**	**0.760**	**0.760**	**0.470**	0.811	0.816
MLP	80	0.754	0.314	0.750	0.754	0.751	0.449	0.803	0.785
	50	0.745	**0.329**	0.740	0.745	0.742	0.427	0.802	0.794
	20	0.759	0.302	0.756	0.759	0.757	0.463	0.801	0.794

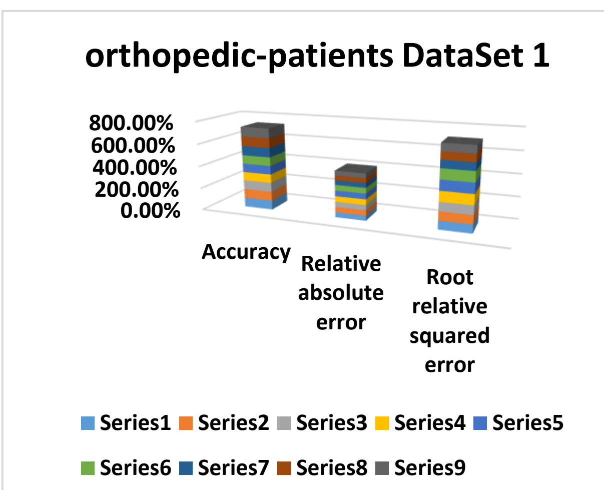


<u>3. Iris</u>

Table shows Accuracy, RAE and RRSE: -

Classifier Name	Data Set Name	Data Ratio	Accuracy	RAE	RRSE
J48	Data Set 1	80	83.87%	45.78%	77.12%
		50	82.90%	47.31%	79.14%
		20	82.26%	46.53%	79.31%
Bayesian	Data Set 1	80	77.74%	50.29%	92.25%
		50	77.42%	**50.49%**	**92.52%**
		20	77.74%	50.17%	92.28%

MLP	Data Set 1	80	85.16%	42.15%	69.12%
		50	85.81%	41.05%	68.06%
		20	**86.45%**	41.57%	67.25%
J48	Data Set 2	80	85.48%	28.61%	64.45%
		50	84.84%	30.00%	66.12%
		20	81.94%	32.48%	**69.36%**
Bayesian	Data Set 2	80	83.55%	33.93%	62.62%
		50	82.90%	**34.20%**	62.94%
		20	83.55%	34.09%	62.78%
MLP	Data Set 2	80	84.52%	31.52%	57.72%
		50	**85.81%**	31.23%	57.89%
		20	84.84%	30.54%	57.25%



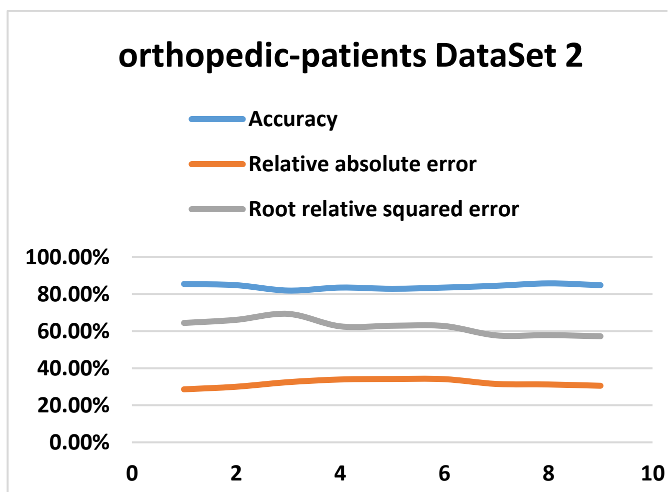
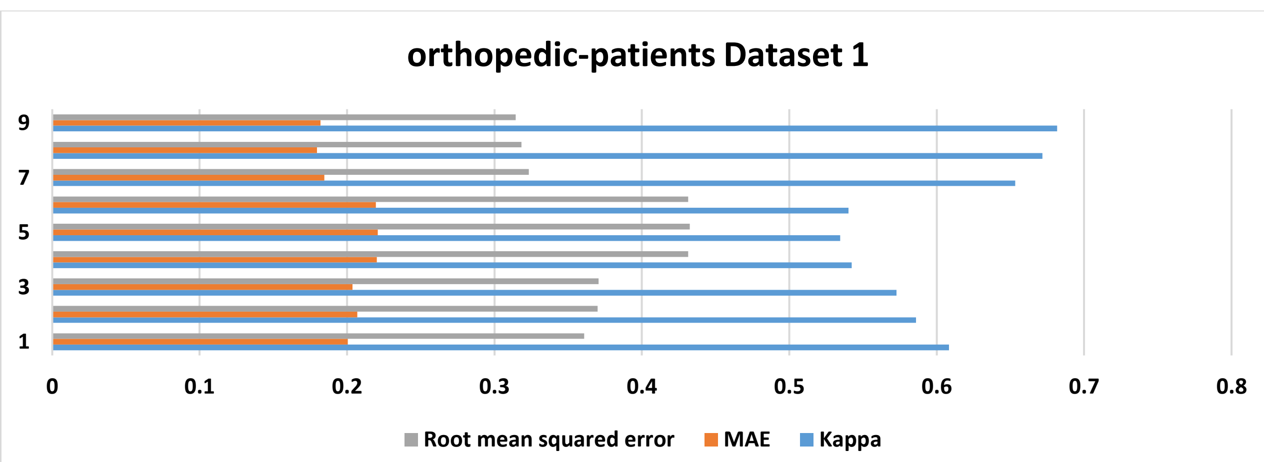


Table Show kappa, MAE and RMSE: -

Classifier Name	Data Set Name	Data Ratio	Kappa	MAE	RMSE
	Data Set 1	80	0.6084	0.2004	0.3608

J48		50	0.5860	0.2070	0.3699
		20	0.5728	0.2036	0.3707
Bayesian	Data Set 1	80	0.5423	0.2202	0.4315
		50	0.5345	**0.2209**	**0.4325**
		20	0.5401	0.2195	0.4314
MLP	Data Set 1	80	0.6532	0.1845	0.3234
		50	0.6718	0.1796	0.3182
		20	**0.6817**	0.1819	0.3144
J48	Data Set 2	80	0.7682	0.1192	0.2942
		50	0.7577	0.1250	0.3017
		20	0.7104	0.1353	**0.3164**
Bayesian	Data Set 2	80	0.7348	0.1414	0.2858
		50	0.7250	**0.1425**	0.2872
		20	0.7354	0.1420	0.2864
MLP	Data Set 2	80	0.7537	0.1314	0.2635
		50	**0.7737**	0.1301	0.2641
		20	0.7578	0.1272	0.2612



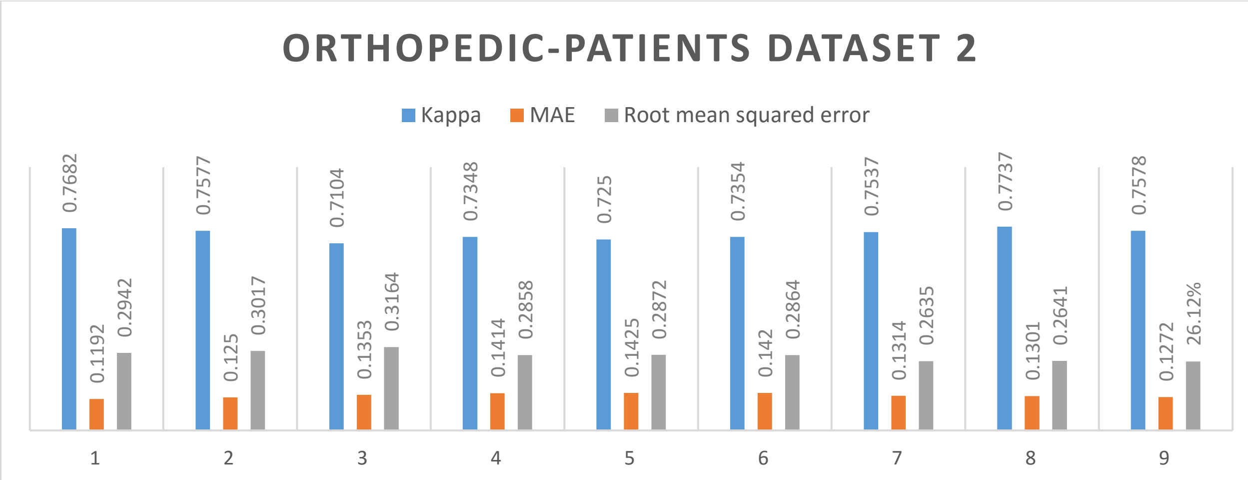
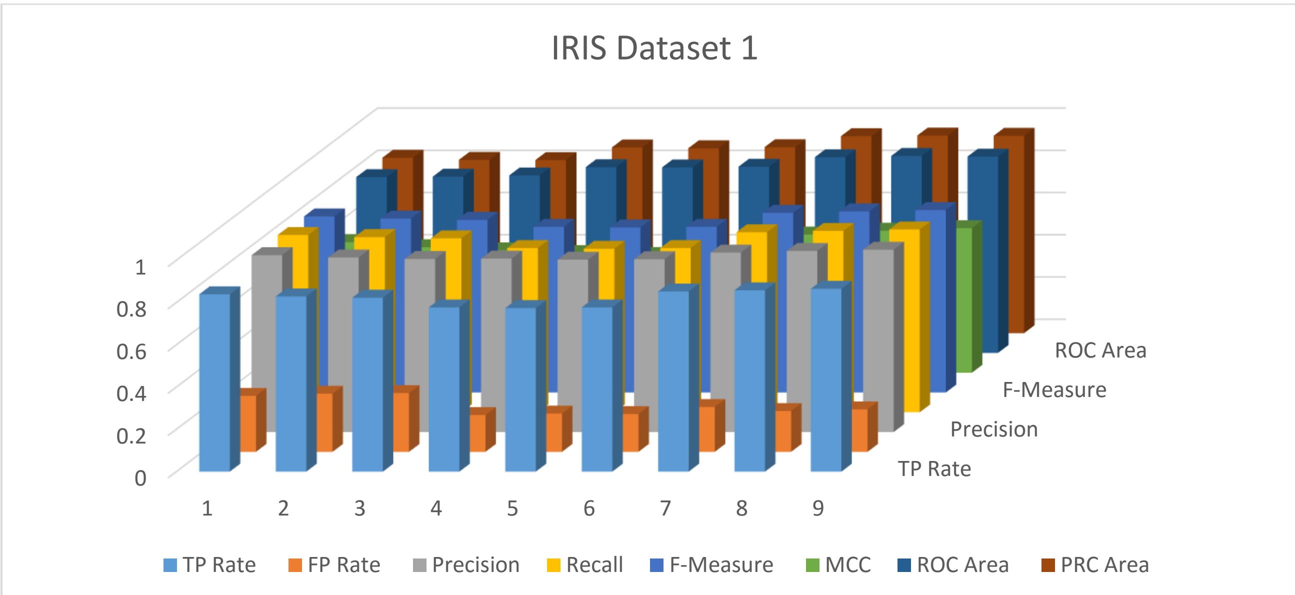
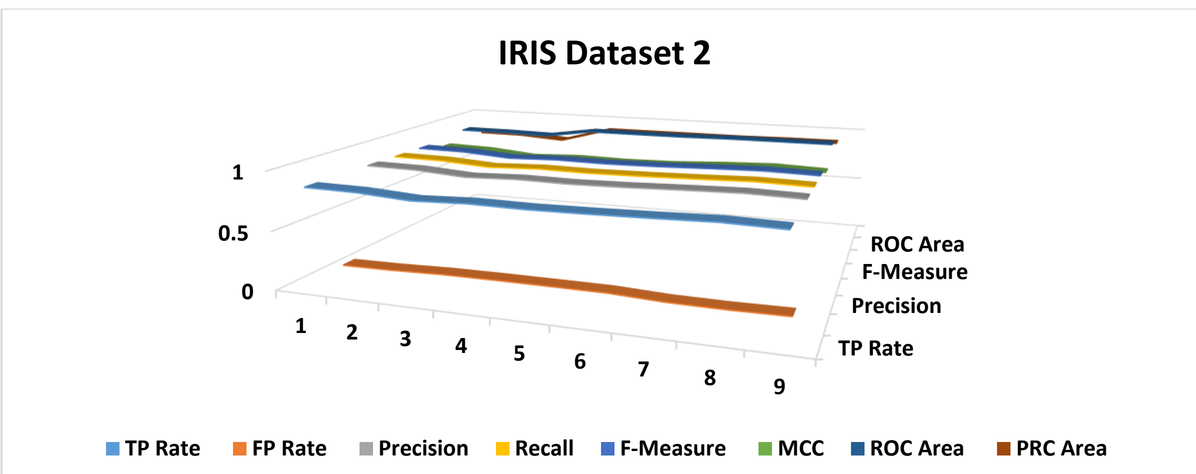


Table present TP and FP Rate, Precision, Recall, F-measure, MCC, ROC and PRC

Classifier Name	Data Set	Data Ratio	TPR	FPR	Precision area	Recall plot	F-score	MCC	ROC plot	PRC plot
J48	Data Set 1	80	0.839	0.265	0.837	0.839	0.833	0.618	0.832	0.830
		50	0.829	0.275	0.826	0.829	0.823	0.594	0.834	0.820
		20	0.823	**0.278**	0.819	0.823	0.817	0.579	0.840	0.819
Bayesian	Data Set 1	80	0.777	0.174	0.821	0.777	0.784	0.566	0.879	0.879
		50	0.774	0.181	0.816	0.774	0.781	0.557	0.878	0.875
		20	0.777	0.179	0.818	0.777	0.784	0.561	0.881	0.880
MLP	Data Set 1	80	0.852	0.212	0.849	0.852	0.850	0.654	0.926	0.932
		50	0.858	0.193	0.857	0.858	0.857	0.672	**0.932**	**0.935**
		20	**0.865**	0.201	**0.863**	**0.865**	**0.863**	**0.684**	0.929	0.934
J48	Data Set 2	80	0.855	0.062	0.857	0.855	0.855	0.793	0.896	0.796
		50	0.848	0.067	0.850	0.848	0.849	0.781	0.894	0.793
		20	0.819	0.080	0.819	0.819	0.818	0.741	0.885	0.764
Bayesian	Data Set 2	80	0.835	0.083	0.834	0.835	0.832	0.759	0.954	0.890
		50	0.829	**0.083**	0.828	0.829	0.826	0.751	0.953	0.890
		20	0.835	0.082	0.836	0.835	0.833	0.759	0.954	0.890

MLP	Data Set 2	80	0.845	0.061	0.850	0.845	0.847	0.783	0.961	0.889
		50	**0.858**	0.057	**0.861**	**0.858**	**0.859**	**0.80**	0.962	0.896
		20	0.848	0.063	0.850	0.848	0.849	0.785	**0.963**	**0.897**



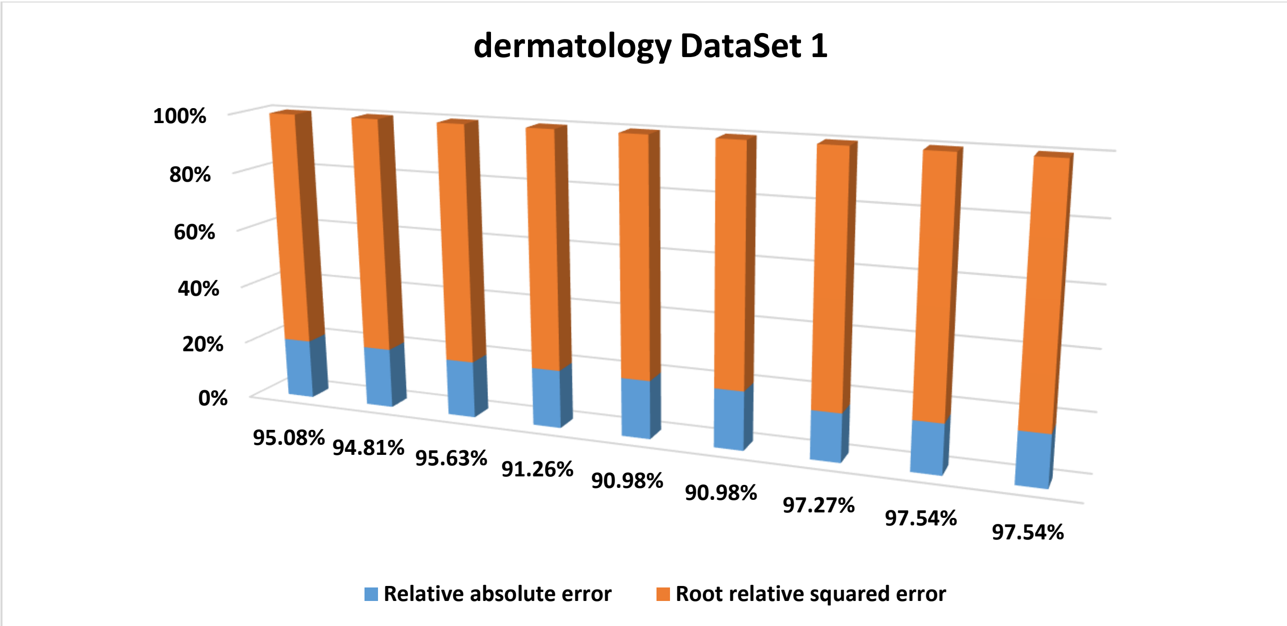


4. Dermatology

Table shows Accuracy, RAE and RRSE: -

Classifier Name	Data Set Name	Data Ratio	Accuracy	RAE	RRSE
J48	Data Set 1	80	95.08%	8.98%	35.08%
		50	94.81%	9.48%	36.53%

		20	95.63%	7.94%	32.90%
Bayesian	Data Set 1	80	91.26%	11.56%	46.60%
		50	90.98%	11.53%	46.51%
		20	90.98%	**11.63%**	**46.75%**
MLP	Data Set 1	80	97.27%	4.73%	24.19%
		50	**97.54%**	4.66%	22.93%
		20	**97.54%**	5.02%	23.81%
J48	Data Set 2	80	**83.43%**	39.94%	72.49%
		50	83.07%	40.64%	73.02%
		20	82.12%	43.03%	74.58%
Bayesian	Data Set 2	80	53.90%	95.46%	106.87%
		50	53.82%	**95.47%**	**106.90%**
		20	53.85%	95.43%	106.86%
MLP	Data Set 2	80	64.23%	76.10%	97.57%
		50	64.97%	74.17%	95.63%
		20	63.70%	76.41%	97.33%



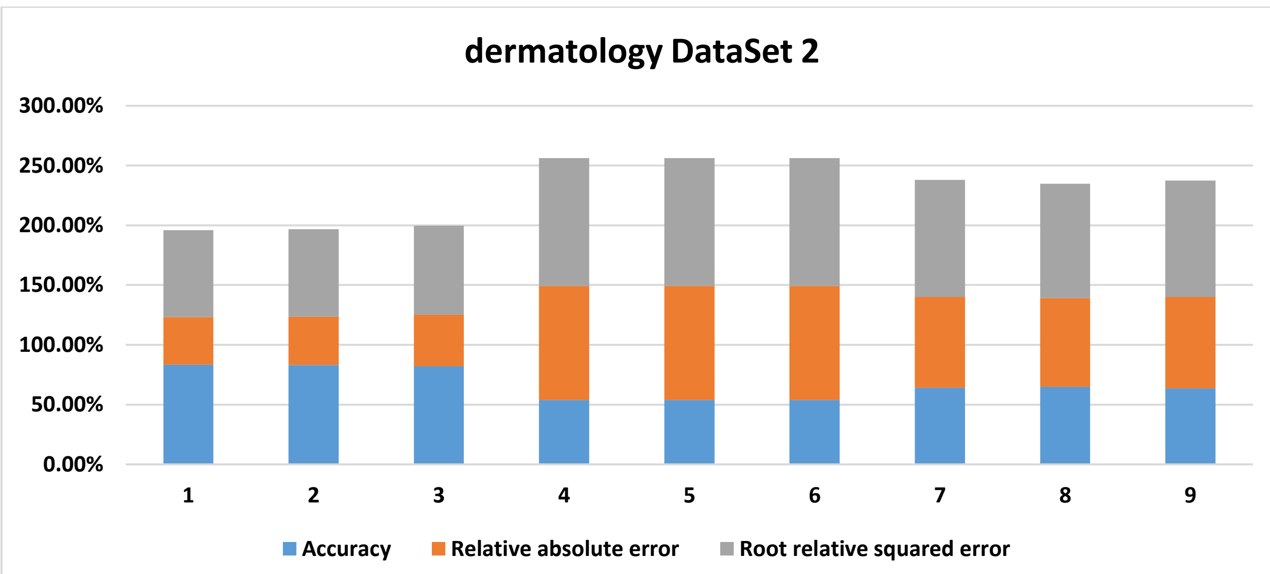
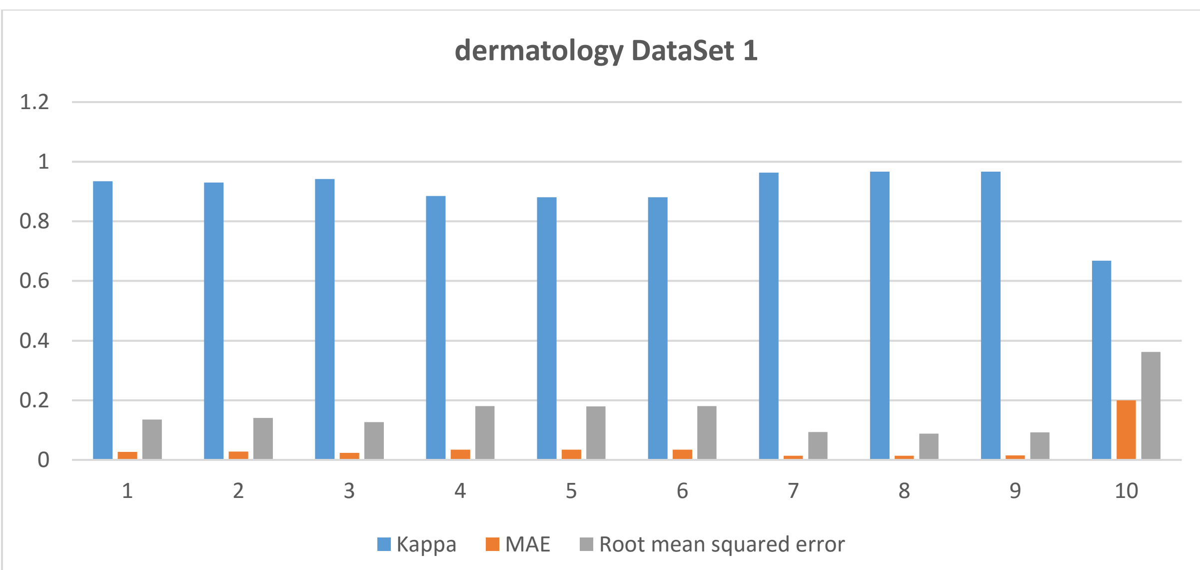


Table Show kappa, MAE and RMSE: -

Classifier Name	Data Set Name	Data Ratio	Kappa	MAE	RMSE
J48	Data Set 1	80	0.9342	0.0269	0.1356
		50	0.9306	0.0283	0.1412
		20	0.9416	0.0237	0.1271
Bayesian	Data Set 1	80	0.8847	0.0346	0.1802
		50	0.8812	0.0345	0.1797

		20	0.8812	**0.0347**	**0.1806**
MLP	Data Set 1	80	0.9633	0.0141	0.0935
		50	**0.9670**	0.0139	0.0886
		20	**0.9670**	0.0150	0.0920
J48	Data Set 2	80	**0.6682**	0.1995	0.3623
		50	0.6609	0.2030	0.3649
		20	0.6418	0.2149	0.3727
Bayesian	Data Set 2	80	0.0950	0.4768	0.5341
		50	0.0935	**0.4769**	**0.5342**
		20	0.0941	0.4767	0.5341
MLP	Data Set 2	80	0.2861	0.3801	0.4876
		50	0.3019	0.3705	0.4779
		20	0.2766	0.3817	0.4864



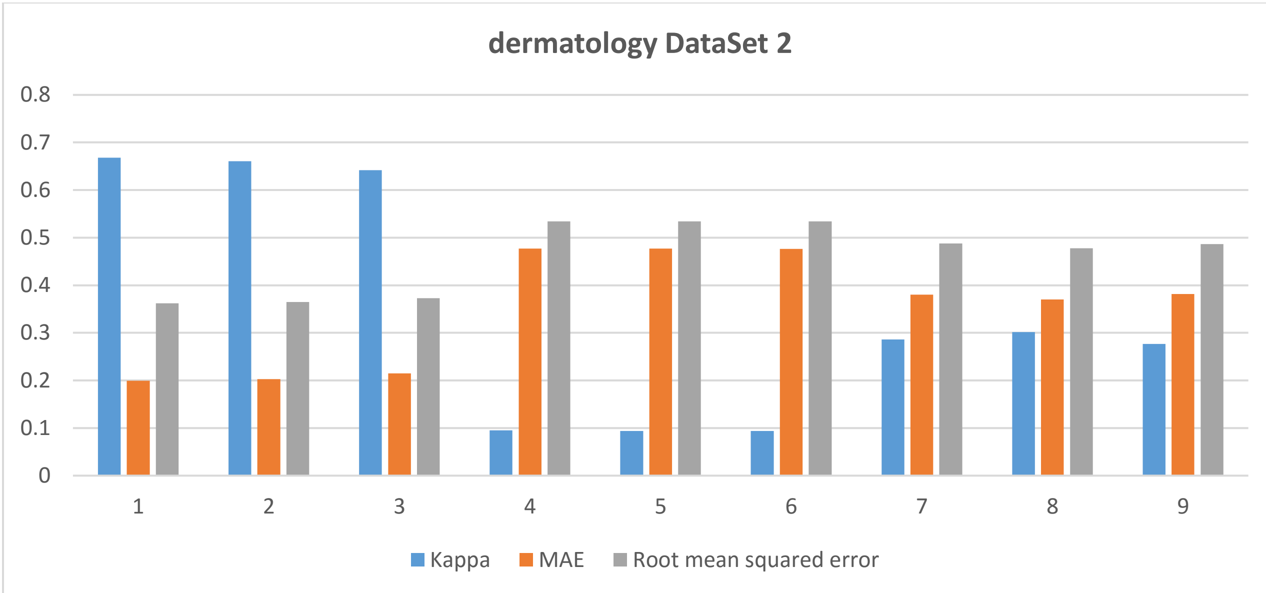
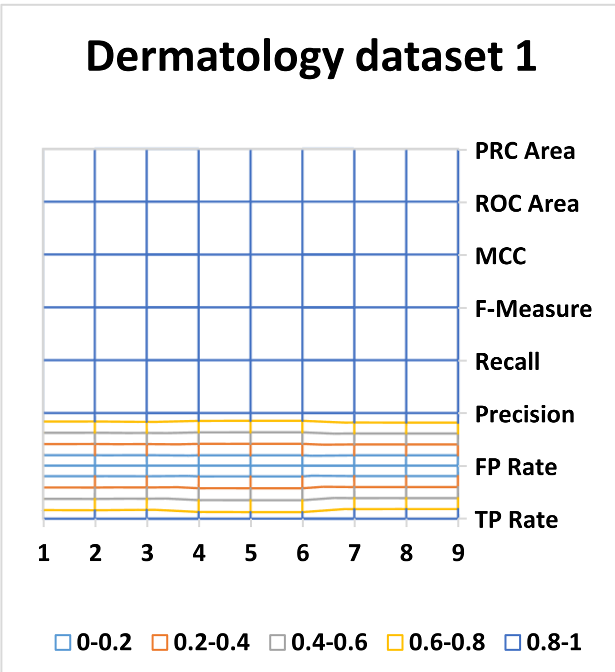
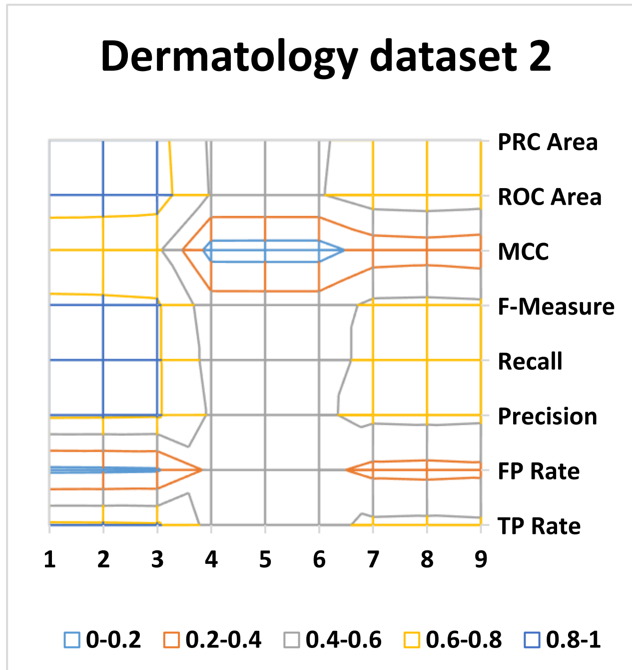


Table present TP and FP Rate, Precision, Recall, F-measure, MCC, ROC and PRC

Classifier Name	Data Set Name	Data Ratio	TPR	FPR	Precision area	Recall plot	F-Score	MCC	ROC Plot	PRC Plot
J48	Data Set 1	80	0.951	0.01	0.952	0.951	0.951	0.936	0.980	0.938
		50	0.948	0.01	0.950	0.948	0.948	0.933	0.978	0.929
		20	0.956	0.01	0.958	0.956	0.957	0.944	0.983	0.944
Bayesian	Data Set 1	80	0.913	0.02	0.937	0.913	0.915	0.894	0.996	0.980
		50	0.910	0.02	0.935	0.910	0.913	0.891	0.996	0.980
		20	0.910	0.02	0.935	0.910	0.913	0.891	0.996	0.981
MLP	Data Set 1	80	0.973	0.01	0.973	0.973	0.973	0.963	0.996	0.990
		50	0.975	0.01	0.975	**0.975**	**0.975**	**0.967**	0.996	**0.992**
		20	**0.975**	0.01	**0.976**	**0.975**	**0.975**	**0.967**	0.996	0.991
J48	Data Set 2	80	**0.834**	0.17	**0.834**	**0.834**	**0.834**	**0.668**	**0.893**	**0.875**
		50	0.831	0.17	0.831	0.831	0.831	0.661	0.891	0.873
		20	0.821	0.18	0.821	0.821	0.821	0.642	0.884	0.866
Bayesian	Data Set 2	80	0.539	0.44	0.577	0.539	0.495	0.120	0.587	0.573
		50	0.538	0.44	0.576	0.538	0.494	0.119	0.586	0.572

		20	0.538	0.44	0.576	0.538	0.494	0.119	0.587	0.573
MLP	Data Set 2	80	0.642	0.36	0.645	0.642	0.642	0.288	0.708	0.704
		50	0.650	0.35	0.654	0.650	0.649	0.305	0.722	0.717
		20	0.637	0.36	0.642	0.637	0.636	0.280	0.706	0.699



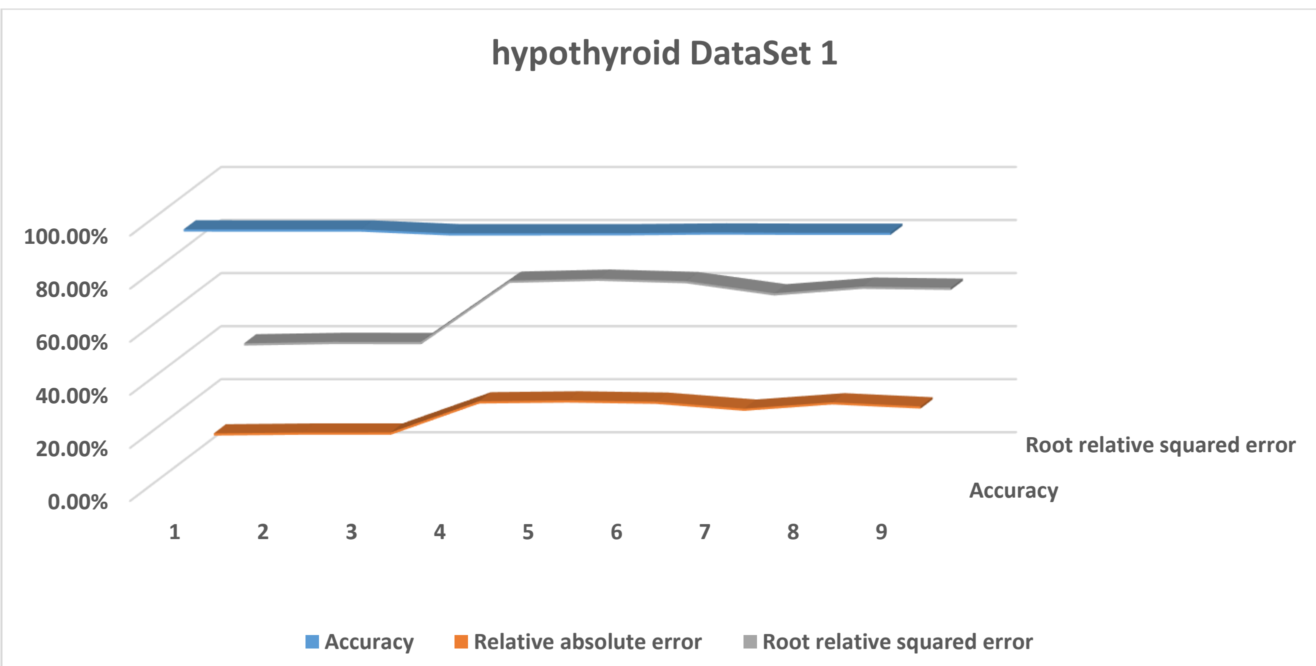


5. Hypothyroid

Table shows Accuracy, RAE and RRSE: -

Classifier Name	Data Set Name	Data Ratio	Accuracy	RAE	RRSE
J48	Data Set 1	80	**99.30%**	13.81%	39.13%
		50	99.27%	14.12%	39.68%
		20	99.27%	14.13%	39.63%
Bayesian	Data Set 1	80	97.85%	25.89%	62.75%
		50	97.85%	**26.24%**	**63.52%**
		20	97.88%	25.75%	62.59%

MLP	Data Set 1	80	98.17%	23.12%	57.97%
		50	98.04%	25.52%	60.49%
		20	98.04%	23.97%	60.08%
J48	Data Set 2	80	80.53%	55.77%	93.47%
		50	80.20%	**56.34%**	**93.64%**
		20	81.19%	53.91%	93.51%
Bayesian	Data Set 2	80	79.21%	48.08%	85.87%
		50	80.20%	48.34%	86.51%
		20	79.87%	48.11%	86.34%
MLP	Data Set 2	80	81.19%	44.64%	87.62%
		50	**84.49%**	39.75%	81.75%
		20	82.51%	43.39%	86.55%



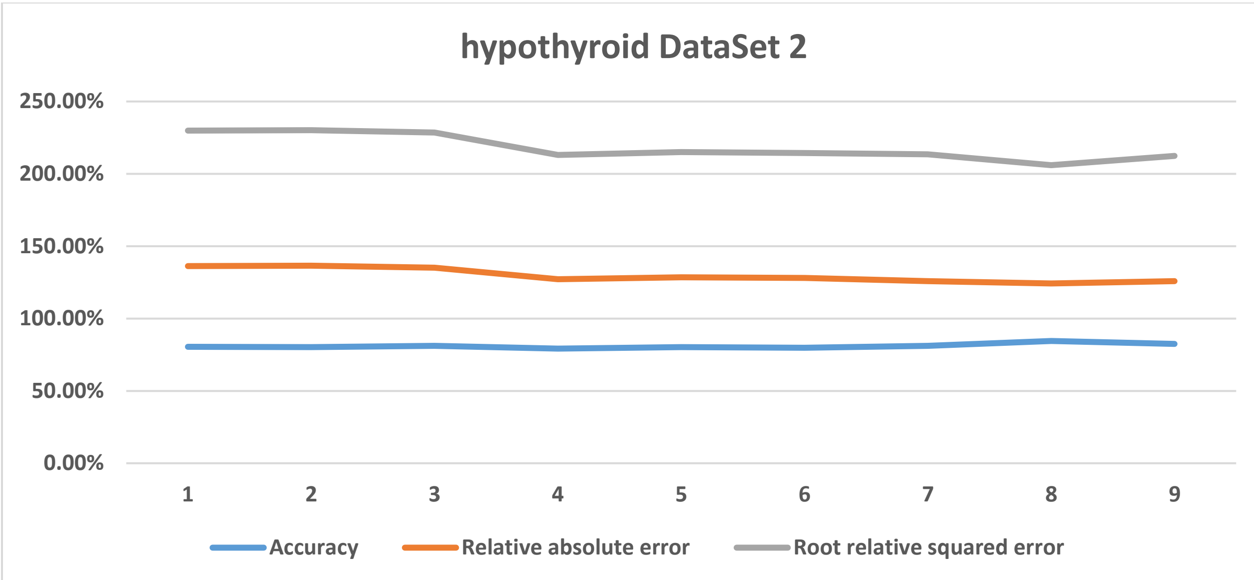
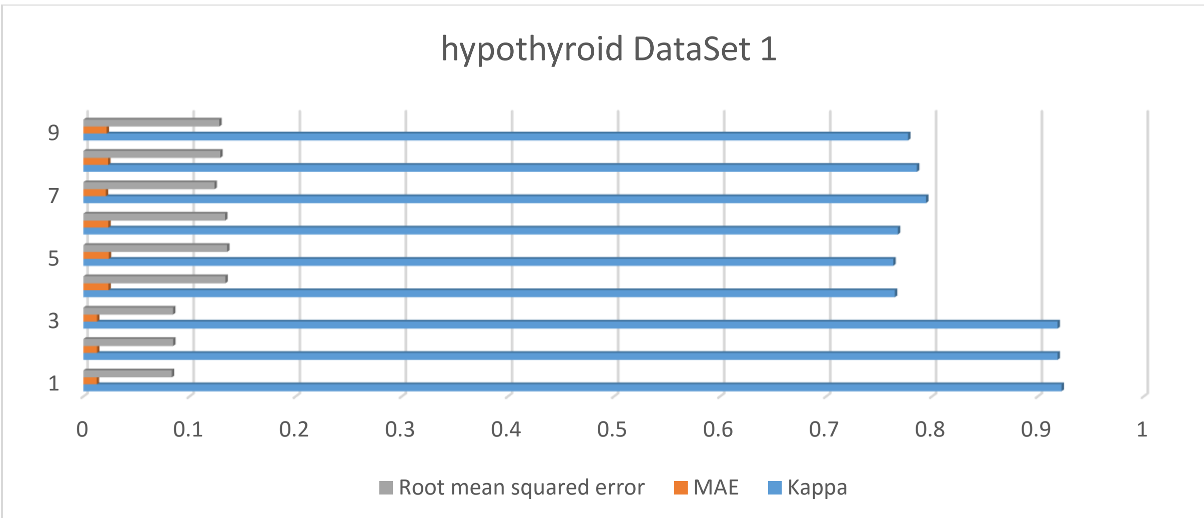


Table Show kappa, MAE and RMSE: -

Classifier Name	Dataset Name	Data Ratio	Kappa	MAE	RMSE
J48	Data Set 1	80	**0.9225**	0.0126	0.0834
		50	0.9187	0.0129	0.0846
		20	0.9187	0.0129	0.0845
Bayesian	Data Set 1	80	0.7650	0.0236	0.1338
		50	0.7635	**0.0239**	0.1354
		20	0.7678	0.0235	0.1335
MLP	Data Set 1	80	0.7944	0.0211	0.1236
		50	0.7858	0.0233	**0.1290**
		20	0.7774	0.0219	0.1281
J48	Data Set 2	80	0.5227	0.2287	0.4230
		50	0.5061	**0.2311**	**0.4239**
		20	0.5452	0.2211	0.4231
Bayesian	Data Set 2	80	0.5352	0.1972	0.3887
		50	0.5531	0.1983	0.3916

		20	0.5442	0.1973	0.3907
MLP	Data Set 2	80	0.5452	0.1831	0.3965
		50	**0.6224**	0.1630	0.3701
		20	0.5771	0.1779	0.3916



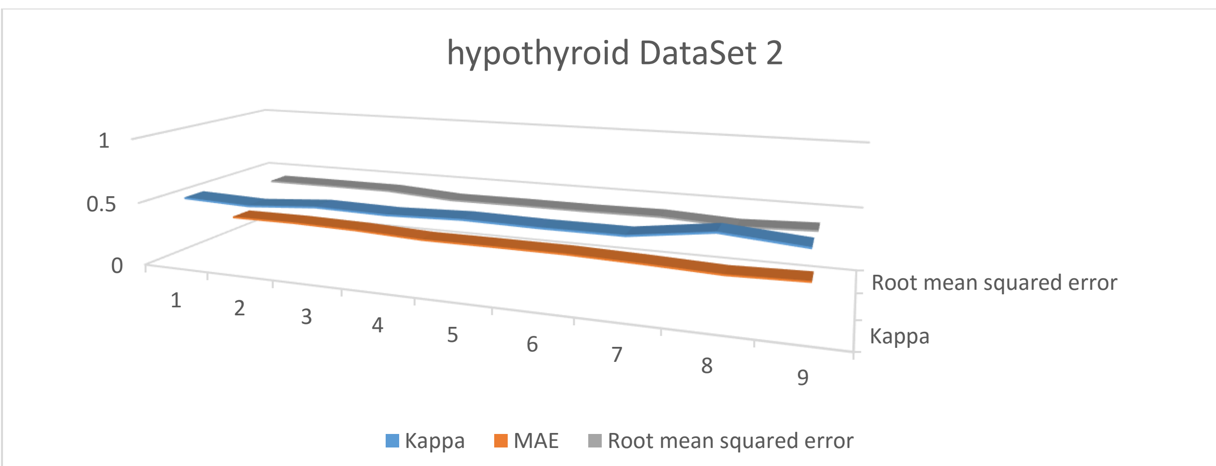
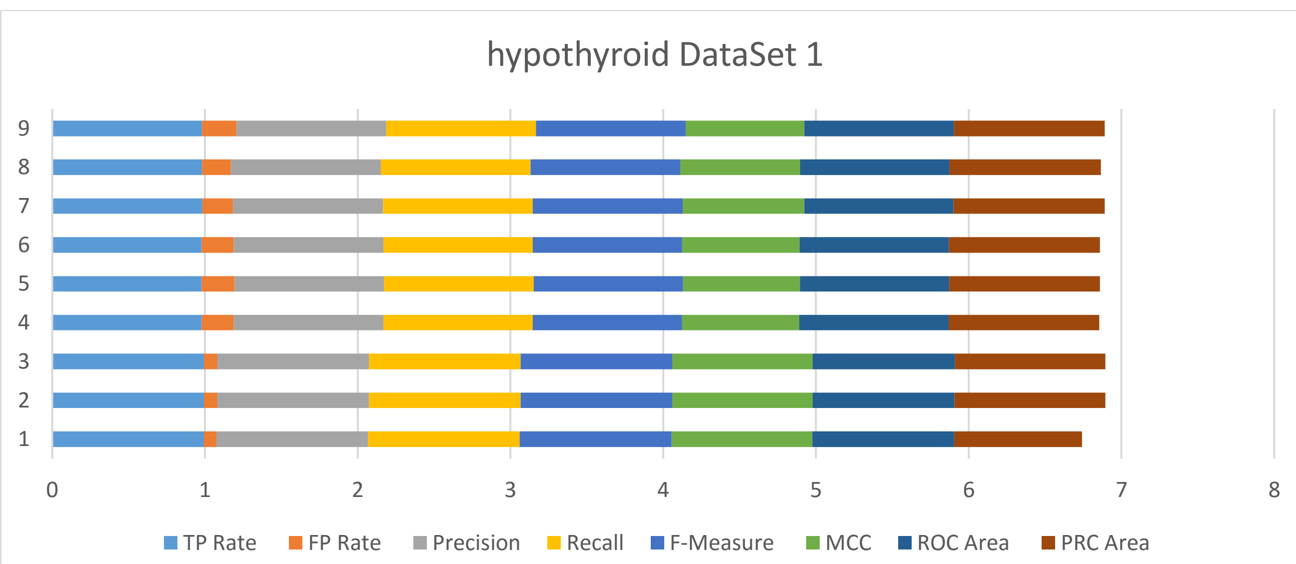
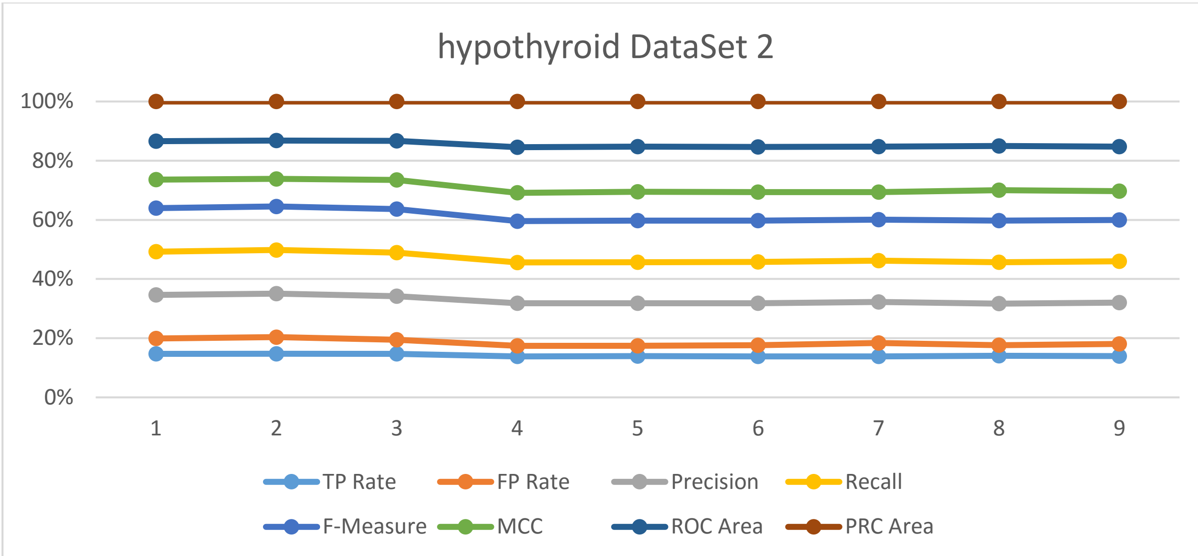


Table present TP and FP Rate, Precision, Recall, F-measure, MCC, ROC and PRC

Classifier Name	Data Set	Data Ratio	TP	FP	Precision area	Recall plot	F-Score	MCC	ROC Plot	PRC Plot
J48	Data Set 1	80	**0.993**	0.082	**0.993**	**0.993**	**0.993**	**0.923**	**0.927**	**0.837**
		50	0.993	0.088	0.993	0.993	0.993	0.919	0.927	0.988

		20	0.993	0.088	0.993	0.993	0.993	0.919	0.932	0.983
Bayesian	Data Set 1	80	0.979	0.209	0.979	0.979	0.979	0.765	0.978	0.987
		50	0.979	0.215	0.979	0.979	0.979	0.764	0.978	0.987
		20	0.979	0.209	0.979	0.979	0.979	0.768	0.978	0.987
MLP	Data Set 1	80	0.982	0.202	0.981	0.982	0.981	0.795	0.977	0.99
		50	0.980	0.190	0.981	0.980	0.980	0.786	0.978	0.991
		20	0.980	**0.227**	0.980	0.980	0.980	0.778	0.976	0.99
J48	Data Set 2	80	0.805	0.284	0.805	0.805	0.805	0.523	0.715	0.733
		50	0.802	**0.306**	0.798	0.802	0.800	0.507	0.703	0.718
		20	0.812	0.261	0.814	0.812	0.813	0.545	0.729	0.736
Bayesian	Data Set 2	80	0.792	0.207	0.819	0.792	0.799	0.546	0.883	0.884
		50	0.802	0.203	0.825	0.802	0.808	0.562	0.879	0.878
		20	0.799	0.212	0.820	0.799	0.805	0.552	0.881	0.882
MLP	Data Set 2	80	0.812	0.261	0.814	0.812	0.813	0.545	0.895	0.895
		50	**0.845**	0.220	**0.845**	**0.845**	**0.845**	**0.622**	**0.902**	**0.905**
		20	0.825	0.242	0.827	0.825	0.826	0.577	0.891	0.899




hypothyroid DataSet 2
100%
80%
60%
40%
20%
0%
1
2
3
4
5
6
7
8
9
TP Rate
FP Rate
Precision
Recall
F-Measure
MCC
ROC Area
PRC Area

CHAPTER 6

RESULTS

The outcome illustrates minimum, maximum and average result by determining error and accuracy of overall medical datasets.

Table: Average accuracy and error		
DATA SET	Accuracy	Error
IRIS	82.15%	46.15%
	84.16%	31.84%
Breast-Cancer	76.87%	84.64%
	91.82%	22.33%
dermatology	94.57%	8.39%
	77.01%	70.74%
diabetes	81.36%	65.39%
hypothyroid	98.41%	21.39%
	81.04%	48.70%
AVG	**85.87%**	**44.40%**
MAX	**98.41%**	**84.64%**
MIN	**76.01%**	**8.39%**

Table 14: Average accuracy and error

The overall accuracy is 85.87% and absolute error is 44.40%

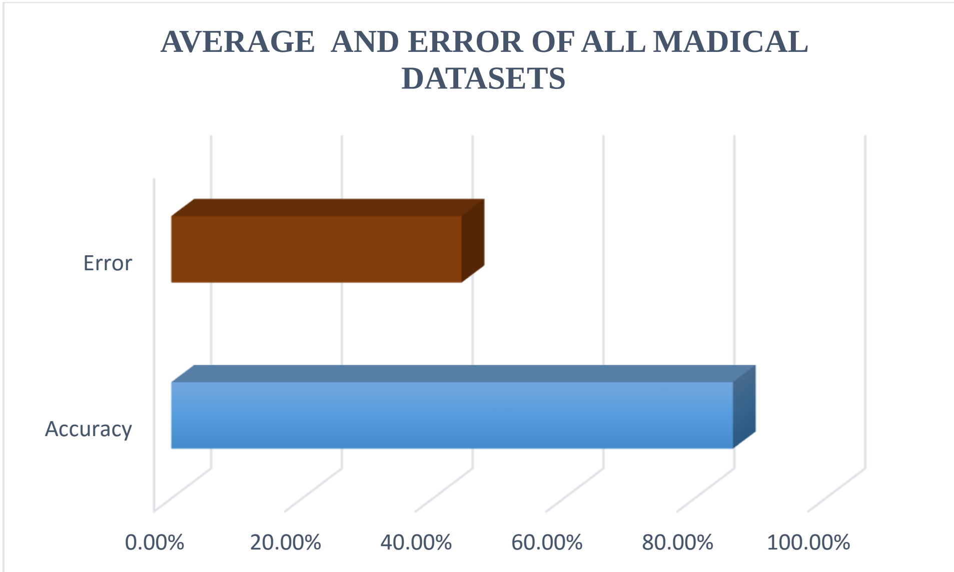


Figure 15: Average accuracy and error

6.1 Maximum and minimum accuracy

The maximum accuracy is hypothyroid dataset-1 98.41% and minimum accuracy is dermatology dataset-2 67.01%

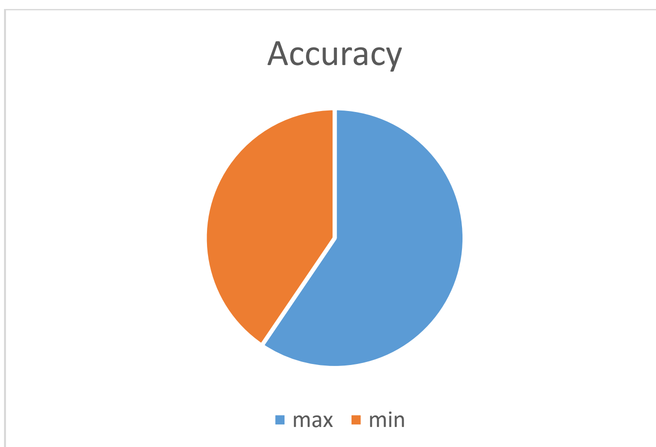


Figure 16: Maximum and minimum accuracy

6.2 Maximum and minimum error

The maximum error is Brest cancer dataset-1 84.64% and minimum error dermatology dataset-1 8.39%

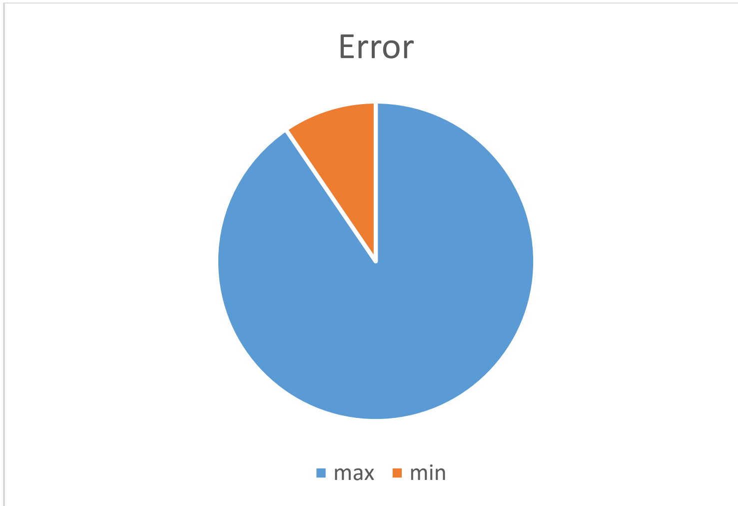


Figure 17: Maximum and minimum error

6.3 Maximum error and accuracy

The maximum accuracy is hypothyroid dataset-1 98.41% and error is in Brest cancer dataset-1 84.64%

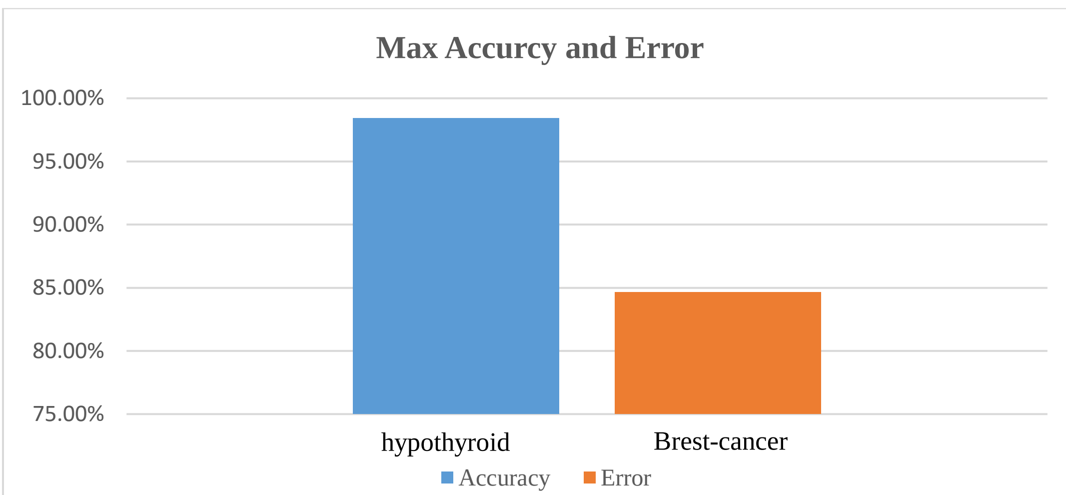


Figure 18: Maximum error and accuracy

6.4 Minimum error and accuracy

The maximum accuracy is dermatology dataset-2 67.01% and error is in dermatology dataset-1 8.39%

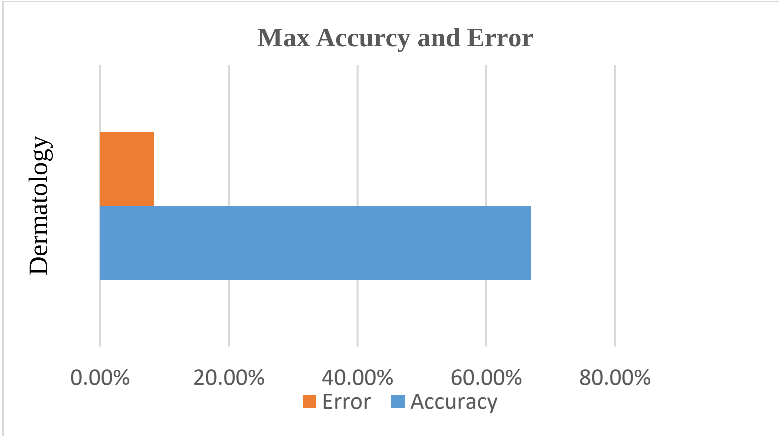


Figure 19: Minimum error and accuracy

6.5 Result on the base on classifier performance

This section describes the accuracy and error between different classifiers applied on the datasets

6.5.1 Decision Tree(J48)

In this section we illustrate an overall accuracy of J48 classifier of datasets

Decision Tree (J48 classifier)		
DATA SET	Accuracy	Error
IRIS	91.55%	38.45%
Breast-Cancer	93.56%	57.96%
dermatology	89.02%	25.00%
diabetes	80.43%	66.85%
hypothyroid	88.46%	34.67%

Table 15: shows overall accuracy of J48 classifier of datasets

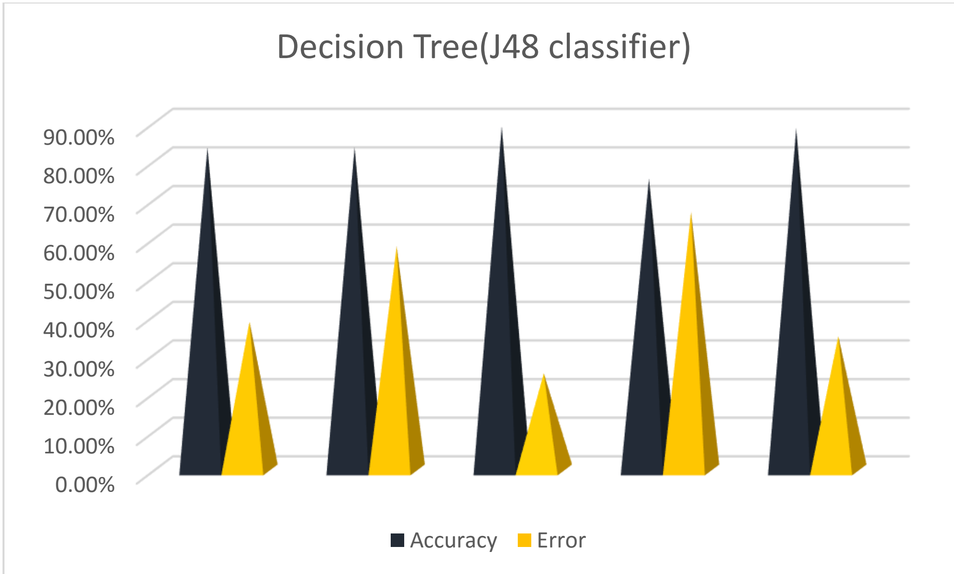


Figure 20: overall accuracy of J48 classifier of datasets

6.5.2 MLP

This part proposes the accuracy calculated with MLP classifier on medical datasets

MLP (Multilayer perceptron)		
DATA SET	Accuracy	Error
IRIS	91.43%	36.34%
Breast-Cancer	86.10%	53.75%
dermatology	88.87%	40.18%
diabetes	92.26%	64.96%
hypothyroid	90.41%	33.40%

Table 16: shows overall accuracy of MLP classifier of datasets

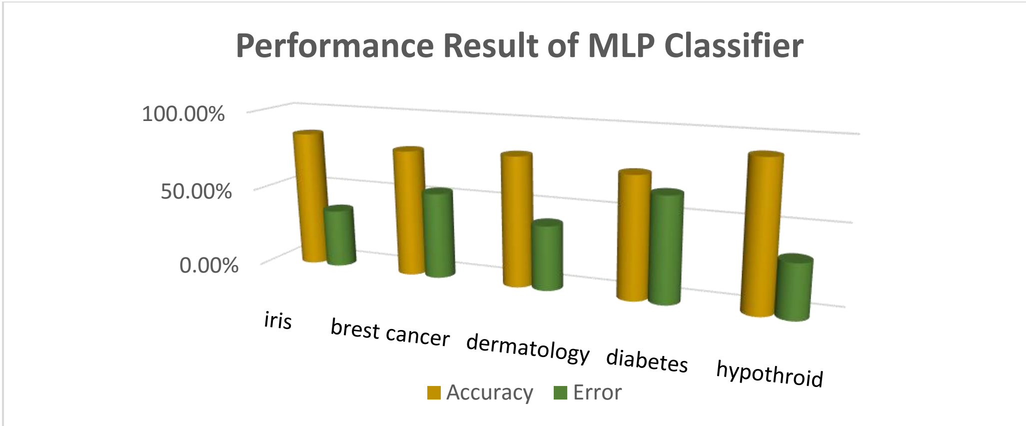


Figure 21: shows overall accuracy of MLP classifier of datasets

6.5.3 Bayesian

The last we determined the accuracy performance evaluated with Bayesian algorithm on medical datasets

Bayesian		
DATA SET	Accuracy	Error
IRIS	90.48%	42.20%
Breast-Cancer	81.38%	51.75%
dermatology	92.47%	41.51%
diabetes	85.91%	64.35%
hypothyroid	88.81%	37.07%

Table 17: shows overall accuracy of MLP classifier of datasets

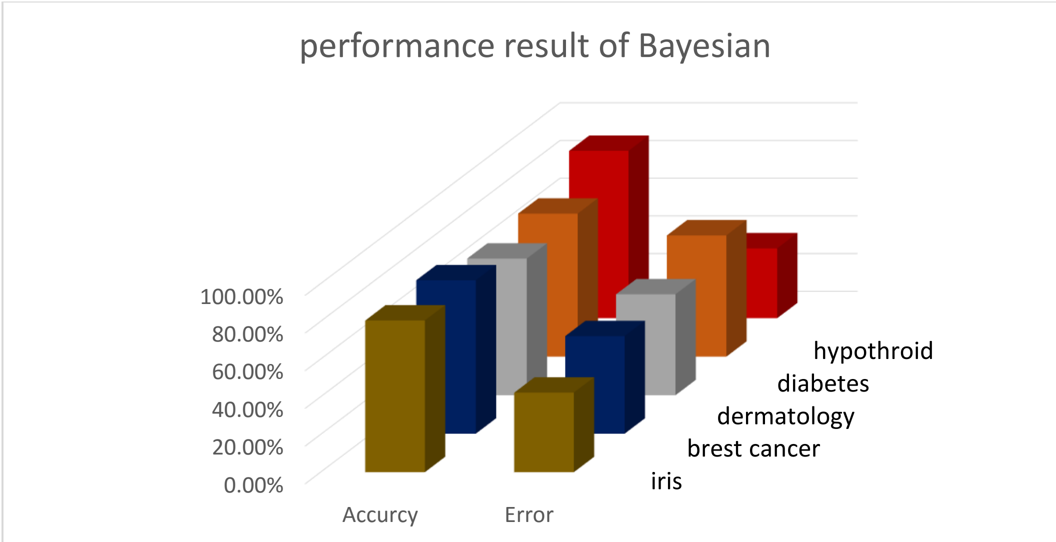


Figure 22: accuracy of MLP classifier of datasets

Finally, the outcome presents the error and accuracy between the five medical datasets. And also predict the maximum, minimum average rates. The dataset test on the three different classifiers j48 from design tree, MLP from function and naive Bayes classifier from Bayes schema.

CHAPTER 7

CONCLUSION AND FUTURE WORK

In this chapter. We will summarize the research, covering the all previous chapters. Following section will be about covering the study and reading the result after analyzing using the WEKA program. Although, the obstacles we have faced during the research.

7.1 Conclusion

As there is some disease that need permanently monitoring. Propose system allows real-time monitoring for the purpose of short- and long-term analysis and diagnosis. The complete system is prepared with IOT devices and AI (Artificial intelligent) technique i.e. machine learning. The intelligent or expert system design to handle missing values, noise, outliers, and so on. Devices connected to e-health users such as the doctor, patient and third party. This goal is to classify medical data, store in electronic health care (EHR). Used in various ways include continue health monitoring of the patient, save the patient lives in an emergency, emergency alert and report is generated, diagnose the chronic disease at an early stage and more doctor can diagnose and decide, view medication side effect, errors and prescribe accordingly. The main target is the classification and prediction of sensitive data and provides the proper and complete information to medical professionals.

This research presents the comparison of three classifiers J48, MLP, and Naive Bayes on five different medical datasets of major disease includes dermatology, hypothyroid, Brest cancer, iris, and diabetes. Dataset are collected separately for each disease to find accuracy and error. The objective includes continue health monitoring of the patient, early stage diagnose and more physician can diagnose and decide, view medication side effect, errors and prescribe accordingly. WEKA tool is used to perform experiment on disease data set by using some classifier and predict the result. The results show's that the how much medical datasets are accurate after testing on the classifiers.

7.2 Future work

The main purpose of this section to set a plan to implement the project in one of the hospitals. In fact, the implementation is going to be rough deal since the systems and infrastructure in the hospitals is very complicated. However, we are going to put all effort we have and use our communications to propose this framework to the hospitals. IOT medical devices are going to invade us eventually. Therefore, it is butter to be ready for this change.

REFERENCES

[1] K. K. Goyal, A. Garg, A. Rastogi, and S. Singhal, "A Literature Survey on Internet of Things (IoT)," *Int J Adv. Netw. Appl.*, vol. 09, no. 06, p. 6, 2018.

[2] M. Elhoseny, G. Ramirez-Gonzalez, O. M. Abu-Elnasr, S. A. Shawkat, A. N, and A. Farouk, "Secure Medical Data Transmission Model for IoT-Based Healthcare Systems," *IEEE Access*, vol. 6, pp. 20596–20608, 2018.

[3] F. Firouzi *et al.*, "Internet-of-Things and big data for smarter healthcare: from device to architecture, applications and analytics," *Future Gener. Comput. Syst.*, vol. 78, pp. 583–586, Jan. 2018.

[4] B. Obama, "United States Health Care Reform: Progress to Date and Next Steps," *JAMA*, vol. 316, no. 5, pp. 525–532, Aug. 2016.

[5] Y.-Y. Chen, J.-C. Lu, and J.-K. Jan, "A Secure EHR System Based on Hybrid Clouds," *J. Med. Syst.*, vol. 36, no. 5, pp. 3375–3384, Oct. 2012.

[6] B. F. Chimieski and R. D. R. Fagundes, "Association and Classification Data Mining Algorithms Comparison over Medical Datasets," p. 8, 2013.

[7] J. Wan, C. Zou, K. Zhou, R. Lu, and D. Li, "IoT sensing framework with inter-cloud computing capability in vehicular networking," *Electron. Commer. Res.*, vol. 14, pp. 389–416, 2014.

[8] D. Dziak, B. Jachimczyk, and W. J. Kulesza, "IoT-Based Information System for Healthcare Application: Design Methodology Approach," *Appl. Sci.*, vol. 7, no. 6, p. 596, Jun. 2017.

[9] J. F. Gomes and S. Moqaddemerad, "Futures Business Models for an IoT Enabled Healthcare Sector: A Causal Layered Analysis Perspective," *J. Bus. Models*, vol. 4, no. 2, Oct. 2016.

[10] M. C. Domingo, "An overview of the Internet of Things for people with disabilities," *J. Netw. Comput. Appl.*, vol. 35, no. 2, pp. 584–596, Mar. 2012.

[11] Y. Lu and J. Cecil, "An Internet of Things (IoT)-based collaborative framework for advanced manufacturing," *Int. J. Adv. Manuf. Technol.*, vol. 84, no. 5, pp. 1141–1152, May 2016.

[12] K. Suresh and M. R. Babu, "Emerging Biomedical Health Care System by Using Internet of Things," *Journal of Biomimetics, Biomaterials and Biomedical Engineering*, 2016. [Online]. Available: https://www.scientific.net/JBBBE.27.103. [Accessed: 20-Oct-2019].

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

[14] M. Grissinger, "Physical Environments That Promote Safe Medication Use," *Pharm. Ther.*, vol. 37, no. 7, pp. 377–378, Jul. 2012.

[15] J. Hopkings Medicine, "Study Suggests Medical Errors Now Third Leading Cause of Death in the U.S. - 05/03/2016," *John Jopkins Med.-News Publ.*, pp. 1–3, 2016.

[16] A. F. Yap, T. Thirumoorthy, and Y. H. Kwan, "Medication adherence in the elderly," *J. Clin. Gerontol. Geriatr.*, vol. 7, no. 2, pp. 64–67, Jun. 2016.

[17] K. N. Barker, E. A. Flynn, G. A. Pepper, D. W. Bates, and R. L. Mikeal, "Medication Errors Observed in 36 Health Care Facilities," *Arch. Intern. Med.*, vol. 162, no. 16, pp. 1897–1903, Sep. 2002.

[18] G. Fernández-Cardeñosa, I. de la Torre-Díez, M. López-Coronado, and J. J. P. C. Rodrigues, “Analysis of Cloud-Based Solutions on EHRs Systems in Different Scenarios,” *J. Med. Syst.*, vol. 36, no. 6, pp. 3777–3782, Dec. 2012.

[19] M. Poulymenopoulou, F. Malamateniou, and G. Vassilacopoulos, “Emergency healthcare process automation using mobile computing and cloud services,” *J. Med. Syst.*, vol. 36, no. 5, pp. 3233–3241, Oct. 2012.

[20] 10. Puri, C., Ukil, A., Bandyopadhyay, S., Singh, R., Pal, A., & Mandana, K, “iCarMa: Inexpensive Cardiac Arrhythmia Management--An IoT Healthcare Analytics Solution,” *Proc. First Workshop IoT-Enabled Healthc. Wellness Technol. Syst.*, pp. 3–8, Jun. 2016.

[21] M. M. Rathore, A. Ahmad, A. Paul, J. Wan, and D. Zhang, “Real-time Medical Emergency Response System: Exploiting IoT and Big Data for Public Health,” *J. Med. Syst.*, vol. 40, no. 12, p. 283, Oct. 2016.

[22] A. Shrivastava, *Exploring the smartwatch as a tool for medical adherence*. 2015.

[23] W. Baqir *et al.*, “Reducing unacceptable missed doses: pharmacy assistant-supported medicine administration,” *Int. J. Pharm. Pract.*, vol. 23, no. 5, pp. 327–332, Oct. 2015.

[24] K. Serdaroglu, G. Uslu, and S. Baydere, “Medication intake adherence with real time activity recognition on IoT,” in *2015 IEEE 11th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, 2015, pp. 230–237.

[25] N. E. Lee, T. H. Lee, D. H. Seo, and S. Y. Kim, “A Smart Water Bottle for New Seniors: Internet of Things (IoT) and Health Care Services,” *Int. J. Bio-Sci. Bio-Technol.*, vol. 7, no. 4, pp. 305–314.

[26] H. Eskelinen, “Improving the productivity of complex electronic systems design by utilizing applied design methodologies,” *IEEE Aerosp. Electron. Syst. Mag.*, vol. 16, no. 10, pp. 26–28, Oct. 2001.

[27] F. Wang and M. J. Hannafin, “Design-based research and technology-enhanced learning environments,” *Educ. Technol. Res. Dev.*, vol. 53, no. 4, pp. 5–23, Dec. 2005.

[28] A. Saini and P. Yammiyavar, “Weak eyesight therapy: A case study in designing an application for m-health systems,” in *2013 International Conference on Human Computer Interactions (ICHCI)*, 2013, pp. 1–8.

[29] M. Ahmad, “Designing for the Internet of Things: A paradigm shift in reliability,” in *2015 IEEE 65th Electronic Components and Technology Conference (ECTC)*, 2015, pp. 1758–1766.

[30] 23. Coetzee, L.; Eksteen, J., “The Internet of Things—Promise for the future? An introduction,” *Proc. 2011 IST-Afr. Conf. Proc. Gaborone Botsw.*, pp. 157–160, May 2011.

[31] D. Ruffieux, M. Contaldo, and C. Enz, “MEMS-based all-digital frequency synthesis for ultralow-power radio for WBAN and WSN applications,” in *2011 IEEE International Symposium of Circuits and Systems (ISCAS)*, 2011, pp. 157–160.

[32] L. Mainetti, L. Patrono, and A. Vilei, “Evolution of wireless sensor networks towards the Internet of Things: A survey,” in *SoftCOM 2011, 19th International Conference on Software, Telecommunications and Computer Networks*, 2011, pp. 1–6.

[33] N. Khalil, M. R. Abid, D. Benhaddou, and M. Gerndt, “Wireless sensors networks for Internet of Things,” in *2014 IEEE Ninth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*, 2014, pp. 1–6.

[34] C. Rotariu and V. Manta, “Wireless system for remote monitoring of oxygen saturation and heart rate,” in *2012 Federated Conference on Computer Science and Information Systems (FedCSIS)*, 2012, pp. 193–196.

[35] and and, “WSN based mobile u-healthcare system with ECG, blood pressure measurement function,” in *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2008, pp. 1533–1536.

[36] S. Tan, J. García-Guzmán, and F. Villa-López, “A wireless body area network for pervasive health monitoring within smart environments,” in *2012 IEEE Second International Conference on Consumer Electronics - Berlin (ICCE-Berlin)*, 2012, pp. 47–51.

[37] J. Wannenburg and R. Malekian, “Body Sensor Network for Mobile Health Monitoring, a Diagnosis and Anticipating System,” *IEEE Sens. J.*, vol. 15, no. 12, pp. 6839–6852, Dec. 2015.

[38] M. Shchekotov, “Indoor localization methods based on Wi-Fi lateration and signal strength data collection,” in *2015 17th Conference of Open Innovations Association (FRUCT)*, Yaroslavl, 2015, pp. 186–191.

[39] J. Youn, H. Ali, H. Sharif, J. Deogun, J. Uher, and S. H. Hinrichs, “WLAN-Based Real-Time Asset Tracking System in Healthcare Environments,” in *Third IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob 2007)*, 2007, pp. 71–71.

[40] W. Chen *et al.*, “Dynamic Indoor Localization Based on Active RFID for Healthcare Applications: A Shape Constraint Approach,” in *2009 2nd International Conference on Biomedical Engineering and Informatics*, 2009, pp. 1–5.

[41] F. Palumbo, P. Barsocchi, S. Chessa, and J. C. Augusto, “A stigmergic approach to indoor localization using Bluetooth Low Energy beacons,” in *2015 12th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2015, pp. 1–6.

[42] J. Wyffels, J. D. Brabanter, P. Crombez, P. Verhoeve, B. Nauwelaers, and L. D. Strycker, “Distributed, Signal Strength-Based Indoor Localization Algorithm for Use in Healthcare Environments,” *IEEE J. Biomed. Health Inform.*, vol. 18, no. 6, pp. 1887–1893, Nov. 2014.

[43] B. Jachimczyk, D. Dziak, and W. J. Kulesza, “RFID - Hybrid Scene Analysis-Neural Network system for 3D Indoor Positioning optimal system arrangement approach,” in *2014 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings*, 2014, pp. 191–196.

[44] S. G and S. G, “An Investigation on IoT Healthcare Analytics,” *Int. J. Inf. Eng. Electron. Bus.*, vol. 9, no. 2, pp. 11–19, 2017.

[45] M. Seera and C. P. Lim, “A hybrid intelligent system for medical data classification,” *Expert Syst. Appl.*, vol. 41, no. 5, pp. 2239–2249, Apr. 2014.

[46] A. M. Anter, A. T. Azar, and K. M. Fouad, “Intelligent Hybrid Approach for Feature Selection,” in *The International Conference on Advanced Machine Learning Technologies and Applications (AMLTA2019)*, 2020, pp. 71–79.

[47] P. Aruna Kumari and G. Jaya Suma, “FGANN: A Hybrid Approach for Medical Diagnosing,” in *Computational Intelligence and Big Data Analytics: Applications in Bioinformatics*, Ch. Satyanarayana, K. N. Rao, and R. G. Bush, Eds. Singapore: Springer Singapore, 2019, pp. 127–137.

[48] M. D. Petty, J. Kim, S. E. Barbosa, and J. J. Pyun, “Software frameworks for model composition,” *Model. Simul. Eng.*, vol. 2014, p. undefined-undefined, 2014.

[49] pubmeddev and H. Z. and G. DF, “A Review of Feature Selection and Feature Extraction Methods Applied on Microarray Data. - PubMed - NCBI.” [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/26170834. [Accessed: 14-Oct-2019].

[50] M. Abernethy, “Classification and clustering,” *IBM Developer*. .

[51] R. Rajesh, J. Maiti, and M. Reena, “Decision Tree for Manual Material Handling Tasks Using WEKA,” in *Ergonomic Design of Products and Worksystems - 21st Century Perspectives of Asia*, P. K. Ray and J. Maiti, Eds. Singapore: Springer Singapore, 2018, pp. 13–24.

[52] D. Morariu, R. Crețulescu, and M. Breazu, “THE WEKA MULTILAYER PERCEPTRON CLASSIFIER,” *Int. J. Adv. Stat. ITC Econ. Life Sci.*, vol. 7, no. 1, Mar. 2018.

[53] R. Soni, B. Kumar, and S. Chand, “Optimal feature and classifier selection for text region classification in natural scene images using Weka tool,” *Multimed. Tools Appl.*, Jul. 2019.

[54] V. Chaurasia, S. Pal, and B. Tiwari, “Prediction of benign and malignant breast cancer using data mining techniques,” *J. Algorithms Comput. Technol.*, vol. 12, no. 2, pp. 119–126, Jun. 2018.

[55] E. Alickovic and A. Subasi, “Normalized Neural Networks for Breast Cancer Classification,” in *CMBEBIH 2019*, 2020, pp. 519–524.

[56] A. Esteva *et al.*, “Dermatologist-level classification of skin cancer with deep neural networks,” *Nature*, vol. 542, no. 7639, pp. 115–118, Feb. 2017.

[57] M. M. F. Islam, R. Ferdousi, S. Rahman, and H. Y. Bushra, “Likelihood Prediction of Diabetes at Early Stage Using Data Mining Techniques,” in *Computer Vision and Machine Intelligence in Medical Image Analysis*, 2020, pp. 113–125.

[58] S. Gaikwad and N. Pise, “An Experimental Study on Hypothyroid Using Rotation Forest,” *Int. J. Data Min. Knowl. Manag. Process*, vol. 4, no. 6, pp. 31–37, Nov. 2014.

TURNITIN PLAGIARISM REPORT

Document Viewer

Turnitin Originality Report

Processed on: 11-Oct-2019 14:42 PKT
ID: 1190719519
Word Count: 22329
Submitted: 1

FinalThesis By Asra Superior

Similarity Index	Similarity by Source	
7%	Internet Sources:	5%
	Publications:	4%
	Student Papers:	6%

exclude quoted include bibliography exclude small matches mode: quickview (classic) report

Change mode print refresh download

1% match (student papers from 30-Sep-2017)
Submitted to University of Kent at Canterbury on 2017-09-30

1% match (Internet from 23-Aug-2019)
https://machinelearningmastery.com/how-to-run-your-first-classifier-in-weka/

<1% match (Internet from 07-Aug-2018)
http://www.ijsrp.org

<1% match (Internet from 09-Sep-2017)
https://kclpure.kcl.ac.uk/portal/files/60862912/2016_Almansour_Amul_Abdullah_1355080_ethesis.pdf

<1% match (student papers from 10-Oct-2019)
Submitted to Higher Education Commission Pakistan on 2019-10-10

<1% match (Internet from 02-Aug-2018)
https://ece.uwaterloo.ca/~z70wang/publications/ssim.pdf

<1% match (publications)
Lakshmi, V. Santhana, and M.S. Vijaya. "Efficient prediction of phishing websites using supervised learning algorithms", Procedia Engineering, 2012.

<1% match (Internet from 25-Jan-2018)
http://docplayer.es

<1% match (student papers from 10-Oct-2019)
Submitted to Higher Education Commission Pakistan on 2019-10-10

https://www.ideals.illinois.edu/bitstream/handle/2142/34587/Mitra_Modhurita.pdf?isAllowed=y&sequence=1

<1% match (publications)
"Internet of Things and Big Data Technologies for Next Generation Healthcare", Springer Science and Business Media LLC, 2017

<1% match (Internet from 18-Nov-2005)
http://homepages.ulb.ac.be

<1% match (Internet from 25-May-2016)
http://uir.unisa.ac.za

<1% match (Internet from 19-Sep-2017)
http://stat.cmu.edu

<1% match (publications)
RodrÍguez-Martos, Alicia, and Elena Santamariña. "Does the Short Form of the Alcohol Use Disorders Identification Test (AUDIT-C) Work at a Trauma Emergency Department?", Substance Use & Misuse, 2007.

<1% match (student papers from 04-Jun-2019)
Submitted to Yeungnam University on 2019-06-04

<1% match (student papers from 11-Feb-2019)
Submitted to Modern Education Society's College of Engineering, Pune on 2019-02-11

<1% match (student papers from 28-Mar-2019)
Submitted to Asia Pacific University College of Technology and Innovation (UCTI) on 2019-03-28

<1% match (student papers from 11-Sep-2009)
Submitted to Higher Education Commission Pakistan on 2009-09-11

<1% match (Internet from 10-Dec-2016)
http://goanna.cs.rmit.edu.au

<1% match (Internet from 25-Apr-2016)
http://academicjournals.org

<1% match (Internet from 15-Jun-2015)
http://users.cs.cf.ac.uk

<1% match (Internet from 05-Sep-2019)
https://link.springer.com/chapter/10.1007%2F978-3-319-92007-8_4

<1% match (Internet from 15-Sep-2009)
http://thesis.lib.ncu.edu.tw