DEVELOPMENT OF A BAYESIAN BASED APPROACH TO MALARIA FEVER DIAGNOSIS

BY

YAYA IBRAHIM ADEKUNLE
(CSC/11/0290)

BEING A PROJECT REPORT SUBMITTED

TO THE

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

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

OCTOBER, 2015.

CERTIFICATION

This is to certify that the Project Report entitled "**Development of A Bayesian Based Approach to Malaria Fever Diagnosis**" was carried out **by YAYA IBRAHIM ADEKUNLE** with matric no. **CSC/11/0290**, in partial fulfilment of the requirements for the award of Bachelor of Science degree in Computer Science, Faculty of Science, Federal University Oye-Ekiti, Ekiti State.

Dr. Adebayo Adetunmbi
Supervisor

22/10/15

DATE

Dr. Oluwadare S.A
Ag. Head of Department

25/10/2015

DATE

DEDICATION

This report is dedicated to the Almighty God who has been so Merciful to me in all ways without Him nothing would have been possible.

To my loving and caring parents, Mr. and Mrs. Yaya and my Siblings for their support, fortitude, patience, hope and perseverance to get me educated.

ACKNOWLEDGEMENTS

My hands stretch towards God both in appreciation and ecstasy that He stood by me all through, such that the journey was both pleasured and success.

My thanks goes to my parent **Mr. and Mrs. Yaya** who has taken the pain to give the greatest ornaments a parent could give to their children education.

I appreciate my supervisor **Dr. Adebayo Adetunmbi** for the help, guidance and for providing me with most essential materials required for the completion of this project. I'm very thankful to him for his indomitable guidance. This inspiration up to the last moment had made things possible in a nice manner.

I also appreciate Mr. Folorunsho and all my lecturers, for the knowledge, wisdom and understanding they impacted on me.

I would like to use this medium to acknowledge the managing director of May System Solutions in person of Mr. Yaya Muritala Adedimeji (Alhaji) and also to my siblings Yaya Adejumoke Ganiyat (Big Sis), Sister Toyin and Yaya Yusuf (Y Square) for their kind gesture to me towards my academic.

I would like to use this medium to appreciate everyone who helped to complete this project most especially Kunle-Oni Temitope Atinuke (Babyluv), Adekola Tolulope Fadekemi, Alatishe Babajide Rufai (Engr Babs) and Adebesin Tobi Tosin (shogy).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Malaria is a deadly disease killing millions of people every year. Different countries of the world, governmental and non-governmental organizations including World Health Organization have taken it as a challenge to address the issue of deaths associated with malaria. Prompt and accurate diagnosis is a major key in medical field. This prompts for the need to develop a Bayesian base approach to malaria fever diagnosis. A machine learning technique Bayesian was used on labelled sets of malaria fever symptoms collected in malaria dataset. The labelled database was divided into five cases of malaria and the classification model for malaria fever diagnosis was developed using WEKA software.

The developed model has been tested and gives a classification accuracy of 66% on training dataset while that of testing data set gives classification accuracy of 84%.

The result shows that the Bayesian is a promising approach and the system hereby recommended for use in areas where cases of malaria fever are prevalence.

CHAPTER ONE

INTRODUCTION

1.1 Background of Study

Malaria Fever has been a long life-threatening parasitic disease transmitted by female anopheles mosquitoes. This has contributed child morbidity in the world. It threatens 2.4 billion people, or about 40% of the world's population living in the world's poorest countries and more than one million deaths are attributable to the disease annually (WHO, 2010).

It is a major public health problem in Africa with over 200 million clinical episodes and nearly one million deaths occurring annually (WHO/UNICEF, 2005). In semi-arid and highland regions of Africa, Malaria Fever is unstable and epidemic Malaria Fever is a common problem causing deaths annually (Worall et al, 2004). However, the risks of morbidity and mortality associated with Malaria Fever, particularly in semi-arid and highland regions, vary spatially and temporally (Snow and Marsh, 2002). Most Malaria Fever infections, particularly in sub-Saharan Africa, are caused by Plasmodium falciparum. Malaria Fever presents a major socio-economic challenge to African countries since it is the region most affected. This challenge cannot be allowed to go unnoticed since good health is not only a basic human need but also a fundamental human right and a prerequisite for economic growth (UN, 2003).

Malaria Fever is caused by a parasite that is transmitted from one person to another through the bite of the Anopheline mosquito (a female Anopheles mosquito). Humans get Malaria Fever from the bite of the Malaria Fever-infected mosquito. When the mosquito bites an infected person, it ingests microscopic Malaria Fever parasites found in the person's blood. The Malaria Fever parasite must grow in the mosquito for a week or more before infection can be passed to another person. Thereafter, if the mosquito bites another person, the

parasites go from the mosquito's mouth into the person's blood.

They feed on the blood cells, multiply inside the liver and thereby destroying the red blood cells causing a cut off in blood circulation which could lead to premature death. (WHO, 2010). Symptoms of Malaria Fever include fever, shivering, pains in the joint, vomiting, anaemia, hemoglobinuria, retinal damage, and convulsions. The classic symptom of Malaria Fever is cyclical occurrence of sudden coldness followed by rigor then fever and sweating lasting four to six hours. This occurs every two days in plasmodium vivax (*P. vivax)* and plasmodium Ovale (*P. ovale*) infections, while every three days for plasmodium Malaria Fevere (P. *Malaria Fevere*) (Nyika, 2009).

Malaria Fever can be prevented by the use of mosquito coils and repellants, spraying the insides of houses (where most *Anopheles* species feed and rest) with insecticides (indoor residual spraying, IRS) and by sleeping under bed nets that have been treated with long-lasting insecticides (long-lasting insecticide nets, LLINs). Mass screening and treatment (MSAT) with effective antiMalaria Feverl drugs can also reduce Malaria Fever transmission (Griffin et al., 2010).

However, the levels of Malaria Fever risk and transmission intensity exhibit significant spatial and temporal variability related to variations in climate, altitude, topography, and human settlement pattern (Abeku et al., 2003). The Ministry of Health (MOH, 2009) records that between 3-3.5 million cases of Malaria Fever are reported each year, over 900,000 of which are children under five years. According to the President Malaria Initiative (PMI) is said to account for 61% of under-five hospital admissions and 8% of admissions of pregnant women.

Four species of the Malaria Fever parasites exist, Plasmodium falciparum, Plasmodium vivax, Plasmodium ovale and Plasmodium Malaria Fever. Epidemiological study shows that

only three species of the Plasmodium are present; Plasmodium falciparum, Plasmodium Malaria Fever and Plasmodium ovale .The Plasmodium falciparum is thus the predominant parasite species carried by a combination of vectors (MOH, 1991).

Climatic factors, particularly rainfall, temperature and relative humidity have a strong influence on the biology of mosquitoes. In Malaria Fever endemic countries, climate factors reportedly contribute to the increased number of mosquitoes and thus make transmission favorable. Once adult mosquitoes have emerged, the ambient temperature, humidity, and rains will determine their chances of survival. Warmer ambient temperatures shorten the duration of the extrinsic cycle, thus increasing the chances of transmission (Jackson, 2010).

1.2 Statement of the Problem

Malaria Fever continues to be an economic burden and a great threat globally and almost impossible to eradicate for the past six decades. It is a mosquito-borne disease causing 1.5 to 2.7 million people to die annually (Breman and Alilio 2004).Malaria Fever vectors have become more resistant to insecticides and the parasites that cause the disease are becoming resistant to chloroquine and other anti-malarial drugs, making prevention and treatment increasingly more difficult and costly. About 40% of the world's populations live in regions where malaria fever transmission is endemic, mainly within the tropical and sub-tropical regions. (Aultman, 2002). Malaria Fever is by far the leading cause of death in the tropical and sub-tropical regions. Twenty five (25%) of children who die before their fifth birthday are killed by the disease, and it claims the lives of many pregnant women too. (Asante and Asenso-Okyere, 2003). As a result, the development of a Bayesian based approach shall be used to address the problems which will effectively provide quality, efficient and accurate diagnosis.

1.3 Aim and Objectives

1.3.1 Aim

The main purpose of this study is to develop a Bayesian based approach to malaria fever diagnosis.

1.3.2 Objectives

The specific objectives are to:

i. Build a Bayesian classification model for malaria fever diagnosis.

ii. Implement the Bayesian classification model for malaria fever diagnosis.

1.4 Significance of Study

It's significance is seen in its ability to:

- Provide diagnostic support.
- Promote better patient care.

1.5 Scope of the Study

This study attempts to build a classification model for malaria diagnosis using Bayesian method and develop a web-based medical diagnosis system for malaria, based on classification model generated by Bayesian approach.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

This chapter focuses on explaining the topic of this research which is development of a Bayesian based approach to malaria fever diagnosis and also the related work that has been done.

2.2 Malaria Fever Situation

Malaria Fever is caused by a parasite called Plasmodium, which is transmitted through the bites of infected female anopheles mosquitoes. In the human body, the parasites multiply in the liver, and then infect red blood cells. There are four of this different species causing the human malaria Fever disease: Plasmodium falciparum, Plasmodium vivax, Plasmodium ovale and Plasmodium malaria. (WHO, 2010).

Malaria Fever is a vector-borne disease that is widespread in the tropical and subtropical areas of the world. This has become a serious challenge for most developing countries where between 300 and 500 million people are infected annually. The disease is a leading cause of infant and child mortality in sub-Saharan Africa (WHO, 2003).

There have been several efforts by government and other development partners in the health sector to eradicate malaria fever but its prevalence rate is still on the increase. This has prompted the question why the malaria fever cases are still on the increase despite these efforts.

The Parasite: Plasmodium

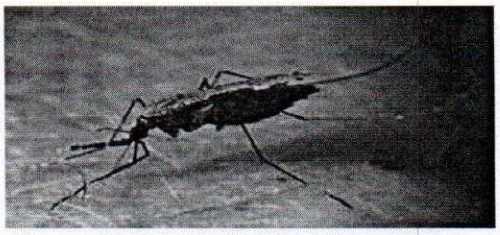

Figure 2.1: Malaria Parasite

The protozoan Plasmodium from the malaria fever parasite in Fig 2.1 is transmitted to humans by mosquitoes of the genus Anopheles. The mosquito picks up the parasite during a blood feeding from an animal with parasitaemia. Plasmodium falciparum causes a large majority of the clinical cases and mortalities. (Bozdech et al., 2003).

2.2.1 Socio-Economic Burden of Malaria Fever

It has been known that malaria fever and underdevelopment are closely linked. As a general rule of thumb, where malaria fever prospers most, human societies have prospered least (Gallup and Sachs, 2001). The disease causes widespread premature death and suffering, imposing financial hardship on poor households, and holds back economic growth and improvements in living standards. Malaria fever flourishes in situations of social and environmental crisis, weak health systems and disadvantaged communities (WHO, 2000).

Developing countries are still having a chunk of their national budget being used for malaria eradication. Studies according have established the fact that malaria fever affects mostly the poor impacting negatively on their socio-economic development. The burden of malaria fever is therefore greatest among the world's poorest countries (Worrall, 2003).

2.3 Biology of the Disease

Understanding the biological basis of the disease aids in unveiling the nature, causes and implication of the disease control and monitoring process.

2.3.1 Vector Biology

Adult females of many mosquito species will bite humans, using the blood meals for egg production. However, only about 60 species of the genus Anopheles can transmit malaria fever. Anophelines generally bite at night and usually rest on a surface (such as the wall of a house) before or after feeding. As with all mosquitoes, the immature stages are aquatic, and they prefer slow-moving or still water in which they can stay close to the water surface with their breathing orifices open to the air (Kathleen, 2002).

2.3.2 Symptoms of Malaria Fever

Symptoms of malaria fever worldwide include fever, headache, and vomiting, and usually appear between 10 and 15 days after the mosquito bite. If not treated, malaria fever can quickly become life threatening by disrupting the blood supply to vital organs (WHO, 2010). The symptoms of clinical malaria fever according to Asenso-Okyere (1994) are yellowish eyeball, chills and shivering, headache, a bitter taste, body weakness and yellowish urine.

2.3.3 Life Cycle of the Malaria Parasite

The life cycle of the malaria fever parasite shown below in Fig 2.2 is as follows:

- A female Anopheles mosquito carrying malaria fever causing parasites feeds on a human and injects the parasites in the form of sporozoites into the bloodstream. The sporozoites travel to the liver and invade liver cells.
- Over 5-16 days, the sporozoites grow, divide, and produce tens of thousands of haploid forms, called merozoites, per liver cell. Some malaria fever parasite species remain dormant for extended periods in the liver, causing relapses weeks or months later.
- The merozoites exit the liver cells and re-enter the bloodstream, beginning a cycle of invasion of red blood cells, asexual replication, and release of newly formed merozoites from the red blood cells repeatedly over 1-3 days. This multiplication can result in

thousands of parasite-infected cells in the host bloodstream, leading to illness and complications of malaria fever that can last for months if not treated.

- Some of the merozoite-infected blood cells leave the cycle of asexual multiplication. Instead of replicating, the merozoites in these cells develop into sexual forms of the parasite, called male and female gametocytes that circulate in the bloodstream.

- When a mosquito bites an infected human, it ingests the gametocytes. In the mosquito gut, the infected human blood cells burst, releasing the gametocytes, which develop further into mature sex cells called gametes. Male and female gametes fuse to form diploid zygotes, which develop into actively moving ookinetes that burrow into the mosquito midgut wall and form oocysts.

Growth and division of each oocyst produces thousands of active haploid forms called sporozoites. After 8-15 days, the oocyst bursts, releasing sporozoites into the body cavity of the mosquito, from which they travel to and invade the mosquito salivary glands. The cycle of human infection re-starts when the mosquito takes a blood meal, injecting the sporozoites from its salivary glands into the human bloodstream.

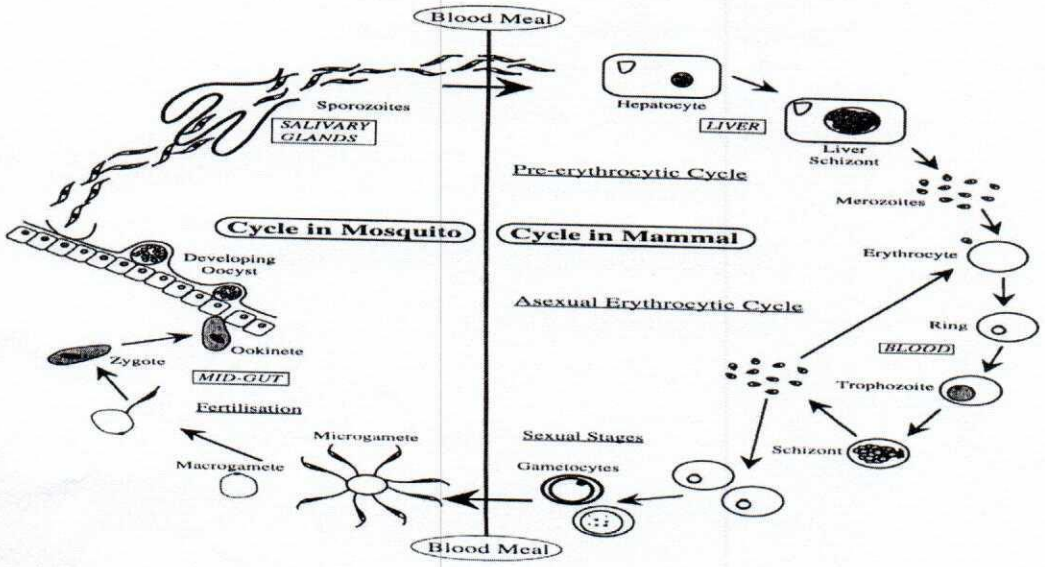

Figure 2.2: Life Cycle of Malaria Fever: (Mendis et al., 2001).

2.4 Treatment and Control

Malaria fever vectors have become more resistant to insecticides and the parasites that cause malaria fever are becoming resistant to chloroquine and possibly other anti-malarial drugs, making prevention and treatment increasingly more difficult and costly than ever (Sharma, 1996). Combination therapy has been shown to increase the efficacy of combining drugs Whitty and Allan (2004) acknowledged the wide spread situation of drug-resistant malaria fever in Africa. Chloroquine resistant malaria fever is now almost universal and resistant to successor drug, sulfadoxine-pyrimethamine (SP) which is growing rapidly (Whitty and Allan, 2004).

Self prescription or medication is a widespread phenomenon. Majority of the malaria fever victims only seek medical examination and treatment from health facilities when the initial attempts have failed resulting in late presentation. (Asenso-Okyere and Dzator, 1995).

2.4.1 Insecticide Treated Nets

Insecticide-treated bed nets (ITNs), including long-lasting insecticidal nets (LLINs), play a primary role in global campaigns to roll back malaria fever in tropical Africa. Effectiveness of treated nets depends on direct impacts on individual mosquitoes including killing and excite-repellency, which vary considerably among vector species due to variations in host-seeking behaviors. While monitoring and evaluation programmes of ITNs have focused on morbidity and all cause mortality in humans, local entomological context receives little attention. Without knowing the dynamics of local vector species and their responses to treated nets, it is difficult to predict clinical outcomes when ITN applications are scaled up across the African continent. Sound model frameworks incorporating intricate interactions between mosquitoes and treated nets are needed to develop the predictive capacity for scale-up applications of ITNs (Weidong et al., 2009).

2.5 Environmental Risk Factors of Malaria Fever

2.5.1 Rainfall

Malaria fever is greatly influenced by rainfall in the tropics. It creates an opportunity for *anopheles* mosquitoes to lay eggs, which can reach adulthood within nine to twelve (9-12) days that are necessary for the mosquito life cycle. Rainfall is one of the climatic variables that aid in the multiplication of mosquito breeding places and increasing humidity, which improves mosquito survival rates. The rainy season is a fertile period for the breeding sites, which are numerous. These species have the highest population density during the rainy season and these accounts for the high incidence of malaria fever at this period of the year (Reid, 2000).

Studies have established complex relationship between malaria fever and rainfall because water is very vital for larval development. A prolonged dry season can decrease mosquito numbers by reducing breeding sites and also reduce malaria fever incidence whiles higher rainfall during the wet season may flush mosquito larvae away (Patz, 2001).

Lindsay et al., (2000) identified a reduction in the infection of malaria in Tanzania as associated with El Niño. It was found out that heavy rainfall may have flushed out Anopheline mosquitoes from their breeding sites thereby increasing the mosquito population. Smith et al (1995) in their study also showed a positive association between the abundance of Anopheles gambiae and rainfall. Increased malaria fever mortality in the Punjab correlated with high rainfall of the previous month (Smith et al, 1995).

In a study of a web-based climate information resources for malaria fever control in Africa, the results showed that rainfall is largely responsible for creating the conditions that allow sufficient surface water for mosquito breeding sites and is therefore recognized as one of the major factors influencing malaria fever transmission (Grover-Kopec et al., 2005).

2.5.2 Topography

Topography generally has a great influence on mosquito replication and thus affect the rate of malaria fever cases. Higher topographies results in cooler temperatures which limits the rate at which the parasite reproduces. Higher elevations therefore result in low rise malaria fever cases as result of the cooler temperatures as you go through higher altitudes thereby elongating the life cycle of the malaria parasite.

Entomologic studies in eight villages to investigate the patterns of malaria fever transmission in different ecologic zones in Eritrea showed a positive relationship between the malaria cases and topography. Mosquito collections conducted for 24 months showed that the biting rates in the higher elevations as a result of the lower temperatures were twice as high as the lowlands (Shillu et al., 2003). The complexity of topography and landscape in the highlands contributes to the spatial heterogeneity of vector abundance and malaria fever transmission intensity. It has implications for the survival of the vector for different altitudes (Minakawa et al., 2002).

Malaria fever has been observed to be a growing problem in African highlands because at high altitudes in the highlands and on hilltops, where malaria fever transmission intensity is low, human populations have poorly developed immunity to malaria fever because exposures are infrequent. (Balls et al., 2004) investigated whether the risk of infection with malaria fever parasites was related to topography in the Usambara Mountains, Tanzania. Clinical surveys were carried out in seven villages, situated at altitudes from 300m to 1650m. Each village was mapped and incorporated into a Digital Terrain Model. Univariate analysis showed that the risk declined with increasing topography and the fact that such elevations washed away water when it rained therefore decreasing potential for water to accumulate. This therefore prevents stagnation of the water that could result in the breeding of

mosquitoes.

Lindsay et al., (2000), discussing the effect of the 1997-98 El Niño on highland malaria fever in Tanzania, discovered quite an opposing results of the malaria incidence with the associated highlands. The study showed that the level of malaria infection was rather following this event than in the previous year, suggesting that heavy rainfall may have washed away mosquito breeding sites. Cohen et al., (2008) in their study of topography-derived wetness indices and household-level malaria fever risk in two communities in the western Kenyan highlands and tried to show the effect of topography on the malaria fever incidence. They found that the transmission of Plasmodium falciparum generally decreases with increasing topography. Knowledge of these local topographic effects may have permitted prediction of regions at high risk of malaria fever within the highlands at small spatial scales. The results indicated that high wetness indices are not merely proxies for valley bottoms, and that hydrologic flow models may prove valuable for predicting areas of high malaria fever risk in highland regions.

2.5.3 Temperature

Malaria Fever incidence is closely linked with temperature. It affects malaria fever transmission in several ways among which we can account for two reasons: either the minimum temperature is so low that it prevents parasite and vector development or the temperature is too high resulting in increased mortality of the vector. A minimum temperature of 16 degrees celsius restricts parasite development and also prevents the development of the vector in its aquatic stages. At 17 degrees celsius parasites develop but not rapidly enough to cause an epidemic (Lindsay and Martens, 1998).

Temperature also plays a fundamental role in the rate of multiplication of the parasite in mosquitoes and directly influences the mosquito development, gonotrophic cycle and

longevity, as well as the duration of the extrinsic cycle of the Plasmodium parasite. In warmer temperatures the mosquitoes develop more rapidly accelerating the mosquito life cycle and replicating rapidly the parasite growth (WHO/AFRO, 2001).

The optimum temperature for the malaria fever parasite extrinsic incubation period is about 20°-27°C whiles the maximum temperature for both vectors and parasites is 40°C. (MARA/ARMA, 1998). Malaria Fever transmission in areas colder than 20°C can still occur because Anophelines often live in houses, which tend to be warmer than external temperatures. Larval development of the mosquito also depends on temperature. Higher temperatures increase the number of blood meals taken and the number of times eggs are laid by the mosquitoes. (Martens et al., 1995).

Broker et al., (2002), studied to see the spatial distributions of Helminth (one type of parasites) in Cameroon. They collected epidemiological and population data. Land surface temperature was derived from NOAA-AVHRR. They used a Logistic regression model to identify significant environmental variables which affect the transmission of infection. The variables used in the regression analysis were mean, minimum and maximum land surface temperature; total annual rainfall and altitude. The result revealed that maximum temperature was an important variable in determining Helminth distribution. At higher temperatures it is realized that female adult mosquitoes feed more frequently and digest blood more rapidly and the Plasmodium parasite matures more rapidly within the female mosquitoes (Githeko et al., 2000).

2.5.4 Climate Change

Climate change and its relation to vector borne disease review was carried out by Githeko et al., (2000) on the whole world. The study revealed that climate variability has a direct influence on vector-borne disease epidemics. A complex interaction exists between man, the

parasite, the vector and the environment and this interaction determines malaria fever's endemicity. Climate therefore has a major impact on vector and the parasite development.

Climate is a major driving force behind malaria fever transmission and climate data are often used to account for the spatial, seasonal and inter- annual variation in malaria fever transmission. The transmission of many infectious diseases varies noticeably by season. For example, the majority of influenza outbreaks in the northern hemisphere occur in mid to late winter (WHO, 2000). In predicting and mapping malaria fever under climate change scenarios, Tonnang et al., (2009) sought information previously generated by entomologists, e.g. on geographical range of vectors and malaria fever distribution in order to build models that will enable prediction and mapping of the potential redistribution of *Anopheles* mosquitoes in Africa. GIS was utilized in this process and it enabled the setting up of an early warning and sustainable strategies for climate change and climate change adaptation for malaria fever vectors control programmes in Africa (Tonnang *et al.*, 2010).

2.5.5 Land Use/Land Cover and Forest

Land use and land cover changes has a significant influence on malaria fever transmission intensity. It affects the spatial and temporal variations in the distribution of anopheline larval habitats. In a study investigated by Munga et al., (2009) the spatial and temporal variations in the distribution of anopheline larval habitats and Land use changes in western Kenya highlands over a 4 year period showed that Anopheles gambiae complex larvae were mainly confined to valley bottoms during both the dry and wet seasons. They were also located in man-made habitats where riparian forests and natural swamps had been cleared. The association between land cover type and occurrence of anopheline larvae was statistically significant.

Forest cover may double the high rate of malaria fever in some of the areas recording high

malaria fever cases. The disease incidence is very high in the forest and forest fringes as compared to plains or urban areas (Sharma, 1991). Mosquitoes in the forested area according to the study were seen to live longer than those in the deforested area in both dry and rainy seasons in the highlands. Forested areas are areas with high humid conditions which favour the ecological reproduction and transmission of the malaria fever parasite (Afrane et al., 2005). Proximity to forest and swamp has both been associated with increased vector density. Broker et al., (2004) established a positive relationship between forest and malaria fever risk. Logistic regression was used to examine the effects of distance to forest on malaria fever prevalence. It was found to escalate the malaria fever rates consistent with other studies. Proximity to forest was found to be a major risk factor for malaria fever.

In a study in the Brazilian Amazon, forest cover can increase malaria fever incidence by nearly 50%. Open spaces and partially sunlit pools of water, typical conditions of deforested landscapes, provide an ideal habitat in which the Anopheles mosquito can thrive. The study revealed that a 4 % change in forest cover was associated with a 48% increase in malaria fever incidence. (Hirschfeld, 2010).

2.6 Related work to the Diagnosis of Malaria Fever

Prompt and accurate diagnosis is the key to effective disease management, one of the main interventions of the Global Malaria Control Strategy (A global strategy for malaria control. Geneva, World Health Organization, 1993). It is thus of concern that poor diagnosis continues to hinder effective malaria control. This is due to a combination of factors, including non-specific clinical presentation of the disease, high prevalence of asymptomatic infection in some areas, lack of resources and insufficient access to trained health care providers and health facilities, and widespread practice of self-treatment for clinically suspected malaria.

Due to the adverse effect of malaria on people and economy, researchers had undergone series of researches to develop systems that could diagnose malaria fever.

Today, approximately 40 percent of the world population mostly that living in the world poorest countries is at risk of malaria (Sola et al., 2010). This makes the early diagnosis of malaria crucial as it reduces the morbidity and mortality rate of this disease. In Nigeria for instance, on the average, each Nigerian suffers at least two or more attacks every year. While millions recover, hundreds of thousands are not so lucky. This single disease accounts for about 60 percent of outpatient visits and 30 percent of hospitalizations; 25 percent of deaths in children under one year old; and 11 percent of maternal death, a heavy burden on Nigerian families, communities, health system, and workforce (Sola et al., 2010).

Intelligent systems have become vital in the growth and survival of the healthcare sector. A good number of systems have been developed to manage tropical diseases. Adekoya et al., (2008) developed an expert system on tropical diseases to assist paramedical staff during training and in the diagnosis of many common diseases presented at their clinics. The system was flexible, friendly, and usable by people without much background in computer operations. The study concluded that the implementation of the system reduced doctor's workload during consultation and eased other problems associated with hospital consultations.

A fuzzy expert system for the management of malaria designed by Djam et al., (2011) attempted to incorporate fuzzy techniques and develop a fuzzy expert system for the management of malaria. Here, the study revealed that the use of fuzzy logic for medical diagnosis provides an efficient way to assist inexperienced physicians to arrive at the final diagnosis of malaria more quickly and efficiently. The developed system provided decision-support platform to assist malaria researchers, physicians and other health practitioners in

malaria endemic regions (Djam and Kimbi, 2011).

Olabiyisi et al., (2011) designed a decision-support model for diagnosing tropical diseases using fuzzy logic. The aim of this research was to detect the disease of a patient based on the patient complaints and also the level of severity of the patient complaints. Cognitive analysis of multiple sclerosis utilizing fuzzy cluster means designed by (Imiavan and Obil, 2012), neuro solutions and crystal reports were used for neural network analysis and graphical representation to aid in the diagnosis of multiple sclerosis, which eliminates the challenges posed by the shortage of medical experts.

Obot and Uzoka, (2008) presented a fuzzy rule-base framework for the management of tropical diseases with main focus on malaria fever. The system involves fuzzification, inference and defuzzification. The system is able to diagnose malaria fever case as either mild, moderate severe or very severe. The symptoms and signs of the diseases were fuzzified based on recommendation of some medical personnel; the fuzzified variables were composed, inferred and later defuzzified to give final diagnosis.

Nicholas et al., (2010). State that an automated image processing method for the diagnosis and classification of malaria fever on thin blood smears was developed. The image classification system is designed to positively identify malaria parasites present in thin blood smears, and differentiate the species of malaria fever. Features based on image characteristics such as colour, texture and geometry as well as original features that mimic the qualities used by microscopists when diagnosing malaria fever is generated from the erythrocytes which are candidates for infection. A tree classifier with two nodes using BFF a neural network is used to determine whether or not a cell is infected, and if so, the species of the malaria. The limitations of this system include inherent limitations of microscopy such as the degradation of slide quality with time apart from the fact that the system is too expensive. It is exclusively

reserve for medical experts to aid in the identification of malaria fever.

According to Uzoka et al., 2009 state that medical decision support system using analytic hierarchy process was used to diagnose malaria fever. The system was designed to overcome the conventional method of malaria diagnosis. Knowledge components used include: patient information, patient characteristics, medical history, patient examination, chemotherapy and symptom intensity. The system is able to diagnose and determine the priority order (ranking) of basic malaria diagnosis criteria. Apart from the fact that the system fails to provide therapy for the diagnosed malaria case, it is not generally accessible for people.

World Health Organization, Decision Tree Analysis is used to diagnose malaria using the basic symptoms for malaria which include, fever, loss of appetite, bitter taste, body and joint pains and headache. The system fails to present ranked list of symptoms according to severity of malaria and it only carries out diagnosis. Malaria Fever is preventable and curable. Early and accurate diagnosis of malaria fever is essential for effective and life-saving treatment. The development of Bayesian based approach is the nearest response to this call because it makes the diagnosis of this deadly disease much easier, faster, and more accurate.

Zeon et al., (2013). Developed a disease prediction system, DOCAID, for predicting typhoid, malaria, jaundice, tuberculosis and gastroenteritis based on patient symptoms and complaints employing Naive Bayes Classifier algorithm. An accuracy of 91% accuracy in predicting the diseases were reported by the authors.

Theodorali et al., (2010). Developed prediction model for predicting the final outcome in patients suffering from severe injuries after accident. The analysis included a comparison of data mining techniques using classification, clustering and association algorithms. Using this analysis they obtained results in terms of sensitivity, specificity, positive predictive value and negative predictive value and compared the results between different prediction models.

2.7 Bayesian Approach to Diagnosis of Malaria Fever

Bayesian approaches are powerful tools for decision and reasoning under uncertain conditions. Naïve Bayes, used in this study, is a straight forward Bayesian learning method based on strong independence assumption.

Naïve Bayes is based on the Bayesian theorem. This classification technique analyses the relationship between each attribute and the class for each instance to derive a conditional probability for the relationships between the attribute values and the class. The principle behind Naïve Bayes for classifications a fairly simple process. During training, the probability of each class is computed by counting how many times it occurs in the training dataset. This is called the "prior probability" P(C=c). In addition to the prior probability, the algorithm also computes the probability for the instance x given c with the assumption that the attributes are independent. This probability becomes the product of the probabilities of each single attribute. The probabilities can then be estimated from the frequencies of the instances in the training set. Numeric attributes can have a large number (possibly infinite) of values and the probability cannot be estimated from the frequency distribution, which tend to reduce the performance of Naïve Bayes (Frank et al., 2000).

Little work has been done in terms of application of Naïve Bayes to solving data mining problems. It has however been used in conjunction with other techniques to solve classification and prediction tasks (Zhou et al., 2003). However, numerous studies have compared Naïve Bayes with other machine learning (data mining) techniques. Naïve Bayes keenly competed with other techniques such as decision tree and neural networks. This may probably be due to the fact that researchers have not paid attention to the capabilities of Naïve Bayes in data mining. Nonetheless, this does not make Naïve Bayes less useful since it

outperformed (in most cases) other techniques when they were compared (Langley et al., 1992, and Frank et al., 2000).

The Bayesian approach serves as the foundation for the derivation of diagnostic candidates, i.e.

- Deducing whether a candidate diagnosis d_k is consistent with the observations, and
- The posterior probability $\Pr(d_k)$ of that candidate being the actual diagnosis.

With respect to (1), rather than computing $\Pr(d_k)$ for *all* possible candidates, just to find that most of them have $\Pr(d_k) = 0$, search algorithms are typically used instead, such as CDA* Williams and Ragno (2007), SAFARI Feldman et al., (2008), or just a minimal hitting set (MHS) algorithm when conflict sets are available, e.g. De Kleer and Williams (1987), but the Bayesian probability framework remains the basis. In this section we will briefly describe the contemporary approach to the derivation of candidates and their posterior probability. In the following, we assume weak fault models.

Consider a particular process, involving a set of components, that either yields a nominal result or a failure. For instance, in a logic circuit a process is the sub-circuit (cone) activity that results in a particular primary output. In software a process is the sequence of software component activity (e.g., statements) that results in a particular return value. The result of a process is either nominal (pass) or an error (fail).

Definition 1: Let $S_f = \{c_j | c_j$ involved in a failing process$\}$, and let $S_p = \{c_j | c_j$ involved in a passing process$\}$, denote the *fail set* and *pass set*, respectively.

Approaches for fault diagnosis that assume persistent, weak fault models often generate candidates based on fail sets (conflict sets), essentially using an MHS algorithm to derive minimal candidates. Recent approaches that allow intermittency also take into account pass sets. A fail set indicts components, whereas a pass set exonerates components. The extent of

indictment or exoneration is computed using Bayes' rule. In the following we assume that a number of pass and fail sets have been collected, either by static modeling (e.g, logic circuits, where each primary output yields a pass or fail set) or by dynamic profiling (e.g, software), where each run yields a pass or fail set, both known as a *spectrum* Abreu et al., (2007).

Definition 2: Let N denote the number of passing and failing processes. Let N_f and N_p, $N_f+N_p = N$, denote the number of fail and pass sets, respectively. Let A denote the $N \times M$ *activity matrix* of the system, where a_{ij} denotes whether component j was involved in process i ($a_{ij} = 1$) or not ($a_{ij} = 0$). Let e denote the *error vector*, where e_i signifies whether process i has passed ($e_i = 0$) or failed ($e_i = 1$).

The observations (A, e) are input to the Bayesian probability update process.

Ranking Diagnoses Let $\Pr(j) = p_j$ denote the prior probability that a component c_j is at fault. Assuming components fail independently the prior probability of a candidate d_k is given by

$$\Pr(d_k) = \prod_{j \in S_N} \Pr(\{j\}) \cdot \prod_{j \in S_P} (1 - \Pr(\{j\}))$$

For each observation $obs_i = (A_{i*}, e_i)$ the posterior probabilities are updated according to Bayes rule (naive Bayes classifying)

$$\Pr(d_k|obs_i) = \frac{\Pr(obs_i|d_k)}{\Pr(obs_i)} \cdot \Pr(d_k)$$

The denominator $\Pr(obs_i)$ is a normalizing term that is identical for all d_k and thus needs not be computed directly. $\Pr(obs_i|d_k)$ is defined as

$$\Pr(d_k|obs_i) = \frac{\Pr(obs_i|d_k)}{\Pr(obs_i)} \cdot \Pr(d_k)$$

As mentioned earlier, rather than updating each candidate only candidates derived from an MHS algorithm are updated implying that the 0-clause need not be considered.

Many policies exist for ε De Kleer (2006). Three policies can be distinguished. The first policy, denoted $\varepsilon^{(0)}$ equals the classical MBD policy for persistent, weak faults, and is defined as follows

$$\varepsilon^{(0)} = \begin{cases} \frac{E_P}{E_P + E_F} & \text{if } e_i = 0 \\ \frac{E_F}{E_P + E_F} & \text{if } e_i = 1 \end{cases}$$

Where $E_P = 2^M$ and $E_F = (2^{|dk|} - 1) \cdot 2^{M-|dk|}$ are the number of passed and failed observations that can be explained by diagnosis d_k, respectively. A disadvantage of this classical policy is that pass sets, apart from making single faults more probable than multiple faults, do not help much in pinpointing the faults, in particular for weak fault models which do not rule out any candidates (the 2^M term in Eq. 1). In addition, there is no way to distinguish between diagnoses with the same cardinality, because the terms are merely a function of the cardinality of the diagnosis candidate.

The next two, intermittent policies account for the fact that components of pass sets should to some extent be exonerated. In the following we distinguish between two policies, De Kleer (2007) and Abreu et al., (2008a) which are defined as

$$\varepsilon = \begin{cases} g(d_k) & \text{if } e_i = 0 \\ d & \end{cases}$$

$$\varepsilon = \begin{cases} g(d_k)^m & \text{if } e_i = 0 \\ 1 - g(d_k)^m & \text{if } e_i = 1 \end{cases}$$

where m =$\prod$ Qj $\in$ dk [aij = 1] is the number of faulty components according to d_k involved in process i. Note that a term $g(d_k)$ is used rather than the real individual component intermittency parameters g_j. As mentioned earlier, this is due to the fact that obtaining g_j from pass and fail sets where multiple intermittent failures are involved was far from trivial.

Instead, an "effective" intermittency parameter $g(d_k)$ is estimated for the candidate d_k by counting how many times components of d_k are involved in pass and fail sets. In both strategies $g(d_k)$ is approximated by

$$g(d_k) = \frac{\sum_{i=1..N} [(\bigvee_{j \in d_k} a_{ij} = 1) \wedge e_i = 0]}{\sum_{i=1..N} [\bigvee_{j \in d_k} a_{ij} = 1]}$$

Where $[\cdot]$ is Iverson's operator Iverson [1962] ([true] = 1], [false] = 0]).

Policy , De Kleer (2007) is a variant of Abreu et al., (2008a), which approximates the probability$\prod j \in dk\, gj$ that all m components in dk exhibit good behavior by g(dk)m assuming that all components of dk have equal g values. This takes into account the fact that the failure probability changes when multiple intermittent faults are involved.

The diagnose and decide model is related to several previously considered problems. Adaptive sensing domains E. K. Chong, C. M. Kreucher, and A. O. Hero (2009) are similar but, in contrast to DAD models, the underlying state may change. In sequential analysis "Wald (1947) problems" involve repeated sampling before making a decision but assume that repeated tests will generate independent outcomes. In Bayesian experimental design there is a prior over the possible hypotheses (i.e., classes) and probabilities for test outcomes, Chaloner and Verdinelli (1995). If the observations are noise-free, this is an optimal decision-tree problem, and greedy methods can produce a solution that utilizes within $O(\log n)$ tests of an optimal approach where n is the number of hypotheses. The closest related work is Golovin et al., (2010) work on Bayesian experimental design in the presence of noisy observations. Their objective is to only consider policies which guarantee that when a decision is chosen that it is identical to the decision that would be made if all tests were performed (risk minimization). A decision tree can then be generated that determines what tests should be

performed to decide which equivalence class is the correct one while minimizing cost. They prove that finding the minimum cost decision tree can be solved using adaptive submodularity. This results in a greedy solution which is efficient and has performance bounds relative to their objective function, but it does not allow the cost of the tests to be balanced with the loss that results from making a decision.

Therefore, there is a pressing need to research into the best methods of diagnosis and using existing approaches, such as Bayesian base approach to diagnose, to have effective and accurate diagnosis on malaria fever. Guerin et al., (2002).

CHAPTER THREE

METHODOLOGY

3.1 Bayesian Classification model for Malaria fever diagnosis

Bayesian approaches are powerful tools for decision and reasoning under uncertain conditions. Bayesian classifiers employ probabilistic concept representations, and range from the Naïve Bayes (NB) to Bayesian networks (Domingos and Pazzani, 1997). Bayesian reasoning is based on the assumption that the relation between attributes can be represented as a probability distribution. Naïve Bayes, used in this study, is a straight forward Bayesian learning method based on strong independence assumption.

3.2 Naïve Bayes Classification

A Bayesian classifier is a probabilistic model of what is happening in data, which estimates the class for new data item. Naïve Bayesian has been successfully applied in solving various problems.

In medical data mining, Naive Bayes classification plays an important role. It is a probabilistic classification based on the Bayes theorem.

The probability of data record X having the class label Ci is

$$P(Cj/X) = \frac{P(X/Ci) * P(Ci)}{P(X)} \qquad (3.1)$$

The class label Ci with largest conditional probability value determines the category of the data record.

It is very practical when the dimensionality of the inputs is high. The word "Naive" implies the independence between all attributes. Naive Bayes (NB) is a machine-learning method that has been used for over 50 years in biomedical informatics, Caruana et al, (2006). It requires

only small amount of training data to estimate the parameter which is very useful for health care applications. Naive Bayes computes conditional probabilities of the classes given with the instance and select the class with highest posterior, Karpagavalli et al., (2009). Regardless of this simplified assumption and naive design, naive Bayes classifier works well in many complex real world situations. Bayes classification is outperformed by current approaches, like boosted trees or random forests.

3.3 Description of Research Work

The research work is structured into 3 stages as represented in figure 3.1. The first stage includes data collection and pre processing and producing training data and analyzing variables. In the second stage we use WEKA tool to check the accuracy of the models. The third stage presents explanation of the prediction mode.

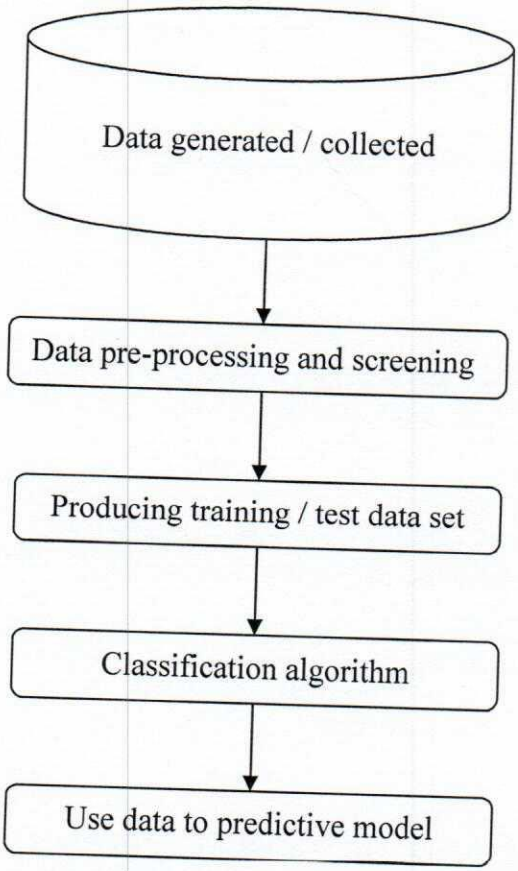

Figure 3.1: Research structure of the work

3.3.1 Description of Data Generated

The data generated were mainly records of patients with malaria cases comprising the symptoms and complaints made by the patients. The data are in two phases, the first set of data was hundred in number and it was used as training set. The second phase consists of fifty sets of data and was used as testing set.

All the data were assigned classes and they grouped the cases of malaria into five classes according to the level of severity. Very High, High, Moderate, Low and Very Low using the symptoms of malaria of each patient. There are nineteen conditional attributes (symptoms) and one decision attribute, shown in the table 3.1 below.

Table 3.1: Attributes of malaria fever (Retrieved from Computer Science and Telecommunications No.1 (33), Adetunmbi et al., (2012)

S/N	ABREVIATION	ATTRIBUTE	ATTRIBUTE TYPE
A1	WKN	Weakness	Discrete
A2	APB	Abdominal Pain	Discrete
A3	COH	Cough	Discrete
A4	BOP	Body Pain	Discrete
A5	FVR	Fever	Discrete
A6	RGR	Rigour	Discrete
A7	COD	Cold	Discrete
A8	ANR	Anorexia	Discrete
A9	HEC	Headache	Discrete
A10	CAH	Catarrh	Discrete
A11	INS	Insomnia	Discrete
A12	YEU	Yellow Urine	Discrete
A13	VOM	Vomiting	Discrete
A14	JOP	Joint pain	Discrete
A15	DSN	Dizziness	Discrete
A16	ILL	Ill-looking	Discrete
A17	COV	Convulsion	Discrete
A18	BOT	Body Temperature	Discrete
A19	DIA	Diarrhea	Discrete
A20	MAL DIAG	Malaria Diagnosed	Discrete

3.3.2 Data Analysis Software

WEKA ("Waikato Environment for Knowledge Analysis") is a popular suite of machine learning software written in Java, developed at University of Waikato, New Zealand. WEKA 3.6.9 is used to analyze data. It contains a collection of algorithms for data analysis and predictive modeling, together with GUI (Graphical User Interface) for easy access to this functionality. WEKA supports several standard data mining tasks like data pre-processing, classification, clustering, association rules, visualization and feature selection. It also offer different test option like Cross validation, using training set, test set, percentage split etc. Naive Bayes method is used to perform the classification process.

3.3.3 Data Pre-processing

The database of this study contains a total of 150 disease cases. After data pre-processing, all data was divided into two sections: The training data set with 100 and test data set with 50 respectively to verify the accuracy model. The data is analyzed and implemented in WEKA tool. Data mining finds out the valuable information hidden in huge volumes of data. WEKA tool is a collection of machine learning algorithms for data mining techniques, written in Java. It consists of data pre-processing, classification, regression, association rules, clustering and visualization tools. Naive Bayes method is used to perform the classification process.

3.4 Experimental Set Up and Results

Bayesian approach used the training data set to build a classification model. It is easier to work around with numbers since the number of bytes that will be reserved for integer numbers will be smaller. For the decision attribute, the data Very Low, Low, Moderate, High and Very High are thus converted to integer numbers 1,2,3,4 and 5 respectively. For the conditional attributes (symptoms) A1 to A19, each symptom is classified as either high or low. Value high is converted to integer 1; Value Low is converted to integer 2 while default exists for symptom not applicable and takes a value of 0.

3.5 Generated Dataset for Training set

Figure 3.2: Training set

S/N	WKN	APB	COH	BOP	FVR	RGR	COD	ANR	HEC	CAH	INS	YEU	VOM	JOP	DNS	ILL	COV	BOT	DIA	DECISION
1	HIGH	LOW	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	HIGH	LOW	HIGH	HIGH	LOW	LOW	LOW	LOW	VL
2	LOW	LOW	LOW	LOW	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	HIGH	LOW	LOW	LOW	VL
3	HIGH	HIGH	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	L
4	LOW	HIGH	LOW	LOW	LOW	HIGH	LOW	LOW	LOW	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	L
5	HIGH	LOW	LOW	LOW	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	L
6	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	HIGH	LOW	LOW	HIGH	LOW	LOW	LOW	L
7	HIGH	HIGH	LOW	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	LOW	HIGH	HIGH	LOW	LOW	LOW	LOW	HIGH	LOW	L
8	LOW	LOW	LOW	HIGH	HIGH	LOW	LOW	LOW	HIGH	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	M
9	LOW	HIGH	LOW	LOW	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	LOW	HIGH	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	M
10	LOW	HIGH	LOW	LOW	HIGH	LOW	HIGH	LOW	HIGH	LOW	LOW	HIGH	HIGH	HIGH	LOW	LOW	LOW	LOW	LOW	M
11	LOW	LOW	HIGH	LOW	HIGH	LOW	LOW	HIGH	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	HIGH	LOW	M
12	LOW	LOW	LOW	HIGH	HIGH	LOW	LOW	LOW	LOW	LOW	HIGH	HIGH	LOW	LOW	LOW	LOW	HIGH	LOW	HIGH	M
13	LOW	HIGH	HIGH	LOW	LOW	HIGH	HIGH	HIGH	HIGH	LOW	HIGH	LOW	LOW	HIGH	LOW	HIGH	LOW	HIGH	HIGH	H
14	LOW	LOW	HIGH	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	LOW	LOW	HIGH	LOW	HIGH	HIGH	H
15	HIGH	LOW	HIGH	LOW	LOW	HIGH	HIGH	HIGH	HIGH	LOW	LOW	HIGH	HIGH	HIGH	HIGH	LOW	LOW	HIGH	HIGH	H
16	HIGH	LOW	HIGH	HIGH	LOW	HIGH	HIGH	LOW	HIGH	LOW	HIGH	HIGH	LOW	HIGH	HIGH	HIGH	LOW	LOW	HIGH	H
17	LOW	HIGH	HIGH	HIGH	HIGH	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	HIGH	LOW	H
18	LOW	HIGH	LOW	HIGH	LOW	HIGH	HIGH	LOW	HIGH	LOW	HIGH	HIGH	LOW	HIGH	LOW	HIGH	LOW	HIGH	HIGH	H
19	HIGH	HIGH	HIGH	LOW	LOW	HIGH	LOW	HIGH	HIGH	HIGH	LOW	HIGH	HIGH	LOW	LOW	HIGH	LOW	LOW	LOW	H
20	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	LOW	LOW	HIGH	HIGH	HIGH	LOW	LOW	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	VH
21	LOW	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	LOW	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	VH
22	LOW	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	LOW	HIGH	HIGH	LOW	HIGH	HIGH	VH
23	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	LOW	LOW	HIGH	HIGH	HIGH	HIGH	VH
24	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	VL
25	LOW	LOW	LOW	LOW	HIGH	HIGH	HIGH	LOW	LOW	LOW	HIGH	LOW	LOW	LOW	LOW	LOW	HIGH	LOW	LOW	VL
26	LOW	LOW	HIGH	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	HIGH	LOW	LOW	HIGH	LOW	LOW	LOW	L
27	HIGH	LOW	LOW	LOW	HIGH	HIGH	HIGH	LOW	LOW	LOW	HIGH	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	LOW	L
28	HIGH	LOW	HIGH	LOW	HIGH	LOW	LOW	LOW	LOW	HIGH	LOW	LOW	LOW	LOW	LOW	HIGH	LOW	LOW	LOW	L
29	LOW	LOW	LOW	HIGH	HIGH	HIGH	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	LOW	LOW	HIGH	HIGH	HIGH	L
30	LOW	LOW	LOW	LOW	LOW	HIGH	HIGH	LOW	HIGH	HIGH	LOW	LOW	LOW	HIGH	HIGH	HIGH	HIGH	LOW	HIGH	L
31	LOW	HIGH	HIGH	LOW	HIGH	HIGH	HIGH	LOW	HIGH	HIGH	LOW	HIGH	HIGH	LOW	HIGH	LOW	HIGH	HIGH	HIGH	M
32	HIGH	LOW	HIGH	LOW	HIGH	HIGH	LOW	LOW	HIGH	LOW	LOW	LOW	LOW	HIGH	LOW	LOW	HIGH	HIGH	HIGH	M
33	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	LOW	HIGH	LOW	LOW	HIGH	HIGH	HIGH	LOW	M
34	HIGH	HIGH	HIGH	LOW	LOW	HIGH	HIGH	LOW	HIGH	LOW	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	HIGH	LOW	HIGH	M
35	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	LOW	LOW	HIGH	LOW	LOW	LOW	LOW	HIGH	HIGH	HIGH	LOW	HIGH	LOW	M
36	LOW	LOW	LOW	HIGH	HIGH	LOW	LOW	LOW	LOW	HIGH	HIGH	HIGH	HIGH	LOW	LOW	LOW	HIGH	LOW	LOW	H
37	LOW	HIGH	LOW	LOW	HIGH	LOW	LOW	HIGH	LOW	LOW	LOW	LOW	LOW	HIGH	LOW	LOW	LOW	LOW	LOW	H
38	LOW	HIGH	LOW	HIGH	LOW	LOW	LOW	HIGH	LOW	LOW	HIGH	LOW	LOW	LOW	LOW	HIGH	HIGH	LOW	LOW	H
39	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	HIGH	HIGH	HIGH	LOW	LOW	HIGH	LOW	LOW	LOW	HIGH	LOW	LOW	H
40	HIGH	LOW	LOW	LOW	LOW	LOW	HIGH	HIGH	HIGH	LOW	LOW	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	LOW	LOW	H
41	LOW	HIGH	HIGH	HIGH	HIGH	LOW	LOW	HIGH	LOW	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	HIGH	LOW	LOW	H
42	LOW	LOW	LOW	HIGH	LOW	LOW	HIGH	LOW	LOW	HIGH	LOW	LOW	LOW	HIGH	HIGH	LOW	HIGH	HIGH	HIGH	H
43	HIGH	HIGH	LOW	LOW	LOW	LOW	HIGH	HIGH	LOW	HIGH	LOW	HIGH	HIGH	LOW	HIGH	LOW	LOW	LOW	LOW	VH
44	HIGH	LOW	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	LOW	HIGH	LOW	HIGH	LOW	LOW	LOW	HIGH	LOW	LOW	VH
45	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	HIGH	LOW	HIGH	LOW	LOW	LOW	VH
46	HIGH	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	HIGH	LOW	LOW	LOW	HIGH	LOW	LOW	LOW	LOW	VH
47	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	HIGH	LOW	LOW	HIGH	VL
48	LOW	LOW	LOW	LOW	HIGH	HIGH	LOW	LOW	LOW	HIGH	HIGH	LOW	LOW	LOW	LOW	LOW	HIGH	LOW	LOW	VL
49	HIGH	HIGH	HIGH	HIGH	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	HIGH	LOW	LOW	HIGH	LOW	LOW	L
50	LOW	LOW	LOW	LOW	HIGH	HIGH	HIGH	LOW	HIGH	LOW	HIGH	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	LOW	L
51	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	LOW	LOW	LOW	HIGH	LOW	LOW	LOW	LOW	LOW	HIGH	LOW	LOW	LOW	L
52	LOW	LOW	LOW	HIGH	HIGH	LOW	HIGH	LOW	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	LOW	LOW	HIGH	HIGH	HIGH	L
53	HIGH	LOW	LOW	LOW	LOW	HIGH	LOW	LOW	HIGH	HIGH	LOW	LOW	LOW	HIGH	HIGH	HIGH	HIGH	LOW	HIGH	L
54	LOW	HIGH	HIGH	LOW	HIGH	LOW	HIGH	LOW	HIGH	HIGH	LOW	LOW	HIGH	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	M
55	LOW	LOW	HIGH	LOW	HIGH	HIGH	LOW	LOW	HIGH	LOW	LOW	HIGH	LOW	HIGH	LOW	LOW	HIGH	HIGH	HIGH	M
56	HIGH	LOW	LOW	HIGH	HIGH	HIGH	LOW	HIGH	HIGH	LOW	HIGH	LOW	HIGH	LOW	LOW	HIGH	HIGH	HIGH	LOW	M
57	HIGH	HIGH	HIGH	LOW	LOW	LOW	LOW	LOW	HIGH	LOW	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	HIGH	LOW	HIGH	M
58	HIGH	LOW	HIGH	HIGH	HIGH	LOW	LOW	LOW	HIGH	LOW	LOW	HIGH	LOW	HIGH	HIGH	HIGH	LOW	HIGH	LOW	M
59	LOW	LOW	LOW	HIGH	HIGH	HIGH	LOW	LOW	HIGH	HIGH	HIGH	HIGH	HIGH	LOW	LOW	LOW	HIGH	LOW	LOW	H
60	LOW	HIGH	LOW	LOW	LOW	LOW	LOW	HIGH	LOW	HIGH	LOW	LOW	LOW	HIGH	LOW	LOW	LOW	LOW	LOW	H
61	LOW	HIGH	LOW	HIGH	LOW	LOW	LOW	HIGH	LOW	LOW	HIGH	HIGH	HIGH	HIGH	LOW	LOW	HIGH	HIGH	LOW	H
62	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	HIGH	HIGH	HIGH	LOW	LOW	HIGH	LOW	HIGH	HIGH	HIGH	LOW	LOW	H
63	HIGH	LOW	LOW	HIGH	LOW	LOW	HIGH	HIGH	HIGH	LOW	HIGH	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	LOW	LOW	H
64	LOW	HIGH	HIGH	HIGH	HIGH	LOW	LOW	HIGH	LOW	HIGH	HIGH	LOW	LOW	HIGH	LOW	LOW	HIGH	LOW	LOW	H
65	LOW	LOW	LOW	HIGH	LOW	HIGH	HIGH	LOW	LOW	HIGH	LOW	LOW	LOW	HIGH	HIGH	LOW	HIGH	LOW	HIGH	H
66	HIGH	HIGH	LOW	LOW	LOW	LOW	HIGH	LOW	LOW	HIGH	LOW	LOW	HIGH	LOW	HIGH	LOW	LOW	LOW	LOW	VH
67	HIGH	LOW	HIGH	LOW	HIGH	HIGH	LOW	LOW	LOW	LOW	HIGH	LOW	HIGH	LOW	LOW	LOW	HIGH	LOW	LOW	VH
68	HIGH	LOW	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	HIGH	HIGH	HIGH	LOW	HIGH	LOW	LOW	VH
69	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	HIGH	LOW	LOW	LOW	HIGH	LOW	LOW	LOW	HIGH	VH
70	LOW	LOW	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	HIGH	HIGH	LOW	LOW	LOW	LOW	VL
71	LOW	LOW	LOW	LOW	HIGH	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	HIGH	HIGH	LOW	LOW	LOW	VL
72	LOW	HIGH	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	HIGH	L
73	LOW	HIGH	LOW	LOW	LOW	HIGH	LOW	LOW	LOW	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	LOW	HIGH	HIGH	L
74	HIGH	LOW	LOW	LOW	HIGH	HIGH	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	L
75	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	HIGH	LOW	LOW	LOW	HIGH	LOW	L
76	HIGH	HIGH	LOW	HIGH	LOW	LOW	LOW	LOW	HIGH	LOW	LOW	LOW	HIGH	HIGH	LOW	LOW	LOW	LOW	LOW	L
77	LOW	LOW	LOW	HIGH	HIGH	HIGH	LOW	LOW	HIGH	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	HIGH	M
78	LOW	HIGH	LOW	LOW	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	LOW	HIGH	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	M
79	HIGH	LOW	LOW	LOW	HIGH	LOW	HIGH	LOW	HIGH	LOW	LOW	HIGH	HIGH	HIGH	LOW	LOW	LOW	LOW	LOW	M
80	LOW	LOW	HIGH	LOW	HIGH	LOW	LOW	HIGH	HIGH	LOW	LOW	LOW	HIGH	LOW	LOW	LOW	LOW	HIGH	HIGH	M
81	HIGH	LOW	HIGH	HIGH	HIGH	LOW	LOW	LOW	LOW	LOW	HIGH	HIGH	LOW	LOW	LOW	LOW	HIGH	LOW	HIGH	M
82	LOW	HIGH	HIGH	LOW	LOW	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	LOW	LOW	HIGH	LOW	HIGH	HIGH	HIGH	H
83	LOW	LOW	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	LOW	LOW	HIGH	LOW	HIGH	HIGH	H
84	HIGH	LOW	HIGH	LOW	LOW	HIGH	HIGH	HIGH	HIGH	LOW	HIGH	LOW	HIGH	HIGH	HIGH	LOW	LOW	HIGH	HIGH	H
85	HIGH	LOW	LOW	LOW	LOW	HIGH	HIGH	LOW	HIGH	LOW	HIGH	HIGH	LOW	HIGH	HIGH	HIGH	LOW	LOW	HIGH	H
86	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	HIGH	LOW	LOW	LOW	LOW	LOW	HIGH	LOW	H
87	LOW	HIGH	LOW	HIGH	LOW	LOW	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	LOW	HIGH	LOW	HIGH	LOW	HIGH	HIGH	H
88	HIGH	LOW	HIGH	LOW	LOW	HIGH	LOW	HIGH	HIGH	HIGH	LOW	LOW	HIGH	LOW	LOW	HIGH	LOW	LOW	LOW	H
89	HIGH	HIGH	LOW	LOW	HIGH	HIGH	LOW	LOW	HIGH	HIGH	HIGH	LOW	LOW	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	VH
90	HIGH	HIGH	LOW	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	HIGH	LOW	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	VH
91	HIGH	HIGH	HIGH	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	LOW	HIGH	HIGH	LOW	HIGH	VH
92	HIGH	LOW	HIGH	HIGH	HIGH	LOW	HIGH	HIGH	LOW	LOW	HIGH	HIGH	HIGH	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	VH
93	LOW	LOW	HIGH	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	VL
94	LOW	LOW	LOW	LOW	HIGH	HIGH	HIGH	LOW	LOW	LOW	HIGH	LOW	LOW	HIGH	HIGH	LOW	HIGH	LOW	LOW	VL
95	LOW	LOW	HIGH	LOW	LOW	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	HIGH	LOW	LOW	HIGH	LOW	LOW	LOW	L
96	HIGH	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	LOW	LOW	LOW	HIGH	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	LOW	L
97	HIGH	LOW	HIGH	LOW	HIGH	LOW	LOW	LOW	LOW	HIGH	LOW	LOW	LOW	LOW	HIGH	HIGH	LOW	HIGH	LOW	L
98	LOW	LOW	LOW	LOW	HIGH	HIGH	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	LOW	LOW	HIGH	LOW	HIGH	L
99	LOW	LOW	LOW	LOW	LOW	HIGH	HIGH	LOW	HIGH	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	LOW	HIGH	L
100	LOW	HIGH	HIGH	LOW	HIGH	HIGH	LOW	LOW	HIGH	HIGH	LOW	HIGH	LOW	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	M

3.5.1 Generated Dataset for Testing

Figure 3.3: Testing set

S/N	WKN	APB	COH	BOP	FVR	RGR	COD	ANR	HEC	CAH	INS	YEU	VOM	JOP	DNS	ILL	COV	BOT	DIA	DECISION
1	HIGH	LOW	HIGH	LOW	HIGH	LOW	LOW	LOW	HIGH	HIGH	LOW	LOW	LOW	HIGH	LOW	LOW	HIGH	HIGH	HIGH	M
2	LOW	LOW	HIGH	HIGH	HIGH	HIGH	LOW	LOW	HIGH	HIGH	HIGH	LOW	HIGH	LOW	LOW	HIGH	HIGH	HIGH	LOW	M
3	HIGH	HIGH	HIGH	LOW	HIGH	HIGH	HIGH	LOW	HIGH	LOW	LOW	LOW	HIGH	HIGH	HIGH	HIGH	HIGH	LOW	HIGH	M
4	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	LOW	HIGH	HIGH	LOW	LOW	HIGH	LOW	M
5	LOW	HIGH	HIGH	HIGH	HIGH	LOW	LOW	LOW	LOW	HIGH	HIGH	HIGH	LOW	LOW	LOW	HIGH	LOW	LOW	LOW	H
6	LOW	HIGH	HIGH	LOW	HIGH	LOW	LOW	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	H
7	HIGH	HIGH	LOW	HIGH	LOW	LOW	LOW	HIGH	LOW	LOW	HIGH	LOW	LOW	HIGH	LOW	HIGH	HIGH	LOW	LOW	H
8	HIGH	LOW	LOW	LOW	LOW	HIGH	LOW	HIGH	HIGH	HIGH	LOW	LOW	HIGH	HIGH	LOW	LOW	HIGH	LOW	LOW	H
9	HIGH	HIGH	HIGH	LOW	LOW	LOW	HIGH	HIGH	HIGH	LOW	LOW	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	LOW	LOW	H
10	LOW	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	LOW	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	HIGH	LOW	LOW	H
11	LOW	HIGH	HIGH	HIGH	LOW	LOW	HIGH	LOW	LOW	HIGH	LOW	LOW	LOW	HIGH	HIGH	LOW	HIGH	HIGH	HIGH	H
12	HIGH	HIGH	LOW	LOW	LOW	HIGH	HIGH	HIGH	LOW	HIGH	LOW	HIGH	HIGH	LOW	HIGH	HIGH	LOW	LOW	LOW	VH
13	HIGH	LOW	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	HIGH	HIGH	LOW	HIGH	HIGH	LOW	LOW	HIGH	LOW	LOW	VH
14	HIGH	LOW	LOW	HIGH	LOW	LOW	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	HIGH	HIGH	LOW	HIGH	LOW	LOW	VH
15	HIGH	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	HIGH	LOW	LOW	LOW	HIGH	HIGH	HIGH	LOW	LOW	VH
16	HIGH	LOW	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	HIGH	LOW	HIGH	LOW	HIGH	LOW	LOW	LOW	VL
17	LOW	LOW	LOW	LOW	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	LOW	HIGH	LOW	LOW	HIGH	LOW	LOW	HIGH	VL
18	HIGH	HIGH	HIGH	HIGH	LOW	LOW	LOW	LOW	LOW	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	L
19	LOW	LOW	LOW	LOW	LOW	HIGH	LOW	LOW	LOW	HIGH	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	L
20	HIGH	LOW	HIGH	LOW	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	LOW	L
21	LOW	LOW	LOW	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	HIGH	HIGH	LOW	HIGH	LOW	LOW	LOW	L
22	HIGH	HIGH	LOW	HIGH	LOW	LOW	HIGH	LOW	HIGH	LOW	LOW	LOW	HIGH	LOW	LOW	LOW	LOW	LOW	HIGH	L
23	LOW	LOW	LOW	HIGH	HIGH	LOW	LOW	LOW	HIGH	HIGH	LOW	HIGH	LOW	LOW	LOW	LOW	HIGH	LOW	LOW	M
24	LOW	HIGH	LOW	LOW	HIGH	LOW	HIGH	LOW	HIGH	HIGH	LOW	HIGH	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	M
25	LOW	HIGH	LOW	LOW	HIGH	LOW	LOW	LOW	HIGH	LOW	LOW	HIGH	HIGH	HIGH	LOW	LOW	LOW	LOW	HIGH	M
26	LOW	LOW	HIGH	LOW	LOW	LOW	LOW	HIGH	HIGH	LOW	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	HIGH	LOW	M
27	LOW	LOW	LOW	HIGH	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	HIGH	LOW	LOW	M
28	LOW	HIGH	HIGH	LOW	LOW	LOW	HIGH	HIGH	HIGH	LOW	HIGH	LOW	LOW	HIGH	LOW	LOW	LOW	HIGH	HIGH	H
29	LOW	LOW	HIGH	HIGH	LOW	LOW	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	LOW	LOW	LOW	HIGH	LOW	HIGH	HIGH	H
30	HIGH	LOW	HIGH	LOW	LOW	LOW	HIGH	HIGH	HIGH	LOW	LOW	HIGH	HIGH	HIGH	LOW	LOW	LOW	HIGH	HIGH	H
31	HIGH	LOW	HIGH	HIGH	LOW	HIGH	HIGH	LOW	HIGH	LOW	LOW	HIGH	LOW	LOW	HIGH	HIGH	LOW	LOW	HIGH	H
32	LOW	HIGH	HIGH	HIGH	LOW	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	H
33	LOW	HIGH	HIGH	HIGH	LOW	HIGH	HIGH	LOW	LOW	LOW	HIGH	HIGH	LOW	HIGH	LOW	HIGH	LOW	HIGH	HIGH	H
34	HIGH	HIGH	HIGH	LOW	LOW	HIGH	LOW	HIGH	HIGH	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	HIGH	LOW	LOW	LOW	H
35	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	LOW	HIGH	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	VH
36	LOW	HIGH	LOW	HIGH	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	VH
37	LOW	HIGH	HIGH	LOW	HIGH	HIGH	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	HIGH	LOW	HIGH	HIGH	LOW	HIGH	HIGH	VH
38	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	LOW	LOW	HIGH	HIGH	HIGH	HIGH	HIGH	LOW	LOW	HIGH	HIGH	HIGH	VH
39	HIGH	LOW	HIGH	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	LOW	HIGH	LOW	LOW	HIGH	LOW	LOW	LOW	HIGH	VL
40	LOW	LOW	LOW	LOW	HIGH	LOW	HIGH	LOW	LOW	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	HIGH	VL
41	LOW	LOW	HIGH	LOW	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	HIGH	LOW	L
42	LOW	HIGH	HIGH	LOW	HIGH	HIGH	LOW	LOW	LOW	HIGH	LOW	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	LOW	L
43	HIGH	LOW	LOW	HIGH	HIGH	LOW	LOW	LOW	LOW	HIGH	LOW	LOW	LOW	LOW	LOW	HIGH	LOW	LOW	LOW	L
44	LOW	LOW	HIGH	LOW	LOW	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	HIGH	LOW	LOW	HIGH	LOW	LOW	HIGH	L
45	LOW	HIGH	LOW	HIGH	LOW	LOW	LOW	LOW	HIGH	HIGH	LOW	HIGH	HIGH	LOW	HIGH	LOW	HIGH	LOW	LOW	L
46	LOW	LOW	HIGH	LOW	HIGH	LOW	LOW	LOW	HIGH	HIGH	LOW	LOW	LOW	LOW	HIGH	LOW	LOW	LOW	LOW	M
47	HIGH	HIGH	LOW	LOW	HIGH	LOW	HIGH	HIGH	HIGH	HIGH	LOW	LOW	HIGH	HIGH	LOW	LOW	LOW	LOW	LOW	M
48	LOW	HIGH	HIGH	LOW	HIGH	LOW	HIGH	LOW	HIGH	LOW	LOW	LOW	HIGH	HIGH	LOW	HIGH	LOW	LOW	LOW	M
49	LOW	LOW	HIGH	LOW	HIGH	LOW	LOW	LOW	HIGH	LOW	LOW	LOW	LOW	LOW	LOW	LOW	HIGH	HIGH	LOW	M
50	HIGH	LOW	LOW	HIGH	HIGH	HIGH	LOW	LOW	LOW	LOW	HIGH	HIGH	LOW	LOW	HIGH	LOW	HIGH	LOW	HIGH	M

3.6 Data Transformation

This is the stage in which the selected data were transformed into forms acceptable to WEKA data mining software. The data file was saved in Comma Separated Value (CSV) file format in Microsoft excel for easy use in WEKA.

3.6.1 Data Training and Testing

The figure 3.4 below shows the flowchart for training and testing the data. The data is selected and converted into Comma Separated Value (CSV) file format and then classified using WEKA and the result is produced.

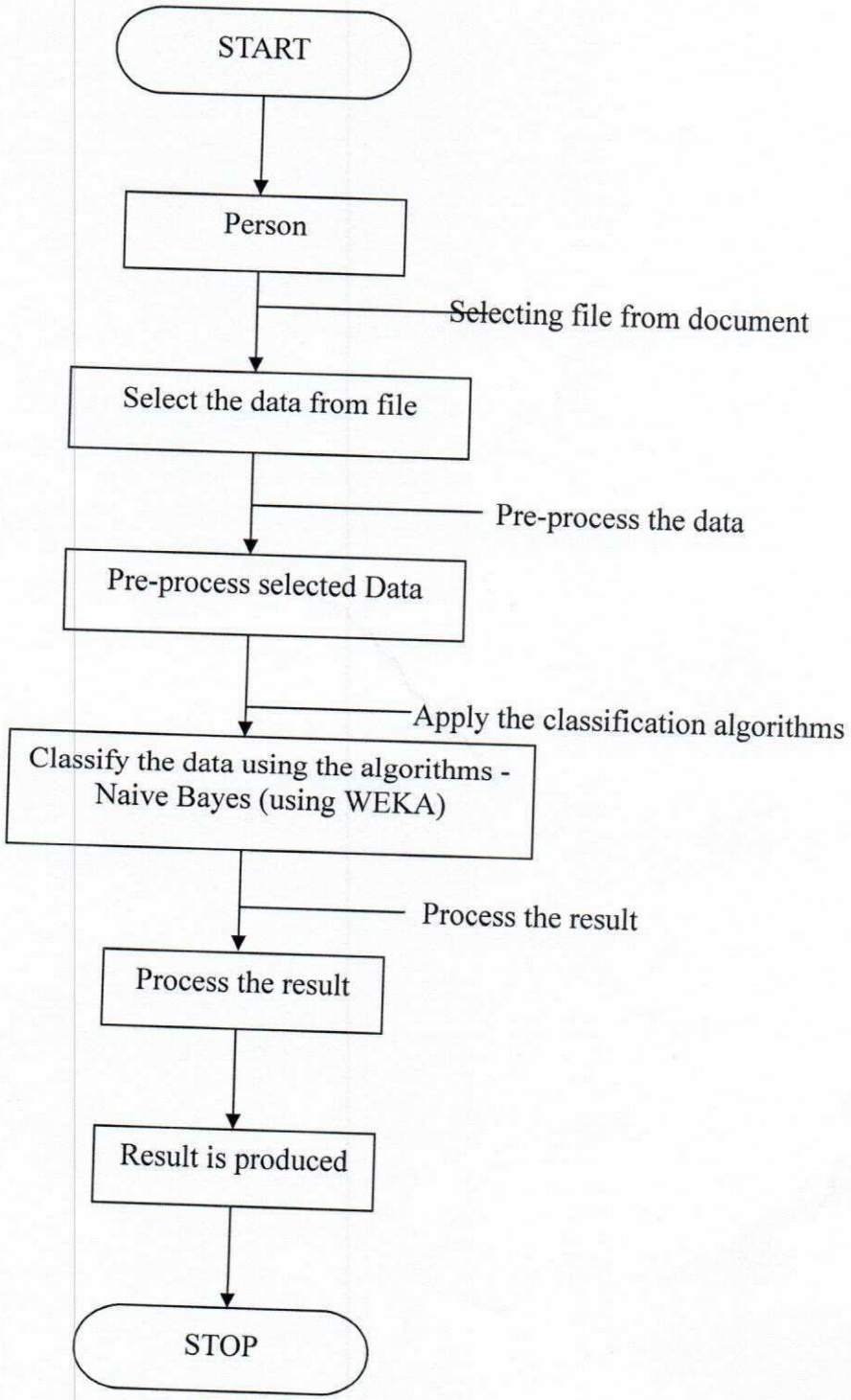

Figure 3.4. Flowchart for training and testing of the data.

3.7 Performance Metrics

The following metrics were used to determine the performance of the model: Time taken to build the model, true positive rate, false positive rate, precision, recall, f-measure, ROC Area, class.

i. **True Positive Rate** is the ability of a test to correctly identify those with the disease.

ii. **False Positive Rate** refers to the probability of falsely rejecting the null hypothesis for a particular test.

iii. **Precision** is the fraction of retrieved instances that are relevant.

iv. **Recall** is the fraction of relevant instances that are retrieved. Both precision and recall are therefore based on an understanding and measure of relevance.

v. **F-measure** is a measure of a test's accuracy. It considers both the precision and the recall of the test to compute.

vi. **Receiver Operating Characteristics:** An alternative method that was used to evaluate classifier performance is the Receiver Operating Characteristic (ROC) analysis. It compares visually the performance of the classifier across the entire range of probabilities. It shows the trade-off between the false-positive rate on the horizontal axis of a graph and the true-positive rate on the vertical axis.

The primary advantage of ROC curve is that they are used to evaluate the performance of a classifier independent of the naturally occurring class distribution or error cost. A good classifier must achieve high TP rate and at the same time less FP rate.

CHAPTER FOUR

IMPLEMENTATION

4.1 Introduction

This chapter discusses the tools used in implementing and it goes further to discuss the result of the training set and testing set of data that are generated.

4.2 System Implementation

In this project work, WEKA data mining software tool was used since this software is open source software that implements a large collection of machine leaning algorithms and is widely used in data mining applications. WEKA contains a collection of visualization tools and algorithms for data analysis and predictive modeling, together with graphical user interfaces.

This dataset was loaded into WEKA explorer. The classify panel enables the user to apply classification to the resulting dataset to estimate the accuracy of the resulting predictive model.

4.2.1 Modules under Training Set

The modules under implementation in the training set are:

a) Pre process module

b) Classify module

c) Visualize module

a) Pre Process Module

The pre process module allows us to choose and modify the data being acted upon and enhances the predictive power of the training set which is shown in figure 4.1.

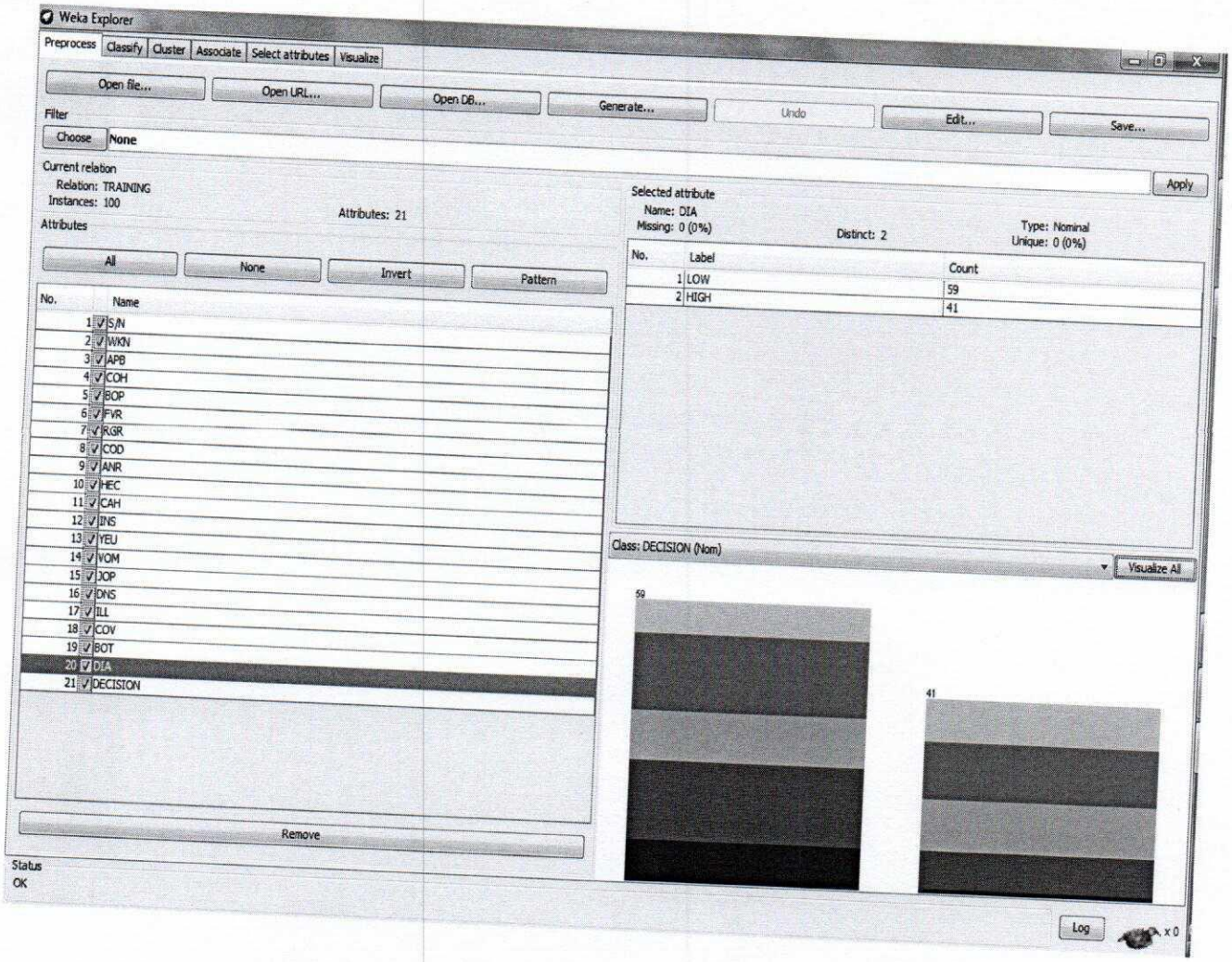

Figure 4.1: Print screen for pre process module on training set.

b) Classify Module

The classify module is used for the selection of algorithm (Naive Bayes) for the training set which is shown in figure 4.2.

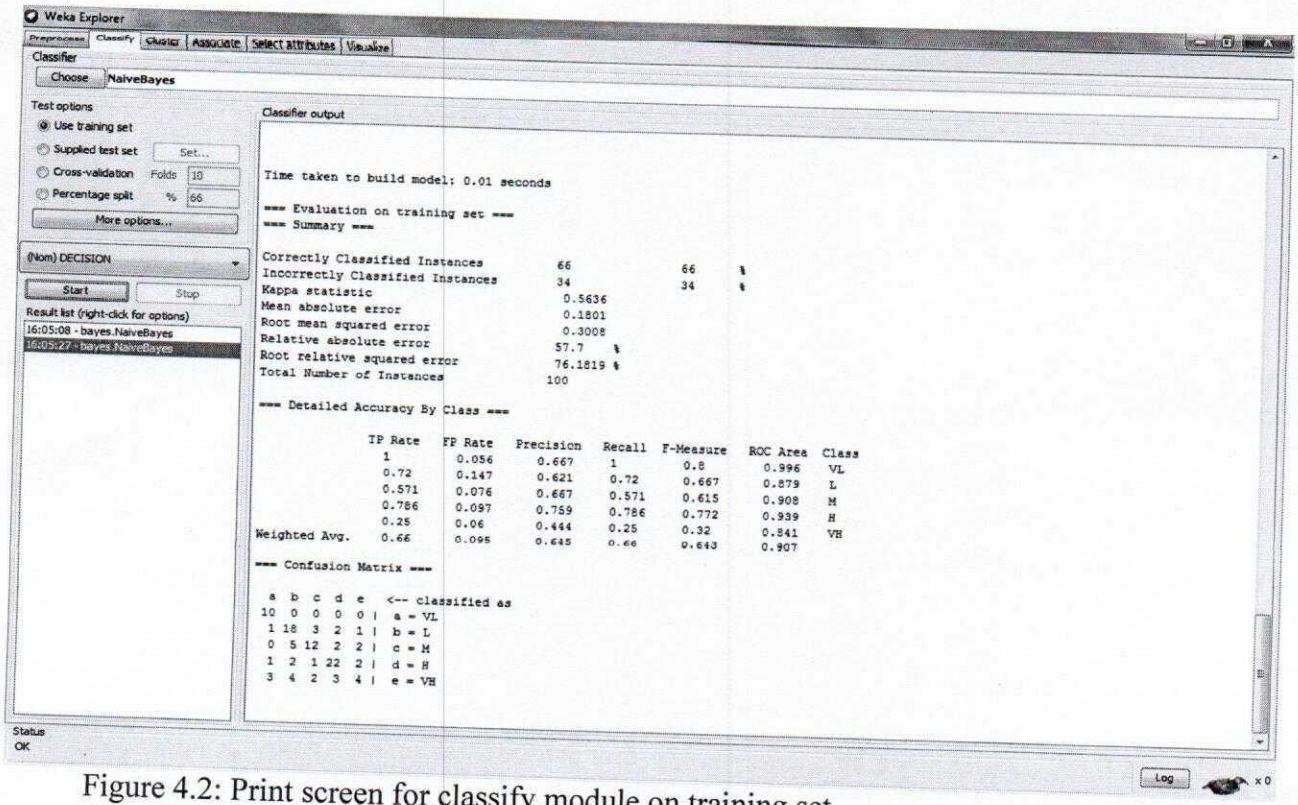

Figure 4.2: Print screen for classify module on training set

c) Visualize Module

It views an interactive 2D plot of the data for training set. The visualize module is seen in figure 4.3.

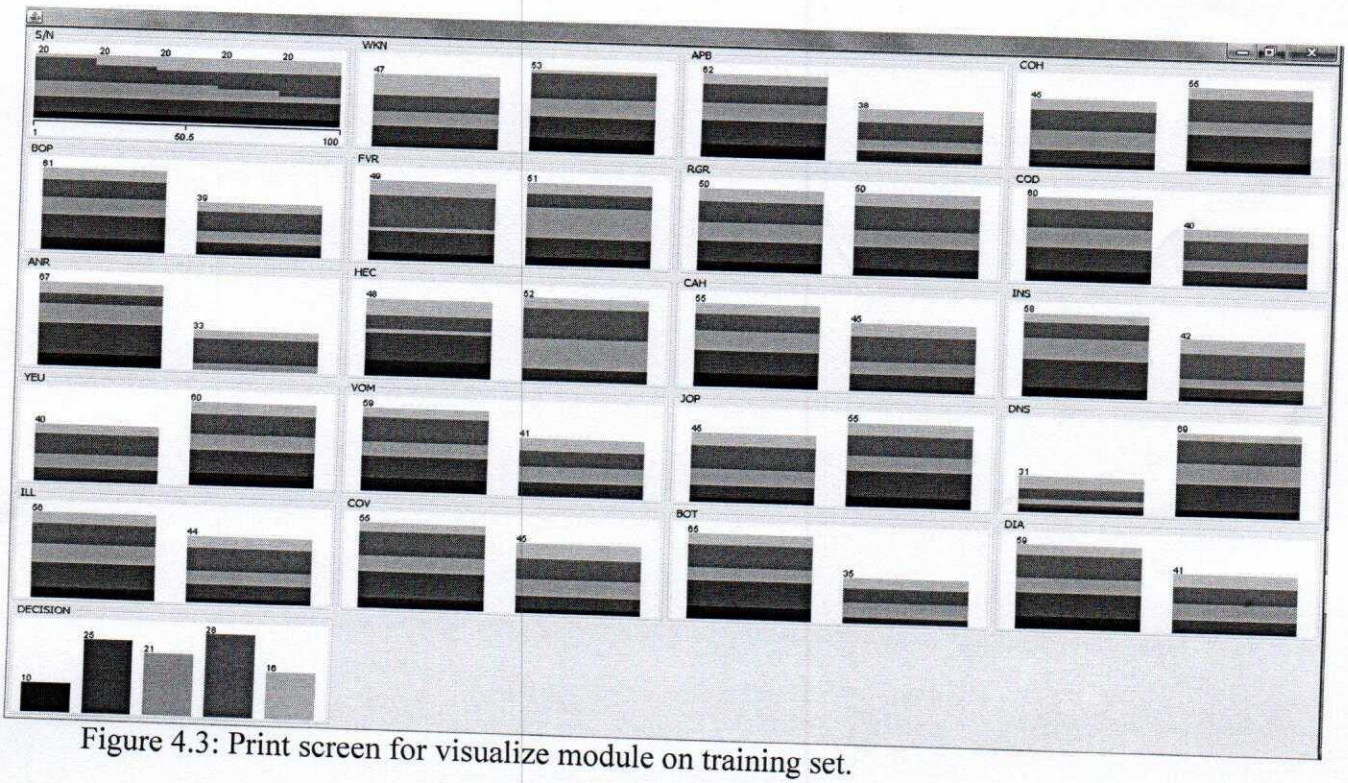

Figure 4.3: Print screen for visualize module on training set.

4.3 Discussion of Results for Training Set

The performance of the system were measured base on two results generated from the system, one on training set and the other on testing set. The first result was obtained from the training set. All the hundred (100) datasets were trained, according to the twenty three rules generated and the confusion matrix of the result is given in the table 4.1 below.

4.3.1 Training Set

Table 4.1: Confusion matrix for the Training Set

Predicted as Actual	Very Low	Low	Moderate	High	Very High
Very Low (10)	10(100%)	0(0.00%)	0(0.00%)	0(0.00%)	0(0.00%)
Low (25)	1(4%)	18(72%)	3(12%)	2(8%)	1(4%)
Moderate (21)	0(0.00%)	5(23.81%)	12(57.14%)	2(9.52%)	2(9.52%)
High (28)	1(3.57%)	2(7.14%)	1(3.57%)	22(78.57%)	2(7.14%)
Very High (16)	3(18.75%)	4(25%)	2(12.5%)	3(18.75%)	4(25%)

TP = Class group correctly classified

TN = Class group wrongly classified

$$\text{Detection Rate} = \frac{TP}{TP + TN} = \frac{66}{66+34} = \frac{66}{100} = 66\%$$

All the ten labels classified as Very Low were correctly predicted which gives 100% prediction. Out of twenty five labels classified as low only Eighteen labels were correctly predicted attaining (72%), one was predicted as Very Low (4%), three was predicted moderate (12%), two was predicted high (8%), while another one was predicted very high (4%). Out of twenty one classified as moderate only twelve were correctly predicted attaining (57.14%), five was predicted as low (23.81%), two was predicted as high (9.52%), while another two was predicted as very high (9.52%). There are twenty eight labels classified as High. only twenty two were correctly predicted attaining (78.57%), one was predicted as very

low (3.57%), two was predicted as low (7.14%), one was predicted as moderate (3.57%), and two was predicted as very high (7.14%). Out of sixteen classified as very high, four were correctly predicted attaining (25%), three were predicted as very low (18.75%), two were predicted as moderate (12.5%), three were predicted as high (18.75%) and four were predicted as very high (25%). The confusion matrix shown in Table 4.1 thus gives 66% Detection Rate.

Table 4.2: Performance of the classifiers

Evaluation Criteria	**Naïve Bayes**
Time taken to build the model	0.02 seconds
Correctly Classified Instances	66
Incorrectly Classified Instances	34
Accuracy	66%

The number of correctly classified instances is often called accuracy or sample accuracy of a model. So Naive Bayes classifier has the accuracy of 66% and it is seen that it takes 0.02 seconds in building the model as shown in table 4.2 above.

Table 4.3: Detailed Accuracy by Class

	TP Rate	**FP Rate**	**Precision**	**Recall**	**F-measure**	**ROC Area**	**Class**
	1	0.056	0.667	1	0.8	0.996	VL
	0.72	0.147	0.621	0.72	0.667	0.879	L
	0.571	0.076	0.667	0.571	0.615	0.908	M
	0.786	0.097	0.759	0.786	0.772	0.939	H
	0.25	0.06	0.444	0.25	0.32	0.841	VH
Weighted Average	0.66	0.095	0.645	0.66	0.643	0.907	

From the table 4.3 above, it is seen that the number of correct positive predictions TP and the number of incorrect negative predictions FN predicts the malaria fever cases better. Also

going by the precision, Recall, F-measure and ROC values, i was able to conclude that the model could be used for easy prediction of malaria fever.

4.4 Modules under Testing Set

The modules under implementation in the testing set are:

a) Pre process testing

b) Classify testing

c) Visualize testing

a) Pre Process Testing

The pre process testing allows us to choose and modify the data being acted upon and enhances the predictive power of the testing set which is shown in figure 4.4.

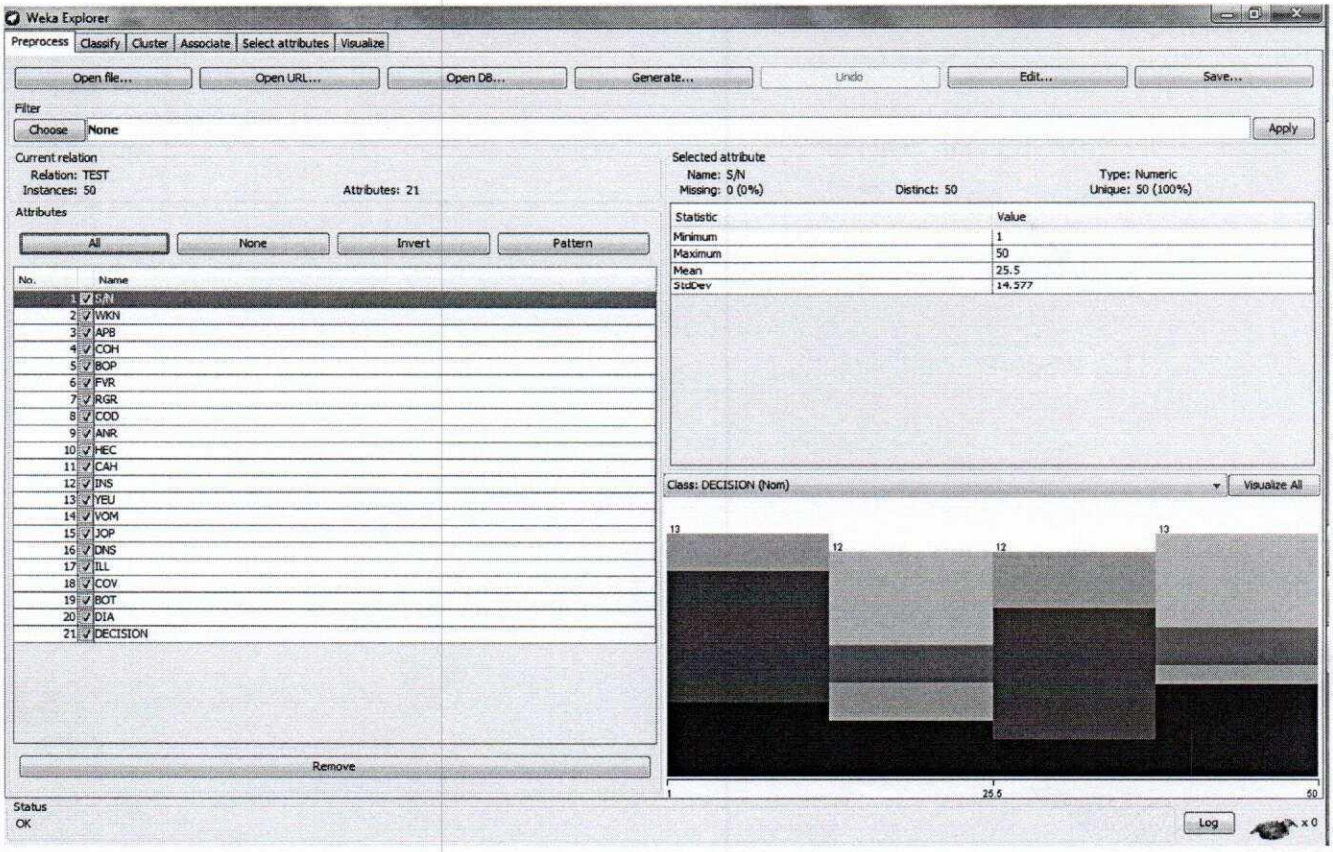

Figure 4.4: Print screen for pre process on testing set

b) Classify Testing

The classify testing is used for the selection of algorithm (Naive Bayes) which is shown in figure 4.5.

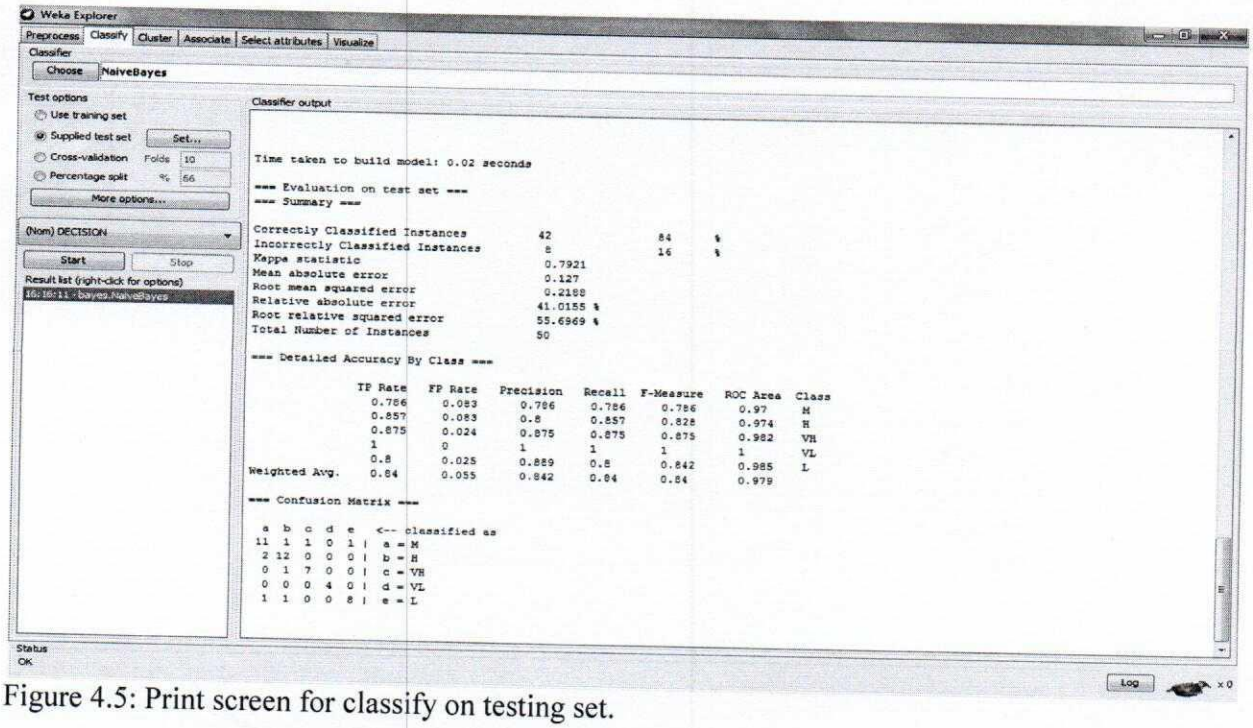

Figure 4.5: Print screen for classify on testing set.

c) Visualize Testing

It views an interactive 2D plot of the data for testing set. The visualize testing is seen in figure 4.6.

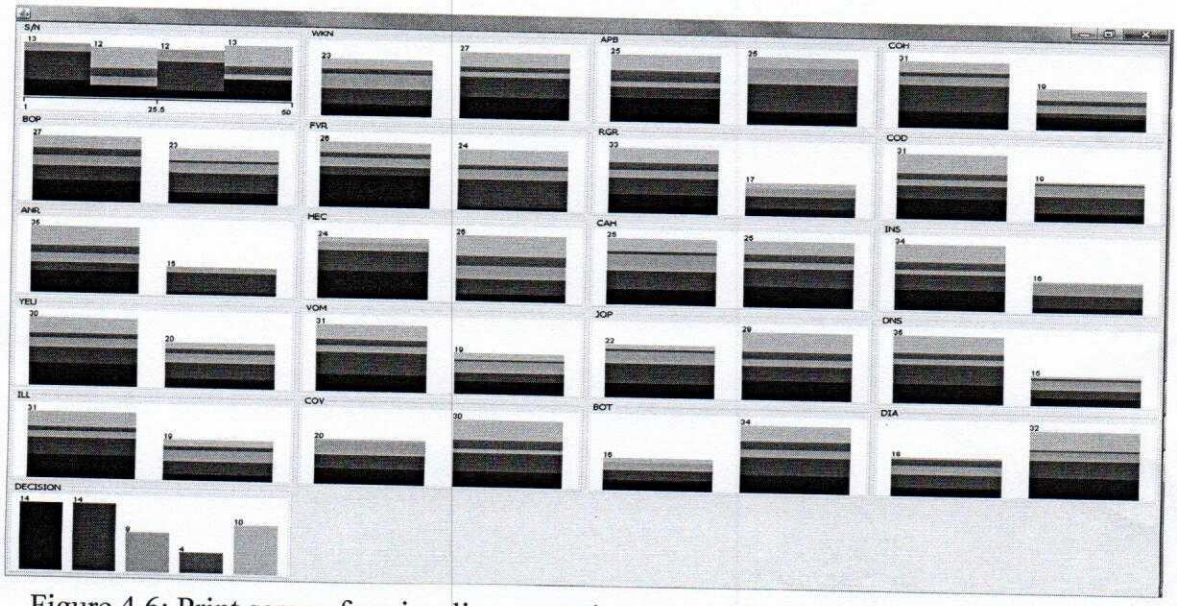

Figure 4.6: Print screen for visualize on testing set.

4.5 Discussion of Results for Testing Set

The second result was obtained from the testing set. All the fifty (50) datasets were tested, according to the twenty three rules generated and the confusion matrix of the result is given in the table 4.4 below.

4.5.1 Testing Set

Table 4.4: Confusion Matrix for the Testing Set.

Predicted as Actual	Very Low	Low	Moderate	High	Very High
Very Low (4)	4(100%)	0(0.00%)	0(0.00%)	0(0.00%)	0(0.00%)
Low (10)	0	8(80%)	1(10%)	1(10%)	0
Moderate 14)	0	1(7.14%)	11(78.57%)	1(7.14%)	1(7.14%)
High (14)	0(0.00%)	0(0.00%)	2(14.29%)	12(85.71)	0(0.00%)
Very High (8)	0(0.00%)	0(0.00%)	0(0.00%)	1(12.5%)	7(87.5%)

TP = Class group correctly classified

TN = Class group wrongly classified

$$\text{Detection Rate} = \frac{TP}{TP + TN} = \frac{42}{42+8} = \frac{42}{50} = 84\%$$

For the testing set, fifty set of data were tested against the twenty three rules. There is only four very low labels and it was correctly predicted, attaining 100% in this case. Out of ten labels classified as Low, only eight were correctly predicted attaining (80%) while one was predicted as Moderate which gives (10%) while another one was predicted as High which gives (10%). Of the fourteen labels classified as moderate, only eleven labels were correctly predicted attaining (78.57%), one was predicted as low (7.14%) while another one was

predicted as high (7.14%) and another one as very high (7.14%). All the fourteen labels classified as high, twelve were correctly predicted (85.71), while two were predicted as moderate attaining (14.29%). In the case of the eight labels classified as Very High, seven were correctly predicted (87.5%), while one was predicted as high which gives (12.5%). The confusion matrix for the testing set shown above in table 4.4 has the Detection Rate of 84%.

Table 4.5: Performance of the Classifiers

Evaluation Criteria	**Naïve Bayes Algorithm**
Time taken to build the model	0.1 seconds
Correctly Classified Instances	42
Incorrectly Classified Instances	8
Accuracy	84%

The number of correctly classified instances is often called accuracy or sample accuracy of a model. So Naive Bayes classifier has the accuracy of 84% and it is seen that it takes 0.1 seconds in building the model as shown in table 4.5 above.

Table 4.6: Detailed Accuracy by Class

	TP Rate	**FP Rate**	**Precision**	**Recall**	**F-measure**	**ROC Area**	**Class**
	1	0	1	1	1	1	VL
	0.8	0.025	0.889	0.8	0.842	0.985	L
	0.786	0.083	0.786	0.786	0.786	0.97	M
	0.857	0.083	0.8	0.857	0.828	0.974	H
	0.875	0.024	0.875	0.875	0.875	0.982	VH
Weighted Average	0.84	0.055	0.842	0.84	0.84	0.979	

From the table 4.6 above, it is seen that the number of correct positive predictions TP and the number of incorrect negative predictions FN predicts the malaria fever cases better. Also going by the precision, Recall, F-measure and ROC values, i was able to conclude that the model could be used for easy prediction of malaria fever.

Based on the result obtained during the implementation of malaria fever diagnosis using Naïve Bayes algorithm and using WEKA software, the success rate is considered good. From the Tables 4.1 and 4.4 above, the experiment shows that the model built with Naive Bayes algorithm for the training set, with all attributes correctly classified (predicted the correct outcome) 66 (66%) instances while 34 (34%) of the instances were classified incorrectly. For the testing set the model built with Naive Bayes algorithm also with all attributes correctly classified (predicted the correct outcome) 42 (84%) instances while 8 (16%) of the instances were classified incorrectly.

CHAPTER FIVE

CONCLUSION AND RECOMMENDATION

5.1 Conclusions

The Bayesian based approach to malaria fever diagnosis system has been designed to proffer good health education and assist doctors. This will in the end help to reduce the number of death rates and increase standard of living. With the detection rate of 66% for the training set and 84% for the testing set, we conclude that this software (WEKA) helps the doctors to clear their confusion when predicting diseases with similar symptoms and helps to take better decision. This study showed that data mining techniques can be used efficiently to model and predict malaria fever cases. The outcome of this study can be used as an assistant tool by medical experts to help them make more consistent diagnosis of malaria fever.

5.2 Recommendation

This study has indicated that data mining techniques can be applied in the diagnosis of malaria fever and the resulting models of this study are worthy of clinical testing. To improve the classification accuracy of the models further researches should be conducted using different classification algorithms and other data mining techniques such as, Decision tree, genetic algorithm etc. can be used for diagnosis. Finally, expanded data set with more distinctive attributes to get more accurate results can also be used to carry out diagnosis to improve the classification accuracy.

REFERENCES

Abeku T.A., de Vlas S.J., Borsboom G.M., Tadege A., Gebreyesus Y., Gebreyohannes H., Alamirew D., Seifu A., Nagelkerke J.D., and Habbema D.F. (2004). Effects of Meteorological Factors on Epidemic Malaria in Ethiopia: A Statistical ModellingApproach Based On Theoretical Reasoning. Parasitology, vol. 1(28), pp. 585-593.

Abreu Z.P., and Van Gemund A. J. C. (2007). On the Accuracy of Spectrum-Based Fault Localization.

Abreu Zoeteweij P., and Van Gemund A. J. C. (2008a). A Dynamic Modelling Approach to Software Multiple-Fault Localization.

Adekoya A.F., Akinwale A.T., and Oke O.E. (2008). A Medical Expert System for Managing Tropical Diseases. Proceedings of the Third Conference on Science and National Development, pp.74-86.

Adetunmbi A.O., Oguntimilehin A., and Falaki S.O (2012). Computer Science andTelecommunications No.1 (33).

Afrane Y.A., Lawson B.W., Githeko A.K., and Yan G. (2005). Effects of microclimaticchanges caused by land use and land cover on duration of gonotrophic cycles ofAnopheles ga mbiae (Diptera: Culicidae) in western Kenya highlands. *Journal ofMedical Entomology*, vol., 42(6), pp. 974–980.

Asenso Okyere W.K., and Dzator J.A (1997). Household cost of seeking malaria care: ARetrospective Study of two Districts in Ghana. