DEVELOPMENT OF AN INTELLIGENT DECISION SUPPORT SYSTEM FOR PROMPT DIAGNOSIS OF EBOLA AND LASSA FEVER DISEASES.

BY

ADE-OJO TOLUWANI

(CSC/14/2209)

BEING A PROJECT REPORT

SUBMITTED TO

THE DEPARTMENT OF COMPUTER SCIENCE, FACULTY OF SCIENCE FEDERAL UNIVERSITY OYE EKITI, EKITI STATE NIGERIA.

IN PARTIAL FULFILMENT OF THE REQUIREMENT FORTHE AWARD OF BACHELOR OF SCIENCE (B.Sc) DEGREE IN COMPUTER SCIENCE.

FEBRUARY, 2019.

DEDICATION

This project is dedicated to GOD because without him I'm nothing. And because of him this project has been a success. I also dedicate this project to my family (Ade-ojo family) and more importantly to my fellow students in Federal University Oye-Ekiti. And also to my friends for their support emotionally and financially throughout the course of this research.

DECLARATION OF ORIGINALITY

This project is all my own work and has not been copied in part or in whole from any other source except where duly acknowledged. As such, all use of previously published work (from books, journals, magazines, internet, etc.) has been acknowledged within the main report to an entry in the References list.

I agree that an electronic copy or hardcopy of this report may be stored and used for the purposes of plagiarism prevention and detection.

I understand that cheating and plagiarism constitute a breach of University Regulations and will be dealt with accordingly.

COPYRIGHT

Signed: Date:

Office Stamp

CERTIFICATION

This is to certify that this project report titled “DEVELOPMENT OF AN INTELLIGENT DECISION SUPPORT FOR PROMPT DIAGNOSIS OF EBOLA AND LASSA FEVER DISEASES” was carried out by ADE-OJO TOLUWANI with the matriculation number CSC/14/2209 in partial fulfilment for the award of Bachelor of Science Degree in Computer Science, Federal University Oye –Ekiti, Ekiti State, Nigeria.

-------------------------------	------------------------------
DR. T.M FAGBOLA	**DATE**
Supervisor	

-------------------------------	------------------------------
Dr. OGUNLEYE G.O.	**DATE**
Ag. Head of Department	

ABSTRACT

Ebola Virus Disease (EVD) and Lassa fever are infectious viral diseases that are very deadly to mankind. These diseases, when handled lightly can quickly degenerate into deadly epidemics. Accurate and prompt diagnosis, and effective treatment of these infectious diseases is a very critical factor in their prevention and containment. The difficulty in differentiating between EVD and Lassa fever at their initial phase can result in wrong diagnosis which can be catastrophic. An intelligent decision support system can help make faster and more accurate diagnosis of these diseases.

In this research, a decision support system for diagnosis of EVD and Lassa fever was developed. Patient clinical history, demographic data, treatment data, and prognosis were obtained from Lagos State University Teaching Hospital (LASUTH) and Irrua Specialist Teaching Hospital, Irrua, Nigeria for 1000 cases of Ebola virus disease and Lassa fever respectively. The data was encoded in a way suitable for input into the learning algorithm. Conditional inference tree and Support vector classifier with sigmoid, polynomial and radial basis function kernel were used to build a classification model for the diseases. 10-fold cross validation technique was used for model validation. Evaluation was based on accuracy, precision, sensitivity, and recall.

Results show that the Conditional inference tree performed best with 99% accuracy and 367 true positive classifications. Support vector machine with Sigmoid and RBF kernels also achieved 99% accuracy, but with lower sensitivity scores.

A web based decision support system was built based on the Conditional inference tree. The developed system can enhance and facilitate decision making process by leveraging on stored

knowledge to correctly classify EVD and Lassa fever and assign necessary treatment to the patients.

TABLE OF CONTENT

LIST OF FIGURES

Figure Page

LIST OF TABLES

ACKNOWLEDGEMENTS

My genuine appreciation goes to God for his grace, mercy and his guidance throughout the course of my project work. All thanks to my parents Mr. and Mrs. Ade-Ojo for their help and financial care during the course of this research.

Importantly, I appreciate my Supervisor Dr. T.M. Fagbola for his fatherly love, feedback and support towards the realization of this project. In addition, my special thanks goes to the Head of the Department Dr. Ogunleye G.O., Dr. Fagbuagun, other all teaching staffs of the department, and non-teaching staffs for the support and contribution towards my progress in life. May God bless you all, Amen.

My appreciation also goes to my friends for their help towards this project.

CHAPTER ONE

INTRODUCTION

1.1 Background to the Study

The major challenge facing the healthcare industry is the provision of quality services at affordable costs (Obanas, 2013)**.** A quality service implies diagnosing patients correctly and treating them effectively (John, 2009). Poor clinical decisions can lead to disastrous results which is unacceptable. Medical diagnosis is known to be subjective, that is, it depends on the physician making the diagnosis (Resul, and Abdulkadir, 2008). Secondly, and most importantly, the amount of data that should be analyzed to make a good prediction is usually huge and at times unmanageable. In this context, machine learning can be used to automatically infer diagnostic rules from descriptions of past, successfully treated patients, and help specialists make the diagnostic process more objective and more reliable (Polat and Gunes, 2007).

The Decision Support System (DSS) terminology refers to a class of computer-based information systems including knowledge based systems that support decision making activities. The DSS that have been developed to assist physicians in the diagnostic and treatment process often are based on static data which may be out of date. A DSS which can learn the relationship between patient history, diseases in the population, symptoms, pathology of a disease, family history and test results, would be useful to physicians and hospitals.

The concept of DSS is very broad because of many diverse approaches and a wide range of domains in which decisions are made. To reduce the diagnosis time and improve the diagnosis accuracy, it has become more of a demanding issue to develop reliable and powerful decision support systems to support the yet and still increasingly complicated diagnosis decision process.

The medical diagnosis by nature is a complex and fuzzy cognitive process, hence soft computing methods, such as decision tree classifiers have shown great potential to be applied in the development of decision support system of Ebola and Lassa fever.

Ebola Virus Disease (EVD) also known as the Ebola hemorrhagic fever is a very deadly infectious disease to humankind (Ayten Kadanali and Gul Karagoz, 2014). Therefore, a safer and complementary method of diagnosis is to employ the use of an expert system in order to initiate a platform for pre-clinical treatments, thus acting as a precursor to comprehensive medical diagnosis and treatments. The deadly, scary spate and debilitating effects of the Ebola Virus Disease (EVD) in the West African sub-region, especially in 2014, left terrifying, untold hardships and discrimination mostly among the affected West African countries. Many are yet to fully recover from the Ebola scare and the psychological trauma it generated. It is a known fact that the Ebola Virus Disease is a very contagious and deadly disease.

Other problems associated with the disease are lack of proper knowledge in diagnosing and managing the disease especially among countries in Sub-Sahara Africa. In some cases, lack of proper training for medical experts to effectively and efficiently manage the disease constitutes a problem.

Also, Lassa fever is a viral hemorrhagic fever that was first described in 1969 in the town of Lassa in the North-East of Nigeria (WHO, 2016). It is endemic in the West African countries of Sierra Leone, Guinea, Liberia, and Nigeria. Cases imported to Europe indicate that Lassa fever also occurs in Côte d'Ivoire and Mali. The causative agent is Lassa virus, an RNA virus of the family Arenaviridae. Its natural host is the rodent Mastomysnatalensis, which lives in close contact to humans. Mastomys shed the virus in urine and contamination of human food is a likely mode of transmission. The virus may be further transmitted from human to human, giving rise to mainly nosocomial epidemics with case fatality rates (CFR) of up to 65%. However, most of the Lassa virus infections in the communities are probably mild.

Lassa fever is difficult to distinguish from other febrile illness in West African hospitals, especially at the initial stage. The frequent symptoms are pharyngitis, cough, gastrointestinal symptoms (Danny *et al*., 2012). Late signs of infection are effusions, facial edema, bleeding, and convulsion, coma, pleural and pericardial. At the extreme stage patients often experience shock, though bleeding usually not of a magnitude to produce shock. The only medication with a proven record of effectiveness in human is the Nucleoside Analogue Ribavirin (NAR). The efficacy of drug reduces if the treatment assignment started at day 7 or later, thus making diagnostic difficult for survival (Danny *et al*., 2012)

Lassa virus can be detected in blood at an early stage of illness. Death occurs about two weeks after onset of illness with fatal cases showing higher levels of viremia than those who survive. In survivors, virus is cleared from circulation about three weeks after onset of symptoms. IgM and IgG antibodies are detectable only in a fraction of patients during the first days of illness, and patients with fatal Lassa fever may not develop antibodies at all making early diagnostics critical for survival.

With these in views, there is need for a practical implementation of a complementary system that can diagnose and provide excellent recommendations to individuals in order to curb the spread of Ebola and Lassa fever diseases. Such system will also act as a supporting tool for medical experts and resident doctors in training.

This work presents a design and implementation of decision support system for the diagnosis of Ebola Virus and Lassa fever diseases.

1.2 Statement of the Problem

Lassa fever and Ebola virus diseases are extremely difficult to distinguish from each other especially at the initial phase of attack. This in turn can lead to wrong diagnosis and treatment process by the health stakeholders and physicians. However, current researches in this retrospect only focus on forecasting and spatio-temporal infectious diseases outbreak prediction (Praker and Stephen, 2017) , (Kyle B and Joshua Proctor, 2017), (Jantien A. and Jacco W, 2014), (Sun, Tsutakawa, and Kim, 2000), (Elisabeth Fichet-Calvet, Thomas Strecker, Stephan Olschlager and Lamina Koivogui, 2017) and (Babasola Olugasa, Eugene Odigie, Mike Lawani and Johnson F.). To savage this problem, this study seeks to develop a DSS for diagnosing Ebola and Lassa fever diseases.

1.3 Aim and Objectives

The aim of this study is to develop a decision support system for the diagnosis of Ebola and Lassa fever diseases to help manage wrong diagnosis. The specific objectives are to:

(i) Design a decision support system for the diagnosis of Ebola and Lassa fever diseases based on Support Vector Machine (SVM) and Conditional Inference Tree (CIT) algorithm using Unified Modeling Language Formalism.

(ii) Implement the design in (i) for diagnosis of Ebola and Lassa fever using R programming language.

(iii) Evaluate the performance of the SVM-based and CIT-based DSS for Ebola and Lassa fever diagnosis using F1 measure, accuracy, specificity and sensitivity as performance metrics.

(iv) Development of the model into web Application with Azure Machine Learning in R –

It is an interface that enables the deployment of web services with the execution of R code.

1.4 Significance of the Study

The outcome of this study will be useful for the doctors and other healthcare providers, patients and the general public given that it will facilitate accurate and prompt diagnosis / management of Ebola and Lassa fever.

This research will be a contribution to the body of literature in the area of design of decision support system for diagnosing and treating Ebola and Lassa fever, thereby constituting the empirical literature for future research in the subject area.

Furthermore, embarking on this research will reduce the mortality rate associated with Ebola and Lassa fever diseases.

1.5 Scope of the Study

The study will only consider the clinical parameters, signs and symptoms and line of diagnosis of Ebola and Lassa fever diseases in the Nigeria context only.

CHAPTER TWO

LITERATURE REVIEW

2.1 Clinical Decision Support System

Clinical decision support system is a branch of decision support system that helps in facilitating valid medical decision making processes; this chapter critically considers review of literature of clinical decision support system and explicitly explains decision support system.The process of identifying disease by analyzing its symptoms is often referred to as medical diagnosis; it could also be define as operation classifier embedding a decision making phases based on medical information (Musa, 2016),.

A decision support system is information based system that support and facilitate organizational decision-making activities. It serves the management, planning and operations levels of organization (mostly, mid and higher management staff) and help in decision making about problems that may be changing rapidly and not easily specified in advance- that is, unstructured and semi-structured decision problems. It can either be human-powered decision, fully computerized or combination of both (Wright, 2008).

While academics have seen decision support system as a tool to facilitate decision making process, the users of this system perceived it as a tool to easily accomplish organizational processes (Keen and Peter, 1990). Across different field the system have been define by some authors to include any system that might support decision making and some include decision making software components. According to Sprague (1980) he defined decision support system:

i. As a system that aimed at the less well structured, underspecified problem that upper level managers typically face;

ii. A system that attempts to combine the use of models or analytic techniques with traditional data access and retrieval functions;

iii. A system that specifically focuses on features which make it easy to be use by non-computer-proficient persons in an interactive mode; and

iv. A system that developed to emphasize flexibility and adaptability to accommodate changes in the environment and the decision making approach of the user.

Knowledge based systems is an aspect of decision support system. A properly designed system is an interactive software-based system intended to help decision makers deduce useful information from a combination of raw data, documents, and personal knowledge, or business models to identify and solve problems and make valid decisions. In this study decision support system was developed to help in facilitating medical decision considering some clinical features (Igwe Sylvester *et al*, 2013).

A clinical decision support system is an information system that support medical decision making activities developed in an application that helps healthcare practitioners in the analysis of data to improve decision making and patients care simultaneously. It focuses on using knowledge based management to derive clinical suggestion based on some set of constraint that consists of patients-related data. Clinical decision support system is a branch of decision support system commonly used to support business management. This system possesses the capacity of a developed workflow and assistance at the time of care (Ida Sim *et al*., 2001).

Data mining is an important aspect of this system; it is a part that keeps patient's medical history in conjunction with relevant clinical features. At such, analysis can help predict potential outcome, such as drug interactions, flag disease symptoms. (TechTarget, 2018), with the integration of data mining the system is guaranteed to proffer accurate predictions. Clinical decision support systems are designed to provide clinicians with knowledge and patient-specific information, presented at appropriate times to optimize decision-making and enhance healthcare (Goggin, Robert and Marcus, 2007). Because Ebola and Lassa fever requires different criteria to be met in order to suspect a diagnosis and refer patients for testing, the use of computer prompts similar to those that alert medical practitioners about issues in prescriptions are a promising avenue to explore (Lipton *et al*.,2004).

Decision support system applied for diagnosing and treating Ebola and Lassa fever could benefit and support clinicians at various stages in the care process, such as preventive care, diagnosis and implementing treatment (Lipton *et al.,* 2004). Some of the benefits of screening are identifying pre-symptomatic individuals at high risk for Ebola and Lassa fever, allowing for targeted screening based on exposure risk, helping in direct management and decision-making. The implementation of decision support system for preventive care on rare diseases is a motivating area to explore, since the optimization of screening practices could improve diagnosis at proper times and prevent delays in treatment of diseases (Goggin, Robert and Marcus, 2007).

Three common features in decision support system are the knowledge base, the inference engine and a mechanism to communicate with the user. The knowledge base is information in the form of rules, the inference engine consists of formulas that combine rules with patient data and the communication mechanism involves the input and output of data used in decision-making (Lipton *et al*., 2004).

These three features were important considerations in the design of a decision support system that could support treatment of diseases.

2.2 Emerging Infectious Diseases

2.2.1 Ebola virus

Ebola virus disease (EVD), also known as Ebola hemorrhagic fever (EHF) or simply Ebola, is a viral hemorrhagic fever of humans and other primates caused by Ebolaviruses (WHO, 2014). Ebola hemorrhagic fever (EHF) is one of numerous Viral Hemorrhagic Fevers (VHFs) including Lassa fever, Rift Valley Fever, Marburg Fever, Crimean-Congo Hemorrhagic fever, and yellow fever (Chandrakant Ruparalia, Curless, Trexler and Black, 2015). Ebola Virus Disease is caused by the Ebola Virus and endemic throughout sub-Saharan Africa. The disease was named after the Ebola River in the Democratic Republic of Congo (DRC) where the first case was recorded in a 44-year-old schoolteacher in 1976. Sporadic outbreaks have occurred since 1976 in DRC, Gabon, Uganda and Republic of Congo. The most recent outbreaks in Nigeria and Senegal were contained within weeks. (Chandrakant Ruparalia *et al.*, 2015).On the basis of available evidence, fruits bats of the family *Pteropdiae*are considered to be the natural reservoir of filoviruses including Ebola. (Chandrakant Ruparalia *et al*., 2015).

Direct transmission from reservoir or secondarily infected animals is rare; bush meat may be a risk. Person-to-person transmission of filoviruses (for example, Ebola) can occur by direct contact with body fluids/excreta (blood, urine, diarrhea, vomit, semen, milk) including percutaneous or mucous membrane routes. Ebola has spread by contact with symptomatic patients or with bodies, particularly in healthcare settings. Airborne spread has not been shown in outbreaks. Ebola virus is a potential bioterrorism agent. Flavi viruses (for example, dengue, and yellow fever) are primarily vector-borne. (Washington State Department of Health, DOH 420-126, March 2018*)*

Signs and symptoms typically start between two days and three weeks after contracting the virus with a fever, sore throat, muscular pain, and headaches. Then, vomiting, diarrhea and rash usually follow, along with decreased function of the liver and kidneys. At this time, some people begin to bleed both internally and externally (WHO, 2014). The disease has a high risk of death, killing between 25 and 90 percent of those infected, with an average of about 50 percent. This is often due to low blood pressure from fluid loss, and typically follows six to sixteen days after symptoms appear (WHO, 2014).

A person infected with the Ebola virus cannot pass it to others before any symptoms appear (Ebola: Minnesota Department of Health Factsheet, 2014).

The virus spreads by direct contact with body fluids, such as blood, of infected human or other animals. This may also occur through contact with an item recently contaminated with bodily fluids. Spread of the disease through the air between primates, including humans, has not been documented in either laboratory or natural conditions. Semen or breast milk of a person after recovery from EVD may carry the virus for several weeks to months. Fruit bats are believed to be the normal carrier in nature, able to spread the virus without being affected by it (WHO, 2014). Other diseases such as malaria, cholera, typhoid fever, meningitis and other viral hemorrhagic fevers may resemble EVD. Blood samples are tested for viral RNA, viral antibodies or for the virus itself to confirm the diagnosis.

Control of outbreaks requires coordinated medical services, alongside a certain level of community engagement. The medical services include rapid detection of cases of disease, contact tracing of those who have come into contact with infected individuals, quick access to laboratory services, proper healthcare for those who are infected, and proper disposal of the dead through cremation or burial. Samples of body fluids and tissues from people with the disease should be handled with special caution. Prevention includes limiting the spread of disease from infected animals to humans. This may be done by handling potentially infected bush-meat only while wearing protective clothing and by thoroughly cooking it before eating it. It also includes wearing proper protective clothing and washing hands when around a person with the disease.

There is no medication that cures Ebola and no vaccine to prevent it (Ebola: Minnesota Department of Health Factsheet, 2014). No specific treatment or vaccine for the virus is available, although a number of potential treatments are being studied (WHO, 2014). Supportive efforts, however, improve outcomes. This includes oral rehydration therapy (drinking slightly sweetened and salty water) or giving intravenous fluids as well as treating symptoms.

2.2.1.1 Structure and Characteristics of Ebola Virus

Ebola Virus is an infectious fatal disease that spread through contact with the infected body fluids by the virus whose normal host specie is bats or fruit bats. They are tubular (80 nm in diameter), matrix and nucleocapsid components that are approximately 970nm long.

They belong to the Filovirus family, and it resembles the length of a thread, glycoprotein is responsible for the attachment and the entrance of new host cells, it also serves as a medium for the virus to reside. During their biosynthesis glycoprotein were usually inserted and the outer envelope of the virion is derived by budding of host cell membrane.

The data used for this research was obtained from Lagos State University Teaching Hospital (LASUTH), Lagos and Ebola Emergency Operation Centre (EOC), Abuja during the peak of Ebola Virus from 20th July to 20th of October 2014 that the country was declared Ebola free with 34 and 966 cases collected from LASUTH and EOC respectively, it consists of the following information;

i. Patient Age: The age of the patient infected with Ebola virus
ii. Gender: The sex of the infected patient either a male or female

iii. Prognosis: The outcome of the patient after being attended to, survived or dead with the date it occurred.
iv. Clinical History: The symptoms each patient possesses at as the time of being infected
v. Management: The treatment, the medical practitioner assigns to the infected patient
vi. Date: The day the patient gets infected with Ebola virus.

2.2.1.2 Technical Diagnosis of Ebola

Technically, the only and reliable way of diagnosing Ebola virus is through the clinical method, a situation where blood sample is collected and tested in the laboratory settings. There are three different categories explained below:

i. Antigen test: This is a test to examine viral protein in the blood.
ii. Serological test: This is a test to examine antibodies gathered to oppose the virus, while
iii. Molecular test: this is a test to examine viral RNA (Ribonucleic acids)

These viral antibodies can reside in its host for years following the period of infection; this therefore render serology minimally effective as a diagnostic method in the acute cases, therefore molecular and antigens test have shown to be effective against acute virus has the period increases. However, no tests have demonstrated the ability to detect Ebola before it started giving symptoms.

2.2.2 LASSA FEVER

Lassa fever is an acute viral illness that occurs in West Africa. The illness was discovered in 1969 when two missionary nurses died in Nigeria. The virus is named after the town in Nigeria where the first cases occurred. The virus, a member of the virus family *Arenaviridae*, is a single-stranded RNA virus and is zoonotic, or animal-borne. (Fact Sheet: Centre for Disease Control and Prevention).Lassa fever is a rodent-transmitted viral hemorrhagic disease of global health concern.

The disease is endemic in West African and responsible for recurrent epidemics of acute hemorrhagic fever in parts of West Africa as well as sporadic disease in Europe, Asia and America (Ogoina, 2013).

Lassa fever has accounted for recurrent outbreaks of acute hemorrhagic fever in Nigeria since the discovery of the virus in Lassa town in the northeastern Nigeria in 1969. The prevalence of antibodies to the virus in Nigeria is 21% as compared to 8-22% in Sierra Leone and 4-55% in Guinea. In the last 50 years more than 28 states in Nigeria and the Federal Capital Territory have experienced one or more outbreaks of Lassa fever (Ogoina, 2013). The last outbreak of Lassa fever in Nigeria began in December 2011 and as at 17th August 2012, a total of 934 suspected Lassa fever cases, 147 Laboratory confirmed and 93 deaths (CFR 9.97%) were reported from 41 LGAs in 23 States (Ogoina, 2013).

Lassa fever, also known as Lassa hemorrhagic fever (LHF), is a type of viral hemorrhagic fever caused by the Lassa virus. Many of those infected by the virus do not develop symptoms (WHO, 2016). When symptoms occur they typically include fever, weakness, headaches, vomiting, and muscle pains. Less commonly there may be bleeding from the mouth or gastrointestinal tract. The risk of death once infected is about one percent and frequently occurs within two weeks of the onset of symptoms. Among those who survive about a quarter have deafness which improves over time in about half (WHO, 2016).

Lassa fever is transmitted mainly through contact with infected secretions of rats. Humans get infected when infected rat secretions (excreta or urine) make contact with non-intact skin (e.g. through cuts or sores) or mucous membranes, and by ingestion of food or liquid contaminated by infected secretions, as well as by inhalation of aerosolized viral particles (Richmond and Baglole,2003)

Human to human transmission of Lassa fever is common in hospital settings and usually follows contact with infected blood, urine, and other body secretions of patients with Lassa fever or through contact with contaminated hospital equipment's, including reused needles. There is also the risk of sexual transmission since the virus is excreted in semen for up to three months after recovery from an acute illness (Ogoina, 2013).

The last decade has seen the emergence and re-emergence of Viral Hemorrhagic Fevers (VHFs) in Nigeria and indeed in the West African sub-region. VHFs pose a great challenge to public health globally due to the high infectivity, morbidity and mortality associated with this group of diseases (Nigeria Center for Disease Control, 2017).

Alia*et al*. (2018) proposed a novel privacy policy single decision tree algorithm for clinical decision support system that assist healthcare for new symptoms diagnosis with the encryption through homomorphic encryption cipher of patients data to different networks using Internet of Things and nuancesto avoid decrypting of each other data since they all using same key pair , it was shown the novel algorithm outperformed the Naïve Bayes algorithm by 46.46% and this model was validated and also meet the requirement of hospitals and diagnosed symptoms.

One significant challenge in West Africa is differentiating between etiologies of febrile illness with similar initial clinical presentations, including malaria, influenza, dengue, yellow fever, and Lassa fever, with limited laboratory facility and reagent availability. Empiric treatment for presumed malaria or bacterial infection is often trialed and Lassa fever only suspected when a patient fails to improve with anti-malarial and antibiotic therapy (Raabe and Koehler, 2017).

In addition to these risks, there is no vaccine. Prevention requires isolating those who are infected and decreasing contact with the rats. Other efforts to control the spread of disease include having a cat to hunt vermin, and storing food in sealed containers. Treatment is directed at addressing dehydration and improving symptoms. The antiviral medication, ribavirin may be useful when given early. These measures improve outcomes. Descriptions of the disease date from the 1950s. Lassa fever is relatively common in West Africa including the countries of Nigeria, Liberia, Sierra Leone, Guinea, and Ghana. There are about 300,000 to 500,000 cases which result in 5,000 deaths a year.

2.2.2.1 Structure and Characteristics of Lassa fever

Lassa fever is a hemorrhagic disease caused by Lassa virus, the virus is a member of the arenavirus family and its source from transmission is rodent. The incubation period ranges from 6–21 days. When it is symptomatic, it starts with fever, general weakness and malaise, followed by headache, sore throat, muscle pain, chest pain, vomiting, diarrhea, cough and abdominal pain after few days, in rare cases facial swelling, bleeding from the mouth and low blood pressure may develop. The features of the virus carrier have been a serious issue overtime; most host species have been attached with Arenavirus. Therefore, understanding the geographical distribution is important to know the epidemiology of human infection. Before hemoglobin electrophorese was used for determining the type of specie, the rodent host of the virus was classified as mastomysnatalensis and antigen was found in Musgenera and Rattus increasing the possibility other rodent genera could also transmit the virus.

The scheme of classification is considered unsolved and the virus carrier specie makes it difficult in that there are 8 identical species recognized as the carrier and lives together where the disease is endemic, the doubt about the accurate natural reservoir with Lassa virus is considered a big challenge.

The Lassa fever data used for this study was obtained from the Centre for the Control and Management of Lassa fever Irrua Specialist Teaching Hospital, Irrua Edo State from 2010 to 2017 and it consist of 1000 data, it contains the following information

2.2.2.2 Technical diagnosis of Lassa fever

Technically, the only defined test for Lassa fever is the clinical based diagnosis by using Reverse transcription-polymerase chain reaction (RT-PCR) though in early stages of infection while using enzyme-linked immunosorbent serologic assays (ELISA) to examine Lassa antigens, IgG (Immunoglobulin G) and IgM (Immunoglobulin M) (MedicalNewToday,2018). However, there is a molecular diagnosis of Lassa fever where symptoms were used as a basis for identifying suspected cases. The development of a clinical decision support system consider the following symptoms; $38^{o}c$ fever for 2 days, exclude typhoid fever and malaria negative 1+ in thick smear, and some or one of the following symptoms chest pain, sore throat, headache, muscle pain vomiting and diarrhea (Danny *et al.*, 2012).

2.2.3 Differences and similarities between Ebola virus and Lassa fever

The carriers of Lassa fever and Ebola are rodent and bats respectively, there are cases of person-to-person transmission with Ebola but there has never been person-to-person transmission in the case of Lassa fever, both Lassa fever and Ebola are hemorrhagic fevers with symptoms of both diseases are similar which is a drive that birth this research, such as fever, headache, muscle pain, weakness, fatigue and vomiting, while we have dissimilarities in symptoms such as difficulty in swallowing and swollen airways while unexplained hemorrhage and unexplained bruising.

2.3 Decision Support System in Diagnosis of Ebola and Lassa fever Virus

It is a technical system developed to help in making decisions, consumes lesser time to solve problems, improve communication and collaboration among team members.

First off, the reason for developing this algorithm is to facilitate decision making in the sector in problem related to difficulty in diagnosis ofLassa fever and Ebola virus.

A case scenario of Nigeria in 2014 for instance where there was an outbreak of Ebola at the time of tackling Lassa fever, these calls for medical practitioner to search for possible means to detect either a patient has been infected with Lassa or Ebola virus in order to assign treatment but in the course of doing this lives were lost, so this calls for the need of a system that helps to make this quick, accurate and effective decisions in classifying a patient to be either Ebola or Lassa fever infected. This system has been integrated in a web application with such a flexible and scalable framework that enables easy accessibility and interaction between the system and user.

2.4**CLASSIFICATION ALGORITHM**

2.4.1**Conditional Inference Tree Technique**

Conditional Inference tree is a data mining method with recursive binary partitioning systems developed for the consideration of statistical measures for splitting covariates. Unlike, both Classification and Regression Tree (CART) and C4.5 Decision tree algorithm that perform an overall possible search, such that it maximizes the information measure of node impurity choosing the covariates with many possible splits. The two algorithm possess basically two fundamental problems; Selection bias and over fitting due to the absence of statistical significance in the algorithm. In view of these, Conditional Inference tree was developed with a framework embedded with defined permutation test that measure the association between responses and covariates which is the basis for unbiased selection among the covariates features. Ultimately, the reason for the selection of this model is because Decision tree cannot handle nominal variables which in our case are the independent variables, while in Conditional Inference tree there is an aspect for nominal variables at response and covariate levels (Hothorn, *et al* 2006).The following explained how Conditional Inference tree works;

1. The weights W test the global null hypothesis of independence between any of the X covariates and the response variables. If this hypothesis cannot be rejected, then stop. Otherwise select the covariates with the strongest association to response (Y) variable.
2. Choose a set $B^* \subset X_j$ in order to split X_j into different disjoint setsB^* and $\frac{X_j}{B^*}$. The weight W_L and W_R define the two subgroups with:

$W_{Li} = W_{Li}I\left(X_{ji} \in B^*\right)$ and $W_R = W_RI\left(X_{ji} \in B^*\right)$2.1

$for\ all\ i = 1, \dots, n\ (I(.)\ denotes\ the\ indicator\ function$

3. Recursively repeat steps 1 and 2 with modified weightsW_L and W_R , respectively.

Where,

B: Is the random dataset that will be model consisting of response and independent variables

W_L: Is the case weight associated with the right node after being split.

W_R: Is the case weight associated with the left node after being split. X_j

2.4.2 SUPPORT VECTOR MACHINE (SVM)

SVM is one of the supervised machine learning technique for solving regression and discrimination problems, it was first introduced by Cortes and Vapnik in 1990 for binary data classification (Cortes and Vapnik, 1995). Today, it is used in several research areas such as face recognition (Osuna, Freund, and Girosi, 1997), speaker recognition (Trabelsi and Ben , 2012), medical diagnosis (Bhatia, Prakash, and Pillai, 2008).

Support Vector Machines (SVM) is a method for classifying data with the use of hyper-planes for separating these data. The technique adopted in SVM is very easily understandable. If we have labeled data, SVM can be used to generate several separating hyper-planes such that the data space is divided into segments and each segment contains only one kind of data. SVM technique is generally useful for data with no regularity which means, data whose distribution is unknown.

2.4.2.1 **Mathematical Definitions of SVM**

The basic idea of SVM is to search for the optimal hyper plane that separate the inputs variable space by their target (class) variable. Figure 2.1 illustrate the high level of how SVM works, hyper plane also called decision boundary is the yellow dashed line that separate the data, the other two lines help to make the right decision boundary.

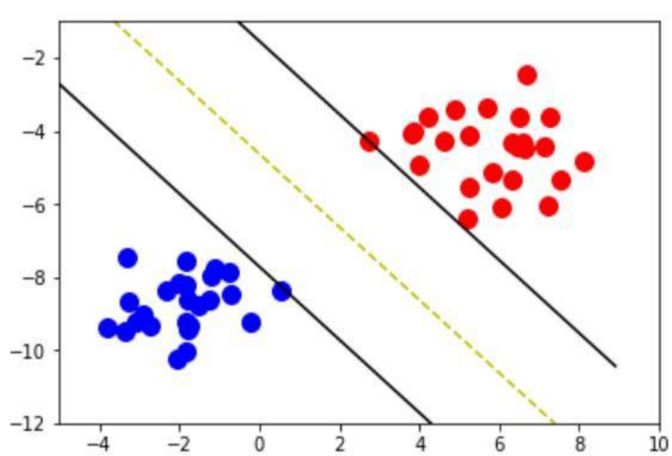


Figure 2.1: Support Vector Machine (Madhu S., 2008)

The purpose of applying this algorithm is to maximize the margin that best separate input variables into distinct sections such that the decision boundary is far away from the data point.

Margin: It is the distance between the left and right hyper plane.

So our aim is to maximize the hyper plane (yellow line) in figure 2.1.

$From\ (2.1), wx + b = 0$

$Where$:

W $= weight$

$w = weight\ vector$

$x = input\ vector$

$b = bias$

$For\ each\ input\ vector\ x_i\ either$:

$w.x + b \geq 1\ For\ x_i\ belonging\ to\ out\ come\ 1$ (2.2)

$w.x + b \leq 1\ For\ x_i\ belonging\ to\ out\ come - 1$ (2.3)

If $w.x + b = 0$ then the decision boundary has been reached; the yellow line in Figure 2.1

If $w.x + b = 1$then the (+) outcome hyper plane for all positive (x) points satisfy ($w.x + b \geq 1$).

If $w.x + b = -1$ then the (-) outcome hyper plane for all negative (x) points satisfy($w.x + b \leq -1$).

Data points in figure 2.1 can be diffused and separated in different ways such as;

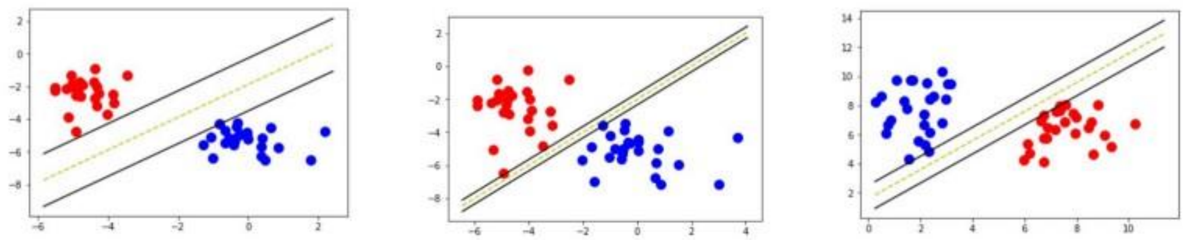


Figure 2.2: Different data points on SVM (Madhu (2008))

From Figure 2.2 to select the classifier with the best maximum margin it has to possess the minimum magnitude, it could be illustrated from figure 2.1

$D_1 = w^T x + b = 1; \quad D_1 = w^T x + b - 1 = 0;$ (2.4)

$D_2 = w^T x + b = -1; \quad D_2 = w^T x + b + 1 = 0;$ (2.5)

D_1: The line touched by the support vectors most times on the left hand side

D_2: The line touched by the support vectors most times on the right hand side

Therefore, the margin is the difference between D_1 and D_2 (D_1- D_2) such as;

$w^T x + b - 1$ - $w^T x + b + 1 = 0$ (2.6)

Solving algebraically,

we have $\frac{2}{|w|}$in other to increase the margin$(D_1 - D_2)|w|$must be minimized such as $\frac{1}{2}|w|$

2.4.2.2 **Kernel Selection in Support Vector Machine**

Kernel selection is important in the use of SVM to get rid of over-fitting in analysis, with an arbitrary dataset there is typically no certainty which kernel will work best. So, the linear kernel works best if the dataset is linearly separable; however, if the dataset isn't linearly separable, a linear kernel is not going to separate it. It should be noted that non-linear kernel can also work fine where linear kernel works, in this case a hyper-parameter search should be setup and compare different kernels to each other, based on the loss function or a performance metric such as accuracy, F1, ROC, AUC, specificity and sensitivity this could determine which kernel is best for the given analysis another distinct feature is that linear kernel is a parametric model while non-linear kernel isn't, such that the number of parameter grows with the number of the training set.

The proposed kernels to be adopted in this research are listed thus:

1. Linear

Linear classifier relies on dot product between vectors, define as

$$K(u, v) = u.v = u^T v \quad (2.7)$$

Provided the data point is mapped to high dimensional space via some transformation the dot product $\Phi: x \rightarrow \emptyset(x)$ the dot product then becomes:

$$K(u, v) = \emptyset(u)^T \emptyset(v)$$

Where

K kernel that analyses the pattern in a dataset

U the input variables in the dataset

V the outcome or class in the dataset

It has no parameter.

2. Radial Basis Function (RBF) also known as Gaussian kernel because it uses Gaussian equation in computation

Mathematical formula:

$$k\left(x, x^i\right) = \exp\left(-\frac{(\|x - x^i\|^2)}{2\sigma^2}\right) \quad (2.8)$$

$$k(x, x^i) = \exp\{-\gamma|x - x^i|^2\}$$

$\| x - x' \|^2$ Can represent the squared Euclidean distance between the two feature vectors. σ Is a free parameter, an equivalent definition involves a parameter. $\gamma = \frac{1}{2\sigma^2}$

$$\exp\left(-\frac{1}{2} \| x - x^i \|^2\right) = \sum_{j=0}^{\infty} \frac{(x^T x^i)^j}{j!} \, exp\left(-\frac{1}{2} \| x \|^2\right) \, exp\left(-\frac{1}{2} \| x^i \|^2\right)$$

$$= \sum_{j=0}^{\infty} \Sigma^{\infty} exp\left(-\frac{1}{2} \| x \|^2\right) \frac{x^{n_1 1}}{\sqrt{n_1! \ldots n_k!}} \, exp\left(-\frac{1}{2} \| x \|^2\right) \, exp\left(-\frac{1}{2} \| x^i \|^2\right) \quad (2.9)$$

Note: γ Is a parameter that sets the spread of the kernel

3. Polynomial

   For polynomial with degree (d)

   $$k(u, v) = (u^T v + c)^d \quad (2.10)$$

   The vectors in the input space are u and v that is train or test set computed from vectors of features with $C \geq 0$ as a free parameter balancing the effect of higher-order versus lower-order terms in the polynomial. When $C = 0$, the kernel is referred to as homogenous.

   In this kernel, an inner product in a feature based on some mapping φ relate with k.

   $$k(u, v) = (\varphi(u), \varphi(v))$$

Sigmoid: It is a mathematical function with a sigmoid curve “S” shaped,

$$\boldsymbol{S_u} = \frac{1}{1+e^{-x}} = \frac{e^x}{e^x + 1} \quad (2.11)$$

It is a real function defined as a non-negative derivative at each point and for real input values.

2.5 Related Works on Application of Decision Support System

The table below presents some previous works on decision support systems used for performing medical task.

S/N	WORK	AUTHOR	YEAR	DISEASE ADDRESSED	COUNTRY OF INCIDENCE	METHOD USED	PERFORMANCE METRIC	RESULT OBTAINED	STRENGTHS	WEAKNESSES
1	Belief rule-based system for clinical risk assessment of cardiac chest pain	Gulian , Dong-Ling, Richard-Bo , Kevin, Simon	**2011**	Chest Pain		Rule-based inference methodology	Found to perform extremely well	Prototype CDSS can deal with uncertainties in both clinical domain	Automatic update of belief rule base (BRB)	Dependent on large data input
2	upSCALE mHealth System Strengthening for Case Management and Disease Surveillance	MassoudMoussavi, Kent, Vilas, Aminata , and YazouméYé.	**2017**	Malaria	Mozambique	upSCALEmHealth System Strengthening for Case Management and Disease Surveillance	This system will assist with data-driven decision making regarding APE program investments, surveillance, and responses to malaria, and early detection	The CommCare application provides image and audio guidance for APEs to assess, classify, treat, and refer patients.	The tool is highly flexible and scalable. It relies on mobile phones, which are ubiquitous, even when electricity may not be.	The tool is highly dependent on ability of national-level officials to use the aggregated data delivered by the tool to shape successful policies.

							of disease outbreaks.			
3	Primaquine Roll Out Monitoring Pharmacovigilance Tool	Massoud *et al*.,	**2017**	Plasmodium Falciparum	Swaziland	Primaquine Roll Out Monitoring Pharmacovigilance Tool, PROMPT	Supports the surveillance of possible adverse events following treatment with single low-dose primaquine; and database compiling recorded information, such as patient characteristics and malaria diagnosis and treatment.	In the pilot, dosage safety was confirmed, and Swaziland's National Malaria Control Program was empowered to adopt the WHO recommendation.	In the long term, PROMPT may be able to assist in treating primaquineside effects through its flow chart treatment algorithm	However, the cost of maintaining the program in terms of person-hours will likely outweigh the benefit.
4	Innovations at Scale for Community Access and	Massoud *et al*.,	**2017**	Malaria	Mozambique and Uganda	Innovations at Scale for Community Access and	Minimum theoretical requirements are one APE worker	The data are collected by an APE, transmitted to a	The tool is highly flexible and scalable. It relies on	A non-functional search option, and its

	Lasting Effects					Lasting Effects (inSCALE)	with a compatible mobile phone and access to the Internet.	server, and then analyzed by supervisors to aid in improving APE efficiency and performance	mobile phones, which are ubiquitous, even when electricity may not be, as documented by the Pew Research Center (2015).	dependence on the availability of electric power and a stable internet connection
5	Lives Saved Tool	Massoud *et al*.,	**2017**	modeling child and maternal mortality	Over 90 Countries	Lives Saved Tool, LiST	The tool can be used at the national and global levels and is available on online	The tool was used to examine which interventions contributed to a change in mortality based on coverage measured in DHIS 2 and Multiple Indicator Cluster	LiST has a user-friendly intuitive interface and provides research-based default data.	The tool is integrated into a proprietary software package (Spectrum) that limits access. The LiST model does not include environmental, economic, and

								Surveys and to support advocacy at the local, national, and global levels.		social factors and does not use an inference mechanism to identify key constraints.
6	Intermittent Prevention and Treatment of Infants Decision Support Tool	Massoud *et al*.,	**2017**	Malaria	African Countries	Intermittent prevention and treatment of infants (IPTi) DST	The IPTi DST provides graphical information on the predicted age distributions of patients with clinical malaria, those admitted to hospital with malaria parasites, and those who will die due to malaria.	The tool can identify whether a malaria control intervention that targets infants is appropriate.	The IPTi DST is based on estimates, and its use is limited to infant populations.	The limitation of the tool is that it can be applied to only a small segment of the population.

7	Disease Data Management System for Enhancing Decision Support for Vector-Borne Disease Control Programs	Massoud *et al*.,	**2017**	Vector-Borne Disease	Benin, Equatorial Guinea, Ethiopia, Ghana, India, Mali, and Zambia.	Disease Data Management System, DDMS	The DDMS offers several unique features that can support the global goals of malaria elimination and control.	DDMS as a multi-disease system facilitates the integration of vector control programs, which can bolster neglected tropical disease elimination efforts and facilitates cross-border collaboration and collective decision making.	The DDMS is flexible and can be adjusted for any vector-borne disease control program. The tool can support decision making from malaria control through elimination phases.	In at least one country (Zambia), the tool was not able to maintain momentum, in part due to support and maintenance issues, and is no longer used.
8	GIS-Based Decision Support System	Massoud *et al*.,	**2017**	Malaria	**Zambia**	GIS-Based Decision Support System	The tool was used in monitoring malaria control	The tool provides an opportunity to integrate up-to-	The GIS-based DSS effectively uses spatial analysis	The tool limits the availability of accurate raw data; limits its

							programs in Zambia to introduce new dimensions to the understanding, prediction, analysis, and dissemination of spatial relations between disease, time, and space.	date information, local knowledge, and historical trends to identify areas where assistance is needed most and options for action.	using GIS technology to control and predict vector-borne malaria transmission, monitor insecticide resistance and impact of interventions, integrate operational and logistical data, and improve regional stratification, leading to better distribution of resources and interventions.	focus to vector control; does not have models to identify environmental, social, and other constraints; and does not offer a knowledge management component.

9	Malaria Decision Support System (MDSS)	Massoud *et al*.,	**2017**	Malaria	Kenya	Malaria Decision Support System (MDSS)	MDSS is a best practice, continuous surveillance system that integrates monitoring and evaluation data from a malaria control program and presents them in a web-based, real time geographic format to assist with intervention planning in Kenya.	The MDSS contains dynamic query tools that allow users to interpret and present data in formats tailored to their needs	The MDSS is a highly customizable tool that aids decision making at subnational and national levels. The tool provides continuous surveillance, monitoring, and evaluation of malarial control programs, including automated alerts. The MDSS is integrated with	The tool does not have a knowledge management component and lacks models to assess the relationships among various factors.

									other health information systems for data input and output, such as DHIS 2	
10	Decision Support Tools for Malaria Prevention and Treatment	Massoud *et al*.,	**2017**	Malaria	Solomon Islands and Vanuatu	GIS-based SDSS	The SDSS framework demonstrates how geospatial systems can support a progressive malaria elimination campaign. The platform rapidly collects, stores, and extracts essential data throughout key phases of	The framework actively locates and classifies transmission to guide swift and appropriate responses.	The SDSS makes extensive use of mobile devices for data collection and uploading to the SDSS database. The system helped with rapid reporting of confirmed cases by household for high-resolution	The SDSS requires purchase of subscriptions because of proprietary software and extensive specialized training needed for data collectors and users as well as system operations and maintenance.

							program implementation.		mapping of areas of interest for targeted response. The SDSS was used at both the national and field levels.	In addition, the system does not have a knowledge management component.
11	Decision Support Tools for Malaria Prevention and Treatment	Massoud *et al*.,	**2017**	Malaria	Bhutan	Spatial Decision Support System. SDSS	It was used to support the distribution of LLINs and IRS in the two subdistricts of SamdrupJongkhar District and was subsequently used to support RACD in the two subdistricts	Informants indicated that the SDSS was an improvement on previous methods for organizing LLIN distribution, IRS, and RACD, and could be easily integrated into routine	The SDSS provides a modernized, high-resolution mapping capacity to support the operational management of scaled-up interventions.	Much of the data are still collected manually, entered in Microsoft Excel, and then uploaded to the SDSS

							of SamdrupJongkhar and two additional subdistricts in Sarpang District.	malaria and other vector-borne disease surveillance systems.		
12	Decision Support Tools for Malaria Prevention and Treatment	Massoud *et al*.,	**2017**	Malaria	Kenya, Tanzania, Uganda	Malaria Decision Analysis Support Tool. MDAST	MDAST calculates the outcomes of the user-defined health delivery strategy by combining parameters describing the malaria context with the health delivery decisions in a systematic modeling framework.	The tool facilitated informed decision making and evidence-based malaria policy development.	MDAST is a comprehensive tool, allowing up to 48 input parameters. It provides evidence-based default values for the parameters where local data are lacking, unreliable, or unavailable to the user.	MDAST requires substantial training and technical support, and the underlying modeling is highly complex.

13	Design and Identification of Tuberculosis using Fuzzy Based Decision Support System	NavneetWalia, and Rahul	**2015**	Tuberculosis	India	Fuzzy inference system	Varying symptoms of 35 patients, the fuzzy basis dependent technique is utilized, which potentially reduces the conservatism of obtained results.	fuzzydiagnosability for tuberculosis of bacterium and formalize reasoning in rule based system.	The designed system can be extended for any number of inputs	The design is only FIS focused
14	Decision Support System for Malaria and Dengue Disease Diagnosis (DSSMD)	Priynka , Singh~~2~~, Manoj and	**2013**	Malaria and Dengue	India	Fuzzy logic toolbox	The performance of the system was analyzed by comparing the result of DSSMD with the clinical report of the patients. Total 69 patient's data	Decision support system aids the diagnosis of disease on the basis of symptoms of disease.	The performance of the system was analyzed by comparing the result of DSSMD with the clinical report of the patients.	Diagnosis of disease is solely based on the non -clinical symptoms of the disease using Artificial intelligence.

							was analyzed in which 35 patients of malaria and 34 patients of dengue disease–Out of 69 patient's data 63 results are positive			
15	Decision Support System for Precluding Coronary Heart Disease (CHD)	Cinetha, Uma	**2014**	Coronary Heart Disease (CHD)	India	Data Mining technique	The proposed system is to build the Decision Support System for precluding Coronary Heart Disease (CHD) using data miningtechniques to identify the level of risk in	This system helps the patients in take precautionary actions to stretch their life span and to assist medical practitioners to diagnose and predict the probable	This system which predicts the possibility of heart disease risk of patient for the next ten years for prevention using clustering algorithm.	Data quantity required.

							coronary heart diseases.	complications well in advance.		
16	CDSS for Osteoporosis	Monika ~~.~~, ., Kevin ~~t~~., Mark., Christine ., David ~~.~~, Sharon ~~.~~	**2011**	Osteoporosis		Osteoporosis disease management tool, ODMT			Improved self-management of Osteoporosis	Threat to internal validity
17	Fuzzy-based DSS for Coronary Heart Disease	Adel , Raja , Roziati	**2012**	Coronary Heart Disease		Multi-objective algorithm to optimize FRBS	Found to improve accuracy and transparency	Humanly understandable rules, able to identify uncertainty of cases	Can specifically improve the ability of FRBS	Threat to internal validity
18	Decision Support System for Diabetes Mellitus	Shaker and Mohammed	2016	Diabetes Mellitus	Egypt	Novel fuzzy KI-CBR framework that handles and exploits imprecise and encoded medical	Semantic performance of 97.67%	The hybridization of CBR with fuzzy ontology and medical ontologies is the most suitable techniques for	Tested to solve complex problems that cannot be solved by traditional systems	Ability to handle termoral data and case adaptation process.

						knowledge through the effective integration of fuzzy logic in the ontology-based CBR paradigm		solving medical diagnosis		
19	Interoperable clinical decision support system for early detection of SIRS in pediatric intensive care using openEHR	Antje , Birge, Erik , Michael , Philip, Thomas	**2017**	Systemic Inflammatory Response Syndrome (SIRS)	Germany	Designed an interoperable concept which enables an easy integration of CDSS across different institutions, by using openEHR Archetypes, terminology	Technical capabilities of the system were evaluated by testing the prototype on 16 randomly selected patients with 129 days of stay and comparing results with the	The use of the openEHRArchytypes and AQL, a feasible approach to bridge the interoperability gap between local infrastructures and CDSS	Approach allows the CDSS to be implemented at other institutions without further modifications of queries or rules	The data has to be sufficient

						bindings and the Archetype Query Language (AQL)	assessment of clinical experts			
20	Healthcare Decision Support System for Administration of Chronic Diseases	Ji-In, Jung and Un-Gu .		Chronic Diseases		Rule based inference methodology	A recommendation message is output through the Web service by receiving patient information from the hospital information recording system and object attribute values as input factors	The system can verify patient behavior by acting as an intellectualized backbone of chronic diseases management	Can help prevent secondary diseases	

Several studies have been conducted on the design and implementation of DSS for the diagnosis and treatment of various diseases, but none has been conducted for the diagnosis and treatment of Ebola and Lassa fever. However, some of the available studies will be reviewed and presented in this section.

A multimedia Based Clinical Decision Support System for Diagnosis of Chronic Heart Diseases (CHLD-MMCDSS) was developed make a meaningful life out of industrial workers undergoing different phase of Chronic Heart Diseases and to aid medical checkup and it uses different components such as Model Base Management (MBM) System, Medical Data Base Access and Management (MDBAM) System, Central Medical Vision Navigator (CMVN) Board, Clinical Vision Technology (CVT) Base comprising of Case Base Reasoning Algorithm based Case Base Reasoning Desk (CBR-Desk) Multi Media Medical Communication (MMMC) Desk, Chronic Disease Queries Support (CDQS) Server and Medical Decision Exchange (MDE) Server based Dialog Management (DiM) System and Clinical Decision Making and User (CDMU) Desk in its analysis (Tomarand, 2013).

Tomar and Singh (2013), conducted Cardio Informatics Portal of Clinical Decision Support System for Diagnosis of Chronic Heart Diseases (CHLD-MMCDSS) for medical checkup of operational workers in pain of Chronic Heart disease, four components were utilized for diagnosing by Medical Diagnosis Capsule (MDiC) of Model Base Management (MBM) System namely Coronary Artery (CAD) Module, Rheumatic Valvular (RHE) Module, Chronic CorPulmonale (CCP) Module, and Congenital (CON) and was developed using Microsoft Visual Studio and SQL Server.

Flores (2015) assessed the feasibility of a clinical decision support system CDSS for the Lynch Syndrome screening process followed in primary care settings. The main objective was to design a CDSS application modeling the clinical guidelines with openEHR (Electronic Health Records) and Guideline Definition Language (GDL). The secondary objectives were to identify the requirements in terms of clinical data, archetypes and rules, and validate them to ensure they delivered the expected results. A qualitative analysis was followed to evaluate the case study of the screening process of Lynch Syndrome. The phases followed were requirements analysis, design and development and testing. The clinical guidelines were analyzed to map the requirements to open EHR archetypes and rules that were used during the development of GDL guidelines. Mock data on 25 patients was used for validation.

Flores (2015), designed using openEHR archetypes three GDL guidelines. The screening stageswas modeled using independent data from family medical history to ascertain the recommendation of referral testing. This system was validated with sample medical records and the obtained results were correct as expected. In addition a user interface was developed to envisage user interaction. This system offer the capacities of designing a clinical decision support system that can interact with patient's screening stages and support accurate referral of Lynch Syndrome.

According to Flores (2015), three GDL guidelines were developed with openEHR archetypes. They modeled the screening process using input data from the patient and relative's medical history to determine if referral for testing was recommended. They were validated with mock patient data and results were accurate with the expected outcome. A user interface prototype was designed to visualize user interaction. OpenEHR and GDL offer the capabilities of developing a CDSS that can model a patient's screening process and support accurate referral of Lynch Syndrome. The architecture of OpenEHR provides the flexibility of further adapting the system to new requirements and additional features.

In a research by Oluwagbenga, Folake and Abimbola (2016), an Ebola fuzzy informatics system was developed for the purpose of diagnosing and providing useful recommendations to the management of the EVD in West Africa and other affected regions of the world. It also acts as a supplementary resource in providing medical advice to individuals in Ebola–ravaged countries. This aim was achieved through the following objectives: gathering of facts through the conduct of a comprehensive continental survey to determine the knowledge and perception level of the public about factors responsible for the transmission of the Ebola Virus Disease; develop an informatics software based on information collated from health institutions on basic diagnosis of the Ebola Virus Disease-related symptoms; adopting and marrying the knowledge of fuzzy logic and expert systems in developing the informatics software.

Necessary requirements were collated from the review of existing expert systems, consultation of journals and articles, and internet sources. Online survey was conducted to determine the level at which individuals are aware of the factors responsible for the transmission of the Ebola Virus Disease (EVD). The expert system developed, was designed to use fuzzy logic as its inference mechanism along with a set of rules. A knowledge base was created to help provide diagnosis on the Ebola Virus Disease (EVD). The Root Sum Square (RSS) was adopted as a fuzzy inference method. The degree of participation of each input parameter was shown using the triangular membership function and the defuzzification technique used is the Center of Gravity (CoG).

The resulting software produced a user-friendly desktop-based, Windows-based, application and the tools used were explained in the results section in three (3) separate phases. First, a comprehensive online survey was conducted over a period of about 3–9 months. 100 Participants participated in the survey on the perception and knowledge analysis of different individuals about Ebola Virus Disease (EVD) transmission factors.

31% of the participants didn't know that there is presently no cure for Ebola. 28% believed that there is presently a cure for Ebola. 43% agreed that Ebola is both air-borne and water-borne, while 33% disagreed, 24% do not know. 23% believed that insects and mosquitoes can help in transmitting the Ebola Virus Disease (EVD), while 30% were completely ignorant. It was noticed that ignorance was a major limiting factor among some participants. Second, a test was conducted among 45 people. Results from a comprehensive testing of the Ebinformatics software by allowing users to operate and use the software, revealed that 60% of them were satisfied, and while 16% were not satisfied with the software, and while 24% were indifferent. 69% of the users were in agreement that Ebinformatics was supportive, 20% disagreed, while 11% were indifferent. 67% found the software easy to use, 13% disagreed, while 20% were indifferent.

Third, the output of the software, showing the various diagnosis and recommendations interfaces were presented. Recommendations were also given with respect to how the system can be extended, and further improved upon.

Ariel (2002) used Fuzzy Support Vector Clustering to identify heart disease. This algorithm applied a kernel induced metric to assign each piece of data and experimental results were obtained using a well-known benchmark of heart disease. Ischemic-heart-disease (IHD) -Support Vector Machines serve as excellent classifiers and predictors and can do so with high accuracy. In this, tree based: classifier uses nonlinear proximal support vector machines (PSVM).

Hela Ltifi *et al*., (2016) evaluated visualization generated by intelligent support systems using fuzzy-logic based method. This study was based was on the system that enable discovery of new trend in data to generate useful knowledge for decision making without any level of numeric measurement such as low, medium and high rather the system identify uncertainty and imprecise users' evaluations in a fuzzy form. Each identification evaluating a particular criterion is converted into the form of fuzzy measure as entered in a fuzzy controller and the outcome is extracted using the inferential method. Therefore, this system allows simultaneously evaluation of the visualization by dealing with both linguistic knowledge and numeric data.

Shaker *et al*. (2014) proposed an open and distributed clinical decision support system that leverage on Electronic Health Record (EHR), data mining methods, clinical databases, available technologies, domain expert knowledge bases and decision-making standard for health care providers. It was deduced from their study that each knowledge base specializes in subjective domain and the model attain cooperation, interoperability and integration between the knowledge bases, in addition the system ensures that all knowledge connected are being updated by connecting data mining system to each local knowledge base.

Morris (2019) explained the safety of clinical environments with decision support tool. He connotes the safety in the clinical environment reduces the tendency of harm, on the basis that improve the probability of actions that increase favorable outcomes. Though he also noted that explicit decision support system will also harm clinical environment. Therefore, he suggested a system that integrated a synthesis of thought and consensus as a complement to the individual decision-making freedom of the past such that the decrease in variation and increase in compliance with evidence-based recommendation will reduce and improves patient safety.

Mathupanee *et al*., (2019) applied machine learning techniques to guided targeted and locally-tailored empiric antibiotic prescribing in children using several machine learning algorithms. He discovered machine learning algorithms to patient data can provide highly informative predictions on antibiotic susceptibilities to guide appropriate empirical antibiotic therapy.

Roosan *et al*., (2015) discussed the complexity of clinical reasoning in infectious diseases using qualitative thematic analysis, that is interview of coauthors to independently note relevant concepts attached with complexity, cognitive goals, adaptive strategies and sense-making.

From the study, it was discovered that decision complexity factors consist of lack of comprehension of the situation, dealing with social and emotional pressures such as anxiety and fear. Therefore, based on this study it was discovered that designing future decision support system for the management of complex patients should give attention to the cognitive strategies to deal with decision complexity found in this study.

Niclas Johansson *et al*., (2018) designed a system for the validation of pre-hospital decision support system for the Emergency Medical Services (EMS), thus allow the discovery of patients with critical infections conditions, this was accomplished with gathering of patient transported by the EMS from electronic patient categories care record system to know previous patient categories, then allow medical expert give respective suggestions and advice for improvement of decision support system thereafter evaluation and validation of the system followed. It was discovered the system give accurate decision support to pre-hospital emergency nurses when mobilizing patients to the peak level of care.

Elalfi, Fouda, and Atta (2016) developed an intelligent system for the diagnosis of some children's diseases to assist both inexperienced and fresh healthcare graduates with medical concepts and knowledge base from various sources such as medical images, knowledge base and domain knowledge base. The system helped in facilitating diagnosis decision by decision makers.

Nassim Douali *et al*., (2015) investigated and evaluated a model framework decisions diagnostic based on a semantic web approach and cognitive process with the use of Urinary Tract Infection

(UTI). Thus, the system proves efficient and effective such that it proposes appropriate diagnosis for each individual case as it diagnose based on record in the database.

Panagiotis *et al* (2014) proposed a clinical decision support system that consists of artificial neural network, a system that intelligently merge the outcome of both classical and ancillary techniques to enhance diagnostic accuracy. The developed system demonstrated high sensitivity (89.4%) and specificity (97.1%) for the detection of neoplasia grade 2 or worse.

Polat and Gunes (2004) designed an expert system to diagnose the diabetes disease based using principal component analysis and adaptive neuro-fuzzy inference system, it consist of two phases. In the first phase, diabetes dimension dataset has 8 features was reduced to 4 features using PCA. In the second phase, the diagnosis of diabetes disease was conducted with ANFIS classifier and the accuracy generated from the system was 89%.

Alexis, Subanatarajan, and Bartz (2017) developed a system that uses conditional inference trees because of its flexibility and statistical fundamental in split selection also its interaction with the data to analyze relevant electrical motor features to monitor the health condition of motors and he concluded that the model was good enough for the study also recommended that much attention should be given to data cleaning because it is a medium that pave way for the influence of the model on the dataset.

CHAPTER THREE

METHODOLOGY

3.1 Research Approach

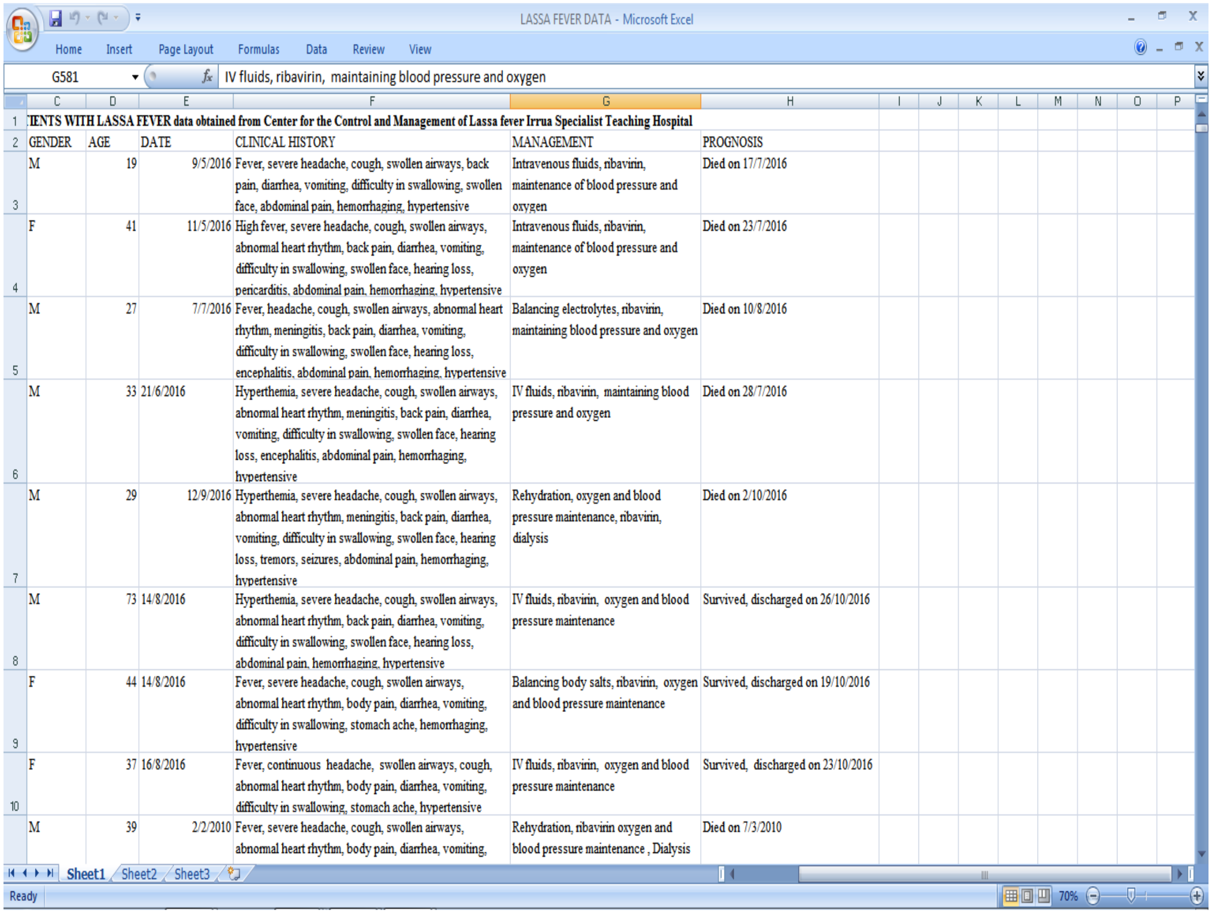
This chapter covers the methods applied in the accomplishment of Decision Support System development for diagnosis and treatment of Ebola and Lassa Fever diseases, the medical practitioner will treat the patient since the model had been equipped with necessary methods in classifying patient based on their respective clinical features, it also explains the method of application of machine learning algorithm to correctly predict patient's ailment with the provision of patient's data.

The following process would be adopted in the implementation of the algorithm.

1. Data Collection
2. Data Preprocessing
3. Model Validation
4. Conceptual Framework
5. Data Partitioning using 10-fold strategy
6. Model Development
7. Design of a DSS for Lassa fever and Ebola diagnosis with CIT and SVM.
8. Model Evaluation

3.1.1 Data Collection

The Ebola and Lassa fever data were collected from Lagos State University Teaching Hospital (LASUTH) and Irrua Specialist Teaching Hospital, Irrua, Nigeria respectively. The data was merged and converted to Comma-Separated Value (CSV) file in excel for importing, manipulation and analysis in R programming. Figure 3.1 display the unstructured form of the data.


LASSA FEVER DATA - Microsoft Excel

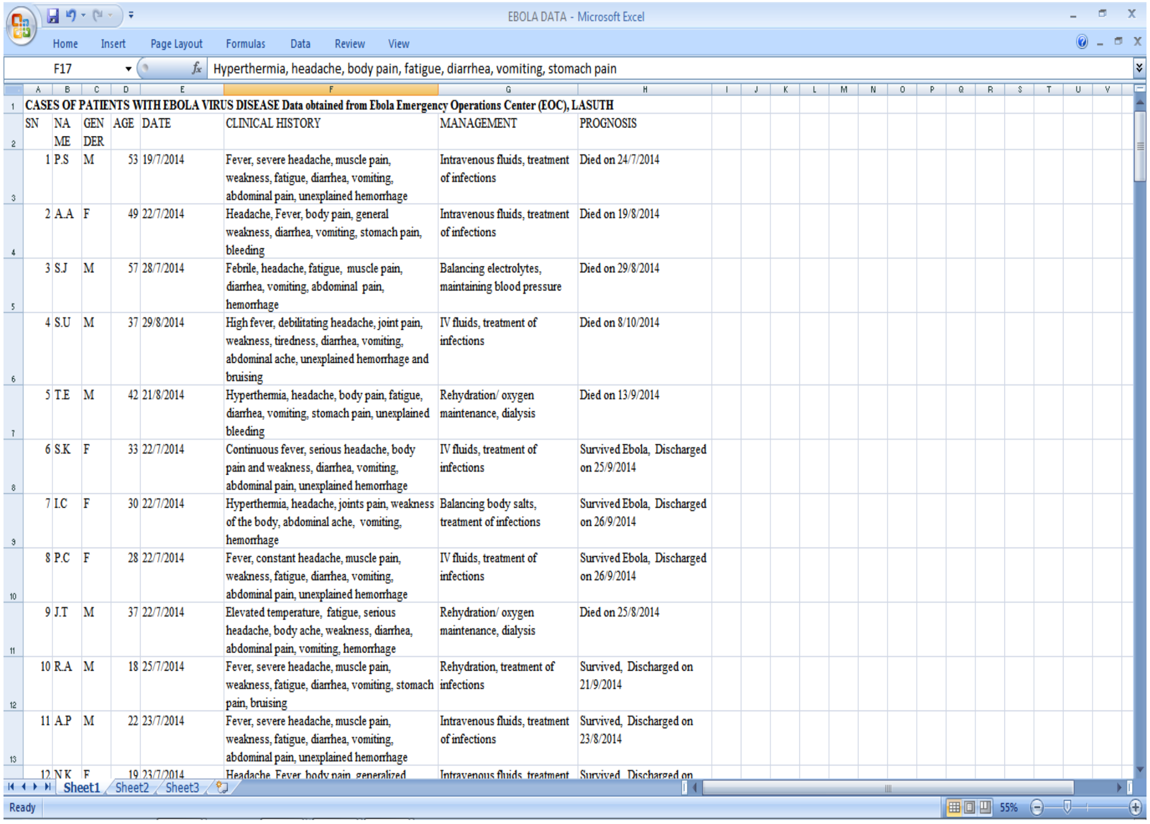
G581 IV fluids, ribavirin, maintaining blood pressure and oxygen

	C	D	E	F	G	H
1	IENTS WITH LASSA FEVER data obtained from Center for the Control and Management of Lassa fever Irrua Specialist Teaching Hospital					
2	GENDER	AGE	DATE	CLINICAL HISTORY	MANAGEMENT	PROGNOSIS
3	M	19	9/5/2016	Fever, severe headache, cough, swollen airways, back pain, diarrhea, vomiting, difficulty in swallowing, swollen face, abdominal pain, hemorrhaging, hypertensive	Intravenous fluids, ribavirin, maintenance of blood pressure and oxygen	Died on 17/7/2016
4	F	41	11/5/2016	High fever, severe headache, cough, swollen airways, abnormal heart rhythm, back pain, diarrhea, vomiting, difficulty in swallowing, swollen face, hearing loss, pericarditis, abdominal pain, hemorrhaging, hypertensive	Intravenous fluids, ribavirin, maintenance of blood pressure and oxygen	Died on 23/7/2016
5	M	27	7/7/2016	Fever, headache, cough, swollen airways, abnormal heart rhythm, meningitis, back pain, diarrhea, vomiting, difficulty in swallowing, swollen face, hearing loss, encephalitis, abdominal pain, hemorrhaging, hypertensive	Balancing electrolytes, ribavirin, maintaining blood pressure and oxygen	Died on 10/8/2016
6	M	33	21/6/2016	Hyperthemia, severe headache, cough, swollen airways, abnormal heart rhythm, meningitis, back pain, diarrhea, vomiting, difficulty in swallowing, swollen face, hearing loss, encephalitis, abdominal pain, hemorrhaging, hypertensive	IV fluids, ribavirin, maintaining blood pressure and oxygen	Died on 28/7/2016
7	M	29	12/9/2016	Hyperthemia, severe headache, cough, swollen airways, abnormal heart rhythm, meningitis, back pain, diarrhea, vomiting, difficulty in swallowing, swollen face, hearing loss, tremors, seizures, abdominal pain, hemorrhaging, hypertensive	Rehydration, oxygen and blood pressure maintenance, ribavirin, dialysis	Died on 2/10/2016
8	M	73	14/8/2016	Hyperthemia, severe headache, cough, swollen airways, abnormal heart rhythm, back pain, diarrhea, vomiting, difficulty in swallowing, swollen face, hearing loss, abdominal pain, hemorrhaging, hypertensive	IV fluids, ribavirin, oxygen and blood pressure maintenance	Survived, discharged on 26/10/2016
9	F	44	14/8/2016	Fever, severe headache, cough, swollen airways, abnormal heart rhythm, body pain, diarrhea, vomiting, difficulty in swallowing, stomach ache, hemorrhaging, hypertensive	Balancing body salts, ribavirin, oxygen and blood pressure maintenance	Survived, discharged on 19/10/2016
10	F	37	16/8/2016	Fever, continuous headache, swollen airways, cough, abnormal heart rhythm, body pain, diarrhea, vomiting, difficulty in swallowing, stomach ache, hypertensive	IV fluids, ribavirin, oxygen and blood pressure maintenance	Survived, discharged on 23/10/2016
	M	39	2/2/2010	Fever, severe headache, cough, swollen airways, abnormal heart rhythm, body pain, diarrhea, vomiting,	Rehydration, ribavirin oxygen and blood pressure maintenance , Dialysis	Died on 7/3/2010

Sheet1 Sheet2 Sheet3

Ready

Figure 3.1: Lassa fever raw dataset presentation



EBOLA DATA - Microsoft Excel

F17 Hyperthermia, headache, body pain, fatigue, diarrhea, vomiting, stomach pain

CASES OF PATIENTS WITH EBOLA VIRUS DISEASE Data obtained from Ebola Emergency Operations Center (EOC), LASUTH

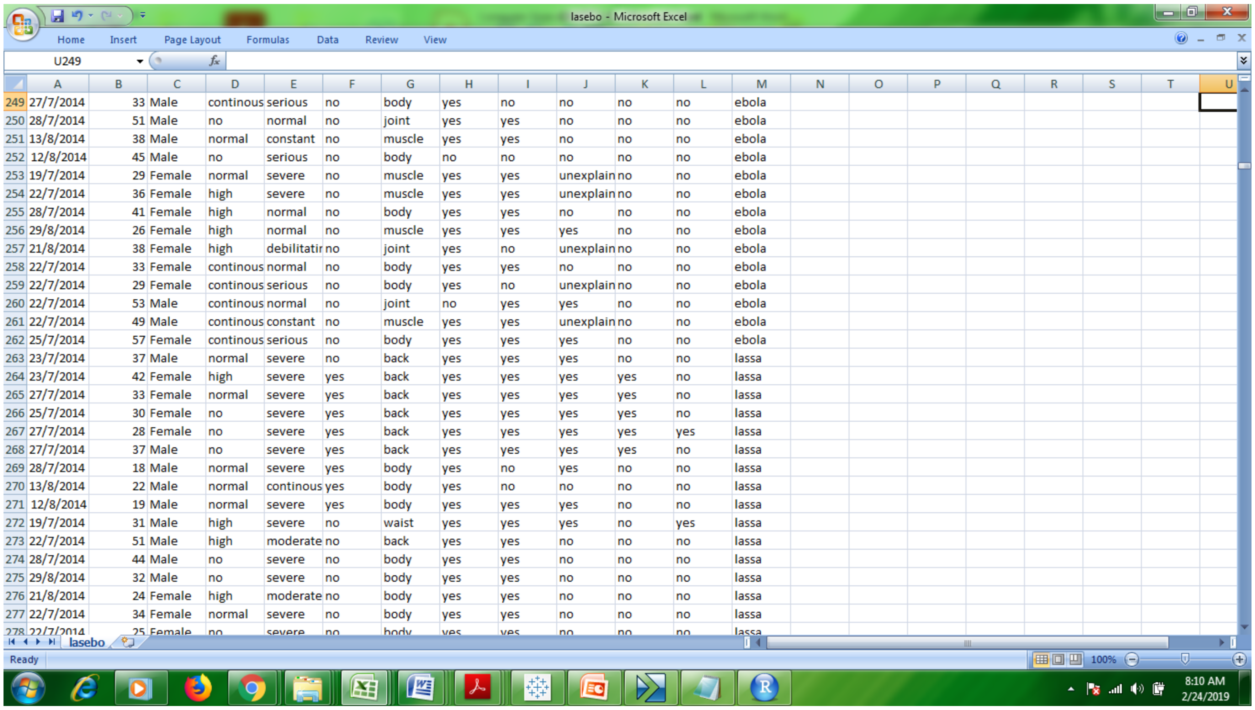
SN	NAME	GENDER	AGE	DATE	CLINICAL HISTORY	MANAGEMENT	PROGNOSIS
1	P.S	M	53	19/7/2014	Fever, severe headache, muscle pain, weakness, fatigue, diarrhea, vomiting, abdominal pain, unexplained hemorrhage	Intravenous fluids, treatment of infections	Died on 24/7/2014
2	A.A	F	49	22/7/2014	Headache, Fever, body pain, general weakness, diarrhea, vomiting, stomach pain, bleeding	Intravenous fluids, treatment of infections	Died on 19/8/2014
3	S.J	M	57	28/7/2014	Febrile, headache, fatigue, muscle pain, diarrhea, vomiting, abdominal pain, hemorrhage	Balancing electrolytes, maintaining blood pressure	Died on 29/8/2014
4	S.U	M	37	29/8/2014	High fever, debilitating headache, joint pain, weakness, tiredness, diarrhea, vomiting, abdominal ache, unexplained hemorrhage and bruising	IV fluids, treatment of infections	Died on 8/10/2014
5	T.E	M	42	21/8/2014	Hyperthermia, headache, body pain, fatigue, diarrhea, vomiting, stomach pain, unexplained bleeding	Rehydration/ oxygen maintenance, dialysis	Died on 13/9/2014
6	S.K	F	33	22/7/2014	Continuous fever, serious headache, body pain and weakness, diarrhea, vomiting, abdominal pain, unexplained hemorrhage	IV fluids, treatment of infections	Survived Ebola, Discharged on 25/9/2014
7	I.C	F	30	22/7/2014	Hyperthermia, headache, joints pain, weakness of the body, abdominal ache, vomiting, hemorrhage	Balancing body salts, treatment of infections	Survived Ebola, Discharged on 26/9/2014
8	P.C	F	28	22/7/2014	Fever, constant headache, muscle pain, weakness, fatigue, diarrhea, vomiting, abdominal pain, unexplained hemorrhage	IV fluids, treatment of infections	Survived Ebola, Discharged on 26/9/2014
9	J.T	M	37	22/7/2014	Elevated temperature, fatigue, serious headache, body ache, weakness, diarrhea, abdominal pain, vomiting, hemorrhage	Rehydration/ oxygen maintenance, dialysis	Died on 25/8/2014
10	R.A	M	18	25/7/2014	Fever, severe headache, muscle pain, weakness, fatigue, diarrhea, vomiting, stomach pain, bruising	Rehydration, treatment of infections	Survived, Discharged on 21/9/2014
11	A.P	M	22	23/7/2014	Fever, severe headache, muscle pain, weakness, fatigue, diarrhea, vomiting, abdominal pain, unexplained hemorrhage	Intravenous fluids, treatment of infections	Survived, Discharged on 23/8/2014
12	N.K	F	19	23/7/2014	Headache, Fever, body pain, generalized	Intravenous fluids, treatment	Survived, Discharged on

Figure 3.2: Ebola raw dataset presentation

3.1.2 Data Preprocessing

The dataset obtained was unstructured and contains column that are not necessary for this work such as name, patient ID, management and prognosis.

The inclusion of these column will render our model useless and unreliable, therefore there is need for preprocessing a situation where the data was section into respective and useful column most importantly clinical history column that contains all the symptoms to be used in this work, took several weeks to partition same symptoms into column in order to maintain meaningful dataset. Figure 2.3 illustrates the processed dataset.

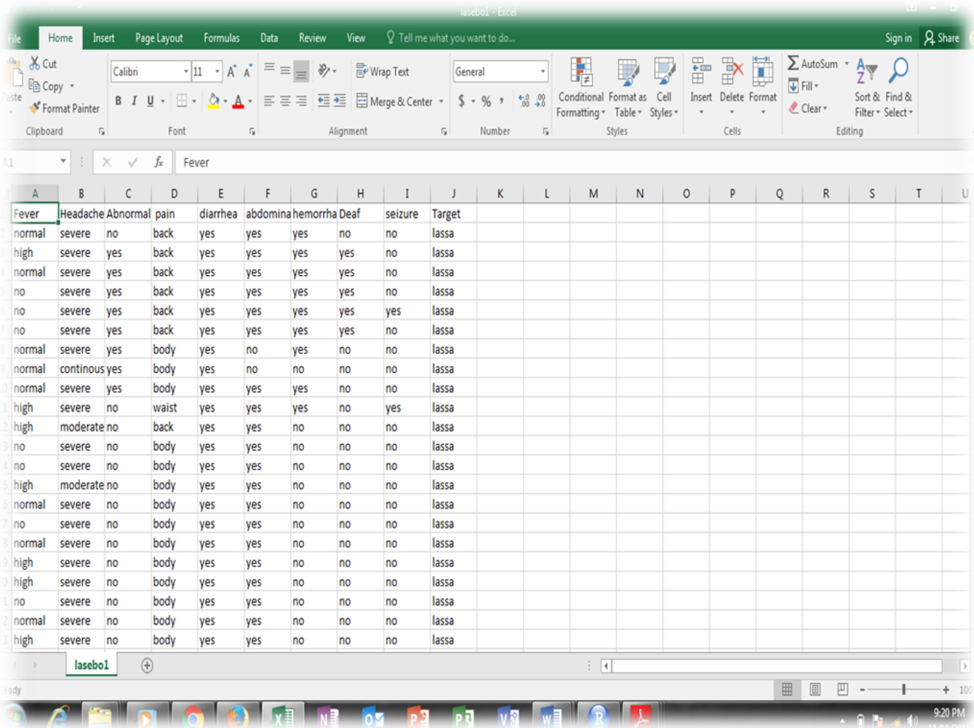


	A	B	C	D	E	F	G	H	I	J	K	L	M
249	27/7/2014	33	Male	continous	serious	no	body	yes	no	no	no	no	ebola
250	28/7/2014	51	Male	no	normal	no	joint	yes	yes	no	no	no	ebola
251	13/8/2014	38	Male	normal	constant	no	muscle	yes	yes	no	no	no	ebola
252	12/8/2014	45	Male	no	serious	no	body	no	no	no	no	no	ebola
253	19/7/2014	29	Female	normal	severe	no	muscle	yes	yes	unexplain	no	no	ebola
254	22/7/2014	36	Female	high	severe	no	muscle	yes	yes	unexplain	no	no	ebola
255	28/7/2014	41	Female	high	normal	no	body	yes	yes	no	no	no	ebola
256	29/8/2014	26	Female	high	normal	no	muscle	yes	yes	yes	no	no	ebola
257	21/8/2014	38	Female	high	debilitatir	no	joint	yes	no	unexplain	no	no	ebola
258	22/7/2014	33	Female	continous	normal	no	body	yes	yes	no	no	no	ebola
259	22/7/2014	29	Female	continous	serious	no	body	yes	no	unexplain	no	no	ebola
260	22/7/2014	53	Male	continous	normal	no	joint	no	yes	yes	no	no	ebola
261	22/7/2014	49	Male	continous	constant	no	muscle	yes	yes	unexplain	no	no	ebola
262	25/7/2014	57	Female	continous	serious	no	body	yes	yes	yes	no	no	ebola
263	23/7/2014	37	Male	normal	severe	no	back	yes	yes	yes	no	no	lassa
264	23/7/2014	42	Female	high	severe	yes	back	yes	yes	yes	yes	no	lassa
265	27/7/2014	33	Female	normal	severe	yes	back	yes	yes	yes	yes	no	lassa
266	25/7/2014	30	Female	no	severe	yes	back	yes	yes	yes	yes	no	lassa
267	27/7/2014	28	Female	no	severe	yes	back	yes	yes	yes	yes	yes	lassa
268	27/7/2014	37	Male	no	severe	yes	back	yes	yes	yes	yes	no	lassa
269	28/7/2014	18	Male	normal	severe	yes	body	yes	no	yes	no	no	lassa
270	13/8/2014	22	Male	normal	continous	yes	body	yes	no	no	no	no	lassa
271	12/8/2014	19	Male	normal	severe	yes	body	yes	yes	yes	no	no	lassa
272	19/7/2014	31	Male	high	severe	no	waist	yes	yes	yes	no	yes	lassa
273	22/7/2014	51	Male	high	moderate	no	back	yes	yes	no	no	no	lassa
274	28/7/2014	44	Male	no	severe	no	body	yes	yes	no	no	no	lassa
275	29/8/2014	32	Male	no	severe	no	body	yes	yes	no	no	no	lassa
276	21/8/2014	24	Female	high	moderate	no	body	yes	yes	no	no	no	lassa
277	22/7/2014	34	Female	normal	severe	no	body	yes	yes	no	no	no	lassa
278	22/7/2014	25	Female	no	severe	no	body	yes	yes	no	no	no	lassa

Figure 3.3: Preprocessed dataset of both (Ebola and Lassa fever data)

3.1.2.1 Data Presentation

This present the processed and labeled dataset used for training and validation of this algorithm, this dataset consists of 2000 observations with 1000 representing Ebola and Lassa fever each. The data was randomized such that the splitting into training and validating set will be random, thereafter it was partitioned into 0.70 and 0.30 for training and test set respectively.



Fever	Headache	Abnormal	pain	diarrhea	abdomina	hemorrha	Deaf	seizure	Target
normal	severe	no	back	yes	yes	yes	no	no	lassa
high	severe	yes	back	yes	yes	yes	yes	no	lassa
normal	severe	yes	back	yes	yes	yes	yes	no	lassa
no	severe	yes	back	yes	yes	yes	yes	no	lassa
no	severe	yes	back	yes	yes	yes	yes	yes	lassa
no	severe	yes	back	yes	yes	yes	yes	no	lassa
normal	severe	yes	body	yes	no	yes	no	no	lassa
normal	continous	yes	body	yes	no	no	no	no	lassa
normal	severe	yes	body	yes	yes	yes	no	no	lassa
high	severe	no	waist	yes	yes	yes	no	yes	lassa
high	moderate	no	back	yes	yes	no	no	no	lassa
no	severe	no	body	yes	yes	no	no	no	lassa
no	severe	no	body	yes	yes	no	no	no	lassa
high	moderate	no	body	yes	yes	no	no	no	lassa
normal	severe	no	body	yes	yes	no	no	no	lassa
no	severe	no	body	yes	yes	no	no	no	lassa
normal	severe	no	body	yes	yes	no	no	no	lassa
high	severe	no	body	yes	yes	no	no	no	lassa
high	severe	no	body	yes	yes	no	no	no	lassa
no	severe	no	body	yes	yes	no	no	no	lassa
normal	severe	no	body	yes	yes	no	no	no	lassa
high	severe	no	body	yes	yes	no	no	no	lassa

Figure 3.4: Dataset used in training set in the algorithm

3.1.3 Model Validation

This section presents the processes of validating the designed model.

3.1.3.1 Conceptual Framework and System Design

This shows the diagrammatic representation of the steps involve in the implementation of Decision Support System in this study.

Figure 3.3 displayed the development of the system with the processes involved and execution with both algorithms

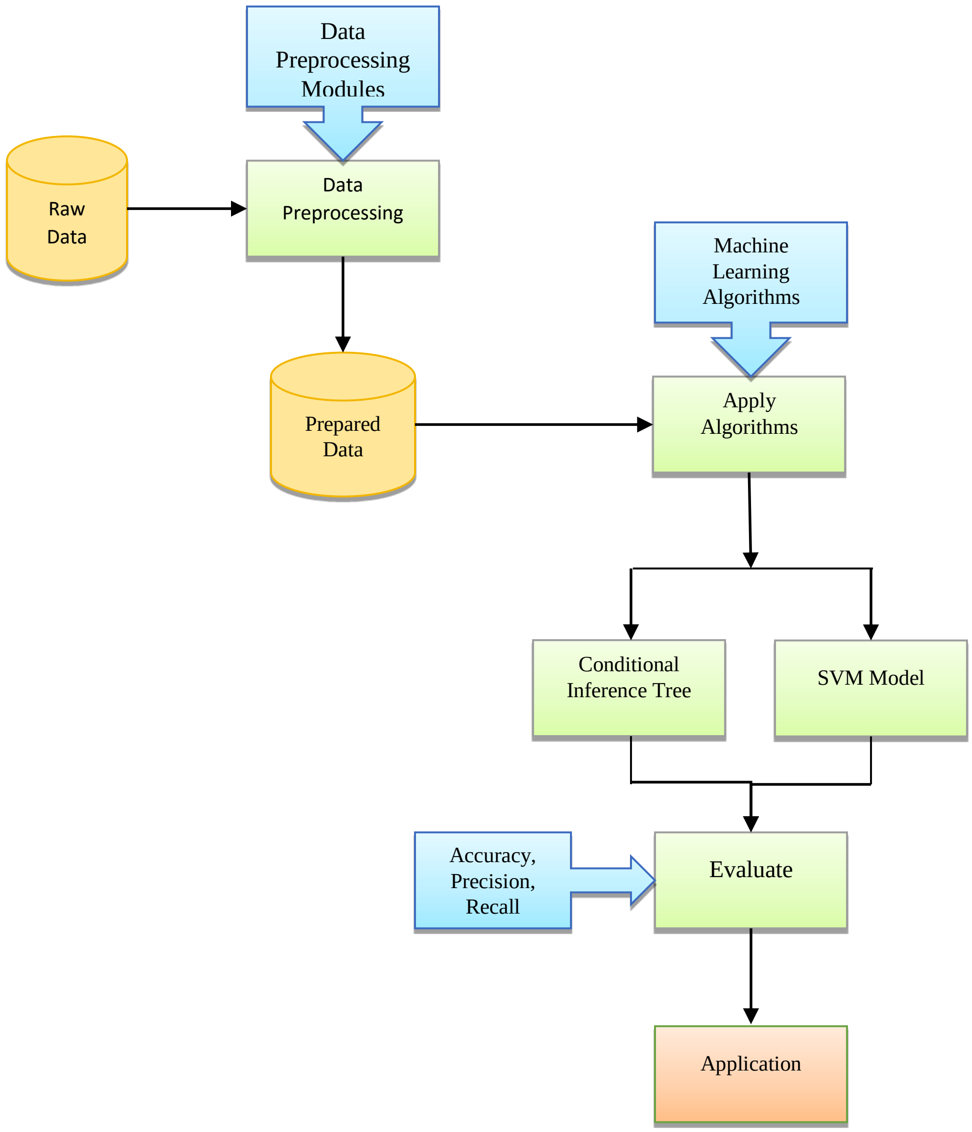
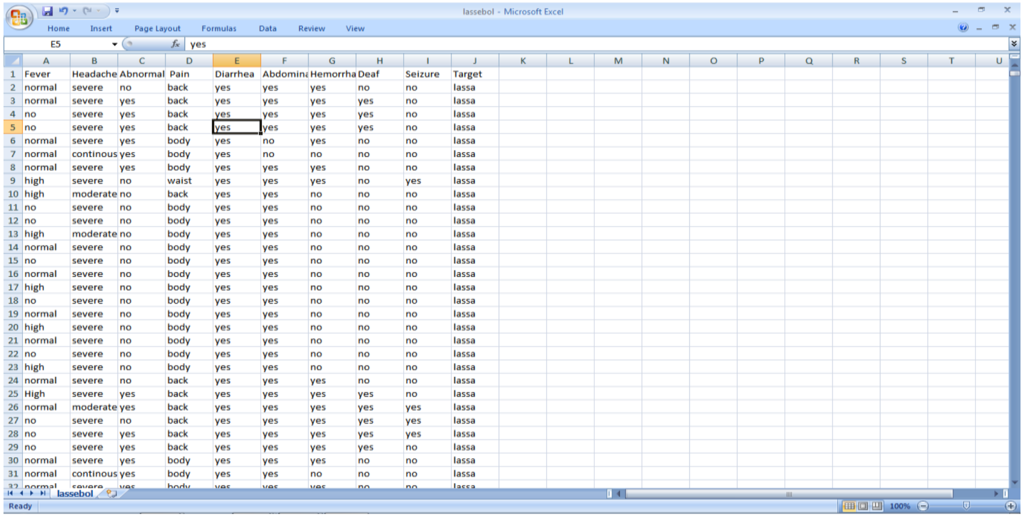


Figure 3.5: Architecture of LASSEBOL Decision Support System

Dataset: this is collection of data. In the case of this research, these are data of Lassa fever and Ebola patients. It is a table with every column representing a patient data and row corresponds to each patient. The total number of datasets on Ebola and Lassa fever are is 2000. Each disease has1000 records each. Both data sets have similar structure. A generalized table structure is described below:



	A	B	C	D	E	F	G	H	I	J
1	Fever	Headache	Abnormal	Pain	Diarrhea	Abdomina	Hemorrha	Deaf	Seizure	Target
2	normal	severe	no	back	yes	yes	yes	no	no	lassa
3	normal	severe	yes	back	yes	yes	yes	yes	no	lassa
4	no	severe	yes	back	yes	yes	yes	yes	no	lassa
5	no	severe	yes	back	yes	yes	yes	yes	no	lassa
6	normal	severe	yes	body	yes	no	yes	no	no	lassa
7	normal	continous	yes	body	yes	no	no	no	no	lassa
8	normal	severe	yes	body	yes	yes	yes	no	no	lassa
9	high	severe	no	waist	yes	yes	yes	no	yes	lassa
10	high	moderate	no	back	yes	yes	no	no	no	lassa
11	no	severe	no	body	yes	yes	no	no	no	lassa
12	no	severe	no	body	yes	yes	no	no	no	lassa
13	high	moderate	no	body	yes	yes	no	no	no	lassa
14	normal	severe	no	body	yes	yes	no	no	no	lassa
15	no	severe	no	body	yes	yes	no	no	no	lassa
16	normal	severe	no	body	yes	yes	no	no	no	lassa
17	high	severe	no	body	yes	yes	no	no	no	lassa
18	no	severe	no	body	yes	yes	no	no	no	lassa
19	normal	severe	no	body	yes	yes	no	no	no	lassa
20	high	severe	no	body	yes	yes	no	no	no	lassa
21	normal	severe	no	body	yes	yes	no	no	no	lassa
22	no	severe	no	body	yes	yes	no	no	no	lassa
23	high	severe	no	body	yes	yes	no	no	no	lassa
24	normal	severe	no	back	yes	yes	yes	no	no	lassa
25	High	severe	yes	back	yes	yes	yes	yes	no	lassa
26	normal	moderate	yes	back	yes	yes	yes	yes	yes	lassa
27	no	severe	no	back	yes	yes	yes	yes	yes	lassa
28	no	severe	yes	back	yes	yes	yes	yes	yes	lassa
29	no	severe	yes	back	yes	yes	yes	yes	no	lassa
30	normal	severe	yes	body	yes	yes	yes	no	no	lassa
31	normal	continous	yes	body	yes	yes	no	no	no	lassa

Figure 3.6: Screenshot Generalized table of the disease's datasets

a. **Fever**: It is a temporary increase in body temperature; it shows that things are not right in the body system

b. **Headache**: It is pain experience in any region of the head

c. Abnormal heart rhythm: This is when the heart beat too slow, irregular or fast

d. **Pain**: It is an unpleasant sensory and emotional experience associated with actual or potential tissue damage.

e. **Diarrhea**: A situation in which feaces are discharged from the bowels frequently and in liquid form.

f. **Abdominal pain**: It is a pain that occurs between the chest and pelvic region.

g. **Hemorrhage**: It is an escape of blood from a ruptured vessel

h. **Deaf**: The inability to hear

i. **Seizure**: It is an uncontrolled, sudden, electrical disturbance in the brain.

i. Training Dataset: is a part of data set used in training the model. Majority of the dataset are used in training the model. The total number of datasets on for training Ebola and Lassa fever are is 2000. Each disease has 1000 records each for training the system.

ii. Classification (building Model): this is a process of organizing the data sets into categories on the basis of training the dataset

iii. Testing Dataset: This is a part of the dataset taken aside for testing our model. After our model has been trained, we test our model by making prediction against the test set.

iv. Trained Model (Knowledge Base): After our model has been trained and tested, our trained model serves as our knowledge base. The knowledge base can also be used to test similar data relating to the dataset used in training the model.

v. Inference: this is the process in which the system reasons using logic rules to deduce information from the knowledge base. The process of inference will be carried out using the decision tree algorithm as a set of logic rules to inference deduce information in the knowledgebase.

vi. User Interface: this act as an intermediary between the user and the inference engine. It helps user communicate with the system effectively and it is also a means of feedback to user.

3.1.3.2 Use Cases and activity

Figure 3.7 and 3.8 illustrates the use cases of the developed system.

The use cases comprise of the following features:

1. Doctor: he/she serves as an intermediary between the system and the patient.
2. Patient Data: The most important feature in the development of the system is patient data, it is the symptoms shown by each patient it will be enter into the system for processing and classification.
3. Prognosis: Prediction is made using the inputted patient data to decide which of the disease the patient is carrying.
4. Treatment: This is the output data of the system. Which can be prescribed by the doctor to the patients.
5. Primary Health care: This is a place where the infected persons are being treated.

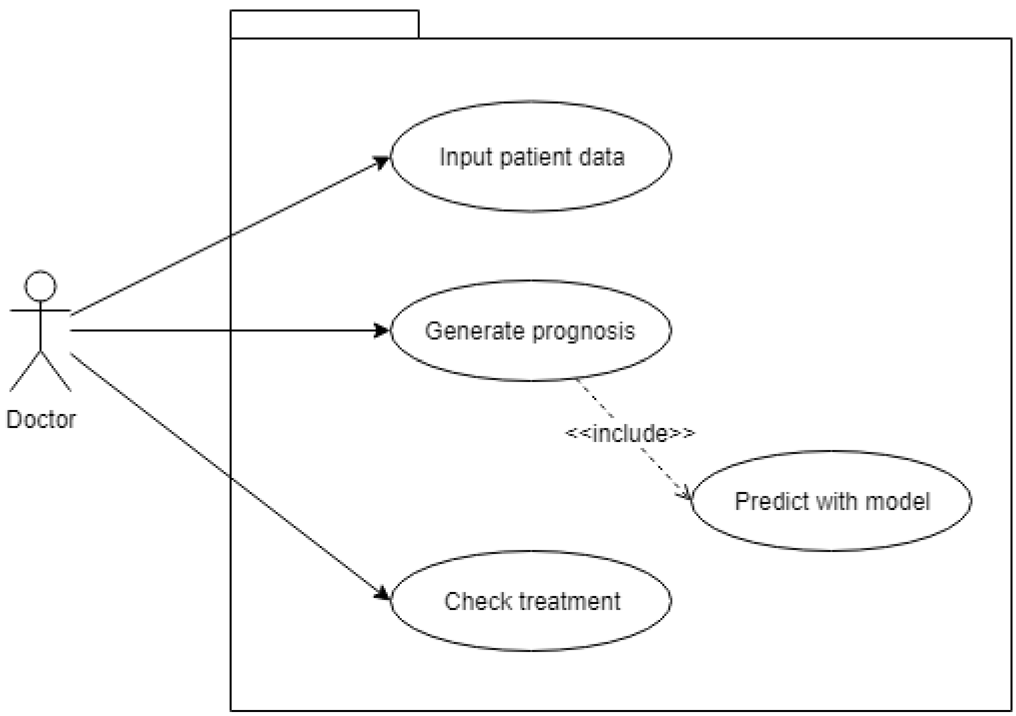


Figure 3.7: Use case for the LASSEBOL decision support system

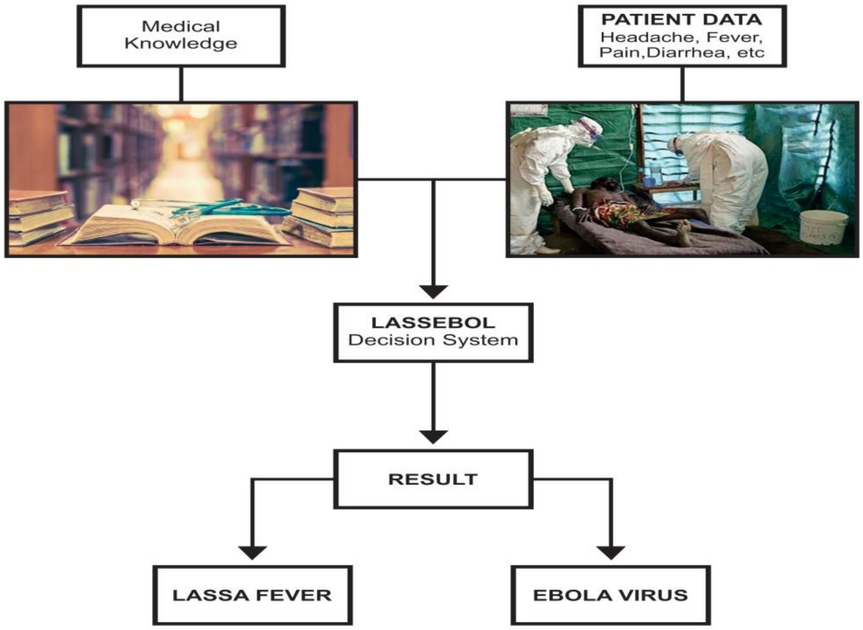


Figure 3.8: Activity diagram for the LASSEBOL decision support system

3.1.4 Data Partitioning

The data collected was moderately classified into the training and validation or testing set, a k-fold 10 was used. A probabilistic classification was applied to understand the level of a patients being selected in a class, the best class possess the highest probability.

3.1.5 MODEL DEVELOPMENT

3.1.5.1 Conditional Inference Tree

The conditional inference tree algorithm

i. Test global null hypothesis H_0 of independence between Y and all X_j with

$H_0: \cap_{j-1}^{m} \ H_0^j \ and \ H_0 : D(Y|X_j) = D(Y)$

$If \ H_0 \ not \ rejected \ \Rightarrow stop$

ii. Select variable X_j with strong association

iii. Search best split point for X_{j*} and partition data

Repeat step i, ii and iii for new split.

Hypothesis test of Independence

- Parametric tests depend on distribution assumption
- Problem: Unknown conditional distribution

$$D(Y|X) = D(Y|X_1, \dots, X_m) = D(Y|f(X_1), \dots, f(X_m)) \tag{3.1}$$

$$\mu_A \neq \mu_B \tag{3.2}$$

Test Statistic

$$T_0 = \bar{\mu}_A - \bar{\mu}_B \tag{3.3}$$

$$H_0: \mu_A - \mu_B = 0 \tag{3.4}$$

$$H_1: \mu_A - \mu_B \neq 0 \tag{3.5}$$

$\bar{\mu}_A$ = mean distribution of feature A

$\bar{\mu}_B$ = mean distribution of feature B

H_0: Null hypothesis such that difference between $\bar{\mu}_A$ and $\bar{\mu}_B$ is zero.

H_1: Alternative hypothesis such that difference between $\bar{\mu}_A$ and $\bar{\mu}_B$ is not zero.

P value and Decision

K= permutation samples:|

$|\bar{\mu}_A perm - \bar{\mu}_B perm| > |\mu_A - \mu_B|$

$$p\ value = \frac{k}{perm} \quad (3.6)$$

Provided the $p\ value < \alpha = 0.05$, H_0 can be rejected but accepted otherwise.

3.1.5.1.1 Model Formulation

$$formula = (target + fever + headache + pain + diarrhea + deaf + abnormal\ heart + abdominal\ pain + hemorrhage + seizure)$$

$$model = ctree(formula, data = train)$$

3.1.5.2 Support Vector Machines

SVM is another supervised machine learning model that can classify features based on the pattern it recognized from the training dataset. It can also be said to be a hyper-plane that divide the training set by a maximal margin. SVM is based on the idea of locating the best hyper-plane that best divides a dataset into two classes, the support vectors are the points very close to the hyper-plane. The dataset was tested on different kernel and their accuracy and other parameter was used to select the best kernel, this does not neglect the importance of other kernel that were not selected because these kernels depends on the structure of the data (Vapnik, 1990).

a. **Model Formulation**

$$formula = (target + fever + headache + pain + diarrhea + deaf + abnormal\ heart\ rhythm + abdominal\ pain + hemorrhage + seizure)$$

$$Model = svm\ (formula, data = train, kernel = "Kernel\ type")$$

b. Linear

Linear classifier relies on dot product between vectors, define as

$$k(u, v) = u.v = u^T v \tag{3.7}$$

$$where\ u = fever + headache + pain + diarrhe + deaf + abnormal\ heart\ rhythm + abdominal\ pain + hemorrhage + seizure\ and\ v = target.$$

Provided the data point is mapped to high dimensional space via some transformation the dot product $\Phi: x \rightarrow \emptyset(x)$ the dot product then becomes:

$$K(u, v) = \emptyset(u)^T \emptyset(v) \tag{3.8}$$

Where

K stands for kernel that analyses the pattern in a dataset

u the input variables in the dataset

v the outcome or class in the dataset

It has no parameter.

Radial Basis Function (RBF) also known as Gaussian kernel because it uses Gaussian equation in computation

Mathematical formula:

$$k(x, x^i) = exp\left(\frac{\|x - x^i\|^2}{2\sigma^2}\right) \tag{3.9}$$

$$k(x, x^i) = \exp\{-\gamma |x - x^i|^2\} \tag{3.10}$$

$\| x - x^{'} \|^2$ Can represent the squared Euclidean distance between the two feature vectors. σ

Is a free parameter, an equivalent definition involves a parameter. $\gamma = \frac{1}{2\sigma^2}$

$$exp\left(-\frac{1}{2} \| x - x^i \|^2\right) = \sum_{j=0}^{\infty} \frac{(x^T x^i)^j}{j!} exp\left(-\frac{1}{2} \| x \|^2\right) exp\left(-\frac{1}{2} \| x^i \|^2\right) \tag{3.11}$$

$$= \sum_{j=0}^{\infty} \Sigma^{\infty} exp\left(-\frac{1}{2} \| x \|^2\right) \frac{x^{n_1}{}_1}{\sqrt{n_1!\ldots n_k!}} exp\left(-\frac{1}{2} \| x^i \|^2\right) \frac{x_1^i n_1 \ldots x_k^i n_k}{\sqrt{n_1!\ldots n_k!}} \tag{3.12}$$

Note: γ Is a parameter that sets the spread of the kernel

$$exp\{-\gamma |x - x^i|^2\} \tag{3.13}$$

3 Polynomial

For polynomial with degree (d)

$$k(u, v) = (u^T v + c)^d \tag{3.14}$$

The vectors in the input space are u and v, which is train or test set computed from vectors of features with $c \geq 0$ as a free parameter balancing the effect of higher-order versus lower-order terms in the polynomial. When $c = 0$, the kernel is referred to as homogenous.

In this kernel, an inner product in a feature based on some mapping φ relate with k (Andrew, 2015).

$$k(u, v) = (\varphi(u), \varphi(v)) \tag{3.15}$$

$$\gamma(u^T v + \ C_0)^d \tag{3.16}$$

Where,

$$u = fever + headache + pain + diarrhea + deaf + abnormal\ heart\ rhythm + abdominal\ pain + hemorrhage + seizure\ and\ v = target$$

Sigmoid: It is a mathematical function with a sigmoid curve “S” shaped,

$$S_u = \frac{1}{1+e^{-x}} = \frac{e^x}{e^x+1} \quad (3.15)$$

$\tanh(\gamma u^T v + C_o)$

Where,

$u = fever + headache + pain + diarrhea + deaf + abnormal\ heart\ rhythm + abdominal\ pain + hemorrhage + seizure\ and\ v = target$

It is a real function defined as a non-negative derivative at each point and for real input values.

Parameter Definition:

i. γ (Gamma): It determine the bias and variance in the model, such that a small gamma means distribution with a large variance so the effect of the support vector is more,. If gamma is large the variance is small meaning the support vector does not have wide-spread effect. The model with low variance and high bias is as a result of large gamma.

ii. c or c_o(cost): It is a parameter for a non-linear support vector machine, it manage each support vectors effect, it involves trading off the influence of one dimension to another dimension to maintain stability.

iii. (d) Degree: The power or order at which the equation is raised.

3.1.6 Model Evaluation

The performance metric to select the algorithm that performs better is the accuracy of the model, with the mathematical formula of the form;

$$\boldsymbol{ACCURACY} = \frac{TP+TN}{TP+FP+TN+FN} \quad (3.16)$$

TP**:** This is the number of correctly predicted object in the positive class

TN: This is the number of correctly predicted object in the negative class.

FP: This is number of incorrectly predicted object of negative class.

FN: This is the number of incorrectly predicted object of positive class

CHAPTER FOUR

RESULTS AND DISCUSSION

This chapter focuses on the presentation of data and the analysis carried out in this study, also the presentation of the performance metric for the selection of accurate, precise and sensitive model that best predict tendency of a disease being Lassa fever or Ebola considering the independent variables features.

4.1 Data Analysis

The algorithm used to develop to solve the classification problem is;

1. **Conditional Inference Tree**: It is a recursive binary partitioning with statistics and weights being the parameter considered for the selection of split.

In R programming environment, party packages provide the ctree function that apply conditional inference tree on the model, using the nine features as the independent variables against the categorical target variable (Ebola or Lassa).

The model predicted the test set with an accuracy of 99%. Figure 4.1 presented the processes involved in the computation of CIT algorithm.

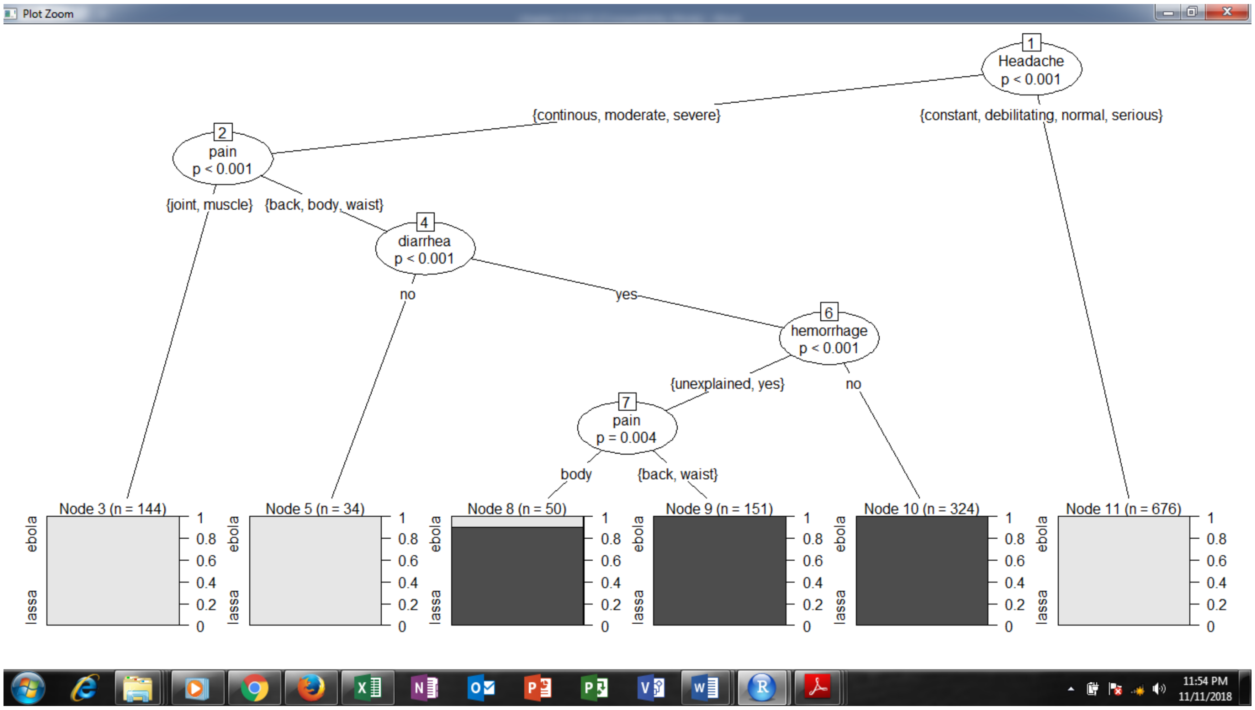


Figure 4.1 Presentation of Condition Inference Tree algorithm

2. **Support Vector Machine**: It basically looks for the optimal separating hyper-plane between two classes by maximizing the gap (margin) between the classes with the closest points, the points that lies on the boundaries of the separating classes is referred to as support vectors and the space between these classes is called optimal separation(Hothorn *et al*, 2006).

In R programming environment, e1071 packages provides the SVM function that apply SVM on the model with an argument in the function to change kernel (Radial Basis Function (RBF), Linear, Sigmoid, Polynomial), using the nine features as the independent variables against the categorical target variable (Ebola or Lassa).

RBF kernel: The kernel had an accuracy of 99% with 217 support vectors also this kernel possesses a single parameter called gamma (γ).

Sigmoid: The kernel had an accuracy of 99% with 331 support vectors, also this kernel possesses two parameters (γ,c0).

Polynomial: The kernel had an accuracy of 92% with 1014 support vectors.

4.1.1 Model Evaluation:

The linear kernel possesses the highest level of accuracy with lowest number of support vectors and without parameter, while other kernel possesses 99% accuracy with assumptions but insufficient number of support vectors. In view of these, conditional inference tree was 99% accuracy and 99% sensitivity, with a framework of splitting nodes base on the features that possesses higher statistics and weights. The caret package in R with confusion Matrix function gives allowance for the printing of Accuracy, Sensitivity, Precision, and Recall while F1-score was computed with F1 score. Table 4.5present the performance metric table.

Table 4.1 PERFORMANCE METRICS OF CIT AND SVM RESPECTIVE CONFUSION MATRIX

Model / Metrics	**Accuracy (%)**	**F1 (%)**	**TP**	**TN**	**FP**	**FN**	**Sensitivity (%)**	**Recall (%)**	**Precision (%)**
Conditional Inference Tree	0.99	1	367	240	3	0	0.99	0.98	1
Support Vector Machine									
Sigmoid	0.99	1	340	240	20	0	0.98	0.98	1
Polynomial	0.92	0.94	366	203	3	37	1	1	0.88
Radial	0.99	1	366	240	3	0	0.99	0.99	1

Table 4.5: Performance metrics of algorith

From the above table, it could be deduced that conditional inference tree perform better than other SVM’s kernel, though they all have higher accuracy of over 90% but the level of sensitivity in CIT is higher than any SVM kernel which is a very important metric since we are dealing with lives. Therefore, conditional inference tree was selected and therefore modeled in the web application with Azure Machine Learning integration with R.

CHAPTER FIVE

CONCLUSION AND RECOMMENDATION

5.1 CONCLUSION

Based on this study, lack of clinical decision support system in this aspect of the medical science has led to loss of lives also easy transmission of these viruses. Therefore, the LASSEBOL system leverage on health professional's knowledge of deciphering symptoms posed by patient as stored information and the system relate with the inputted data to classify these diseases and immediately will be follow up by assigning necessary treatment to the patients. Unlike the traditional method of sampling blood which most patients died before the arrival of result, LASSEBOL will enhance and facilitate decision making in this respect and in turn save lives because of its model sensitivity and accuracy attained.

5.2 RECOMMENDATIONS

For further research work in this field, I hereby suggest the following:

i. Future research should apply the clinical decision system to life-threatening disease in Africa such as Malaria

ii. Future work is expected to be very careful in handling the input variables so as to overcome problem of multi-collinearity among these variables.

REFERENCES

1) Ayten Kadanali and GulKaragoz (2015), 'An Overview of Ebola Virus Disease'. Northern Clinics of Instanbul, Vol. 2(1), pp: 81–86, https://10.14744/nci.2015.97269.
2) Olugbenga Oluwagbemi, Folakemi Oluwagbemi, Oluyemi Abimbola.(2016), 'Ebinformatics: Ebola fuzzy informatics systems on the diagnosis, prediction and recommendation of appropriate treatment for Ebola virus disease (EVD)'. Vol. (2), pp: 12-37. Elsevier
3) DimieOgoina, 'A Clinical and Epidemiological Review' (2013). Niger Dalta Journal of Medicine and Medical Research. Vol. (1), pp: 1-10. Niger Delta University
4) Vanessa Raabe, Jeffrey Koehler. (2017)'Laboratory Diagnosis of Lassa Fever'. Journal of Clinical Microbiology Volume 55(6), pp: 1629-1637, https://10.1128/JCM.00170-17.
5) Shaker H. El-sappagh, Samir El-masri.(2014). 'A distributed clinical decision support system architecture'. Journal of King Saud University- Computer and Information Sciences. Volume 26(1). Pp: 69-78. https://doi.org/10.1016/j.jksuci.2013.03.005.
6) Alexis Sardá-Espinosa, SubanatarajanSubbiaha, Thomas BartzBeielstein Conditional Inference, 2017 "Trees for the Knowledge Extraction from Motor Health Condition Data". Volume (62) Pages 26-37.https://doi.org/10.1016/j.engappai.03.008.
7) Chandrakant Ruparelia, Melanie Curless, Polly Trexler, and Meredith Black (2015)." Prevention and Control of Ebola Virus Disease in Health Care Facilities with Limited Resources". John Hopkins Medicine.JHpiego Corporation

8) Alia Alabdulkarim, Mznah Al-Rodhaan, Tinguai Ma, Yuan Tian. (2019). ‘ A Novel Privacy-Policy Single Decision Tree Algorithm for Clinical Decision Support Systems Using IoT Devices. Journal Sensors, Vol 19(1), pp: 1-20. https://doi.org/10.3390/s/9010142.

9) Roosan Islam, Charlene R. Weir, Wakoto Jones, Guilherme Del Fiol, Matthew H. Samore (2015).”Understanding Complex Clinical Reasoning in Infectious diseases for Improving Clinical Decision Support Design”. Medical Informatics and Decision Making. Vol. 15(100), pp: 1-12. https://doi.org/10.1186/s12911-015-0221-z.

10) Hothorn, T., Hornik, K., Zeileis, A., 2006. Unbiased recursive partitioning: A conditional inference framework. Journal of Computational and Graphical Statistics 15, 651–674.

11) Musa Peker (2016) “A decision support system to medical diagnosis using a combination of k-medoids clustering based attribute weighting and SVM”.Journal of Medical Systems. Vol. 40 (5).pp:1-16. https://doi.org/10.1007/s10916-016-0477-6. “MuglaSıtkıKocman University”.

12) Prem Pal Singh Tomar&Ranjit Singh (2013). Design of Clinical Decision Support System for Chronic Heart Disease Diagnosis Using Case Base Reasoning. Afr J. of Comp & ICTs. Vol 6, No. 1. Pp 149-160.

13) Polat K, Gunes (2007) A hybrid approach to medical decision support systems: combining feature selection, fuzzy weighted pre-processing and AIRS. Computer Methods and Programs in Biomedicine. Vol. 88(2).Pp:164-178. https://doi.org/10.1016/j.cmpb.2007.07.013.

14) Flores (2015), Design and development of a decision support system for screening of Lynch syndrome using openEHR. Stockholm University

15) Niclas Johansson, Carl Spindler, John Valik, and Veronica Vicente (2018). "Developing a Decision Support System for Patients with Severe Infection Conditions in Pre-hospital Care". International Journal of Diseases. Vol. (72), pp: 40-48. https://10.1016/j.ijid.2018.04.4321.

16) Elalfi A., Fouda M, Atta A.A (2016). "Developing an Intelligent Decision Support System for the Diagnosis of some Children's Diseases". International Journal of Computer Application. Vol. 151(2).pp: 32-38.

17) A.H Morris (2019). "Decision Support and Safety of Clinical Environments". British Medical Journal. Vol. 11(1).pp:69-75. https://dx.doi.org/10.1136/qhc-11.1.69. Epidemiology and Community Health.

18) Mathapanee Oonsivilai, Yin Mo, NantasitLungusanatip, YuelLubell, ThylMiliyaPisey Tan, LornLoeuk, Paul Turner, Ben S. Cooper (2019)." Using Machine to Guide Targeted and Locally-Tailored Empiric Antibiotic Prescribing in a Children's Hospital in Cambodia".WellcomeOpenResearch,Vol.3(131).pp:1-18. https://doi.org/10.12688/wellcomeopenres.14847.1.

19) HelaLtifi, SabberAmri, Mounir Ben Ayed (2016)."Fuzzy logic-based evaluation of visualization generated by intelligent decision support systems". Sage Journal. Vol. 17(1). pp:3-21. https://doi.org/10.1177/1473871616674046.

20) Panagiotis Bomtis, Maria Haritou, Abraham Pouliakis, NikiMargari, Maria Kyrgiou, ArisSpathis, Asimaki Pappas, LoannisPanayiotides, Evangelos A. Paraskevaids (2014). "An intelligent Clinical Decision Support System for Patient-Specific Predictions to Improve Cervical Intra Epithelial Neoplasia Detection". Biomed Research International. Vol. (2014).pp:1-20. https://dx.doi.org/10.1155/2014/341483.

21) Ida Sim, Paul Gorman, Robert A. Greene, Bryan Haynes, Bonnie Kaplan, Harold Lehmann, Paul Tang. Journal of the American Medical Informatics Association "Clinical Decision Support System for the Practice of Evidence Based Medicine". Vo. 8(6). pp:527-534.

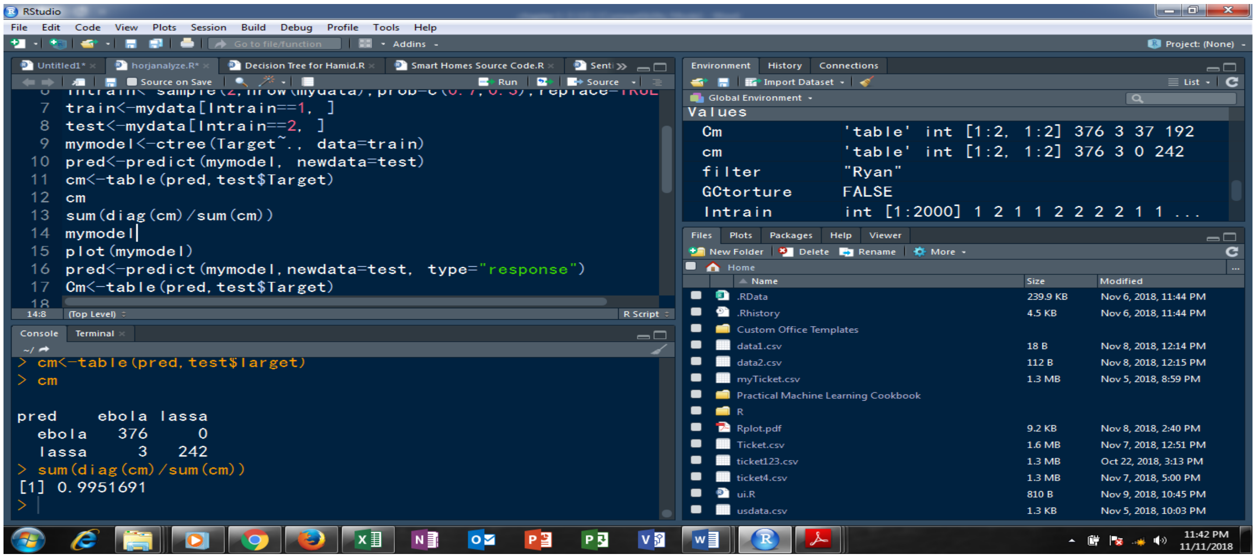
22) Nassim Douali, Elpiniki l. Papageorgiu, Jos De Roo, Hans Cool and Marrie-Christine Jaulent. "Clinical Decision Support based on fuzzy cognitive maps". Computer Science and System Biology, Vol.8 (1).pp;112-120. https://10.4172/jcsb.1000177.

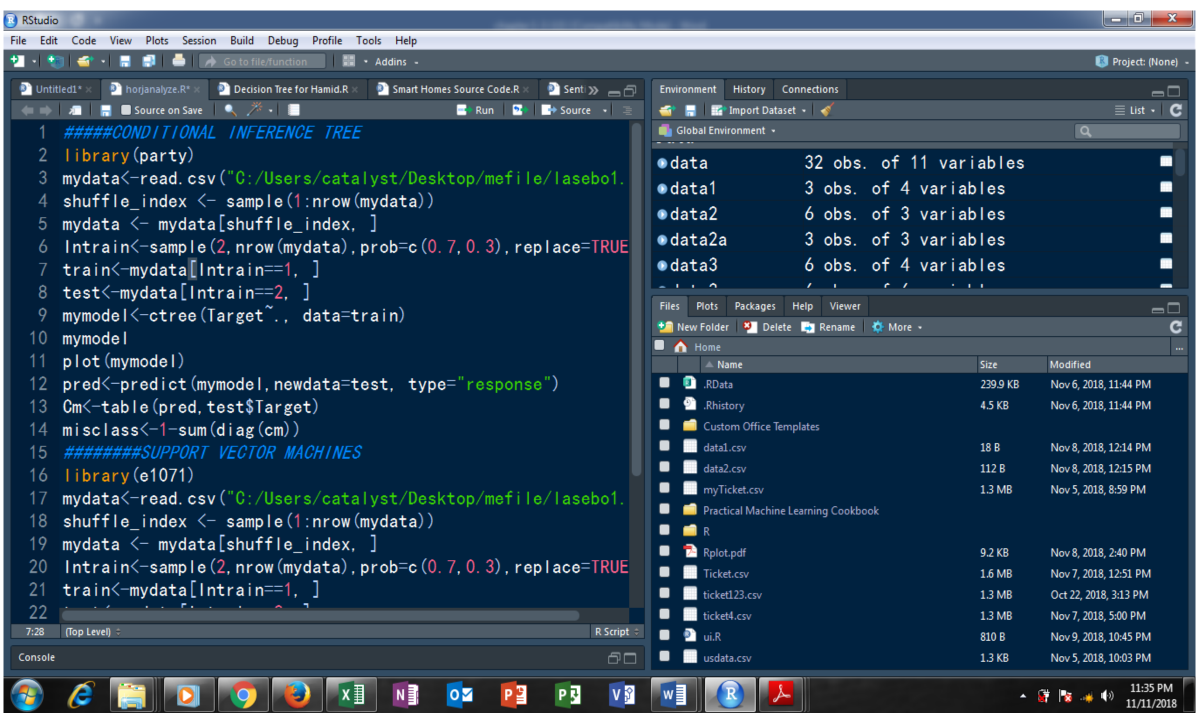
23) Igwe Sylvester Agbo, Nwele Anamechi Ogu."The Need for Introducing Decision Support System in Nigerian Universities Management and Administraton". Information and Knowledge Management. Vol. 3(10). pp:96-101

24) Resul T.I, Abdulkadir (2008), "Effective diagnosis of heart disease through neural networks ensemble". An International Journal ELSEVIER Vol.(36). pp: 7675-7680.

Appendix A

System's Output





SOURCE CODE

```
#####CONDITIONAL INFERENCE TREE

library(party)

library(e1071)

library(caret)

mydata<-read.csv("C:/Users/PC/Documents/lasebo1.csv")

shuffle_index<- sample(1:nrow(mydata))

mydata<- mydata[shuffle_index, ]

Intrain<-sample(2,nrow(mydata),prob=c(0.7,0.3),replace=TRUE)

train<-mydata[Intrain==1, ]

test<-mydata[Intrain==2, ]

mymodel<-ctree(Target~., data=train)

pred<-predict(mymodel, newdata=train, type="response")

cm<-precision(table(pred,test$Target))

Cm<-confusionMatrix(table(pred,train$Target))

sum(diag(cm)/sum(cm))

mymodel

plot(mymodel)
```

```
pred<-predict(mymodel,newdata=test, type="response")

cm<-confusionMatrix(table(pred,test$Target))

cm<-precision(table(pred,test$Target))

Cm<-table(pred,test$Target)

misclass<-1-sum(diag(cm))

########SUPPORT VECTOR MACHINES

library(e1071)

mydata<-read.csv("C:/Users/PC/Documents/lasebo1.csv")

shuffle_index<- sample(1:nrow(mydata))

mydata<- mydata[shuffle_index, ]

Intrain<-sample(2,nrow(mydata),prob=c(0.7,0.3),replace=TRUE)

train<-mydata[Intrain==1, ]

test<-mydata[Intrain==2, ]

mymodel<-svm(Target~., data=train, kernel="radial")

mymodel

plot(mymodel)

pred<-predict(mymodel,newdata=test, type="response")

tab<-precision(table(pred, test$Target))
```

```
tab1<-recall(table(pred, test$Target))

Cm<-confusionMatrix(table(pred,test$Target))

misclass<-1-sum(diag(cm))

######### Sigmoid

mydata<-read.csv("C:/Users/PC/Documents/lasebo1.csv")

shuffle_index<- sample(1:nrow(mydata))

mydata<- mydata[shuffle_index, ]

Intrain<-sample(2,nrow(mydata),prob=c(0.7,0.3),replace=TRUE)

train<-mydata[Intrain==1, ]

test<-mydata[Intrain==2, ]

mymodel<-svm(Target~., data=train, kernel="sigmoid")

mymodel

plot(mymodel)

pred<-predict(mymodel,newdata=test, type="response")

tab<-precision(table(pred, test$Target))

tab1<-recall(table(pred, test$Target))

Cm<-confusionMatrix(table(pred,test$Target))

misclass<-1-sum(diag(cm))
```

```
########### Polynomial

mydata<-read.csv("C:/Users/PC/Documents/lasebo1.csv")

shuffle_index<- sample(1:nrow(mydata))

mydata<- mydata[shuffle_index, ]

Intrain<-sample(2,nrow(mydata),prob=c(0.7,0.3),replace=TRUE)

train<-mydata[Intrain==1, ]

test<-mydata[Intrain==2, ]

mymodel<-svm(Target~., data=train, kernel="polynomial")

mymodel

plot(mymodel)

pred<-predict(mymodel,newdata=test, type="response")

tab<-precision(table(pred, test$Target))

tab1<-recall(table(pred, test$Target))

Cm<-confusionMatrix(table(pred,test$Target))

misclass<-1-sum(diag(cm))
```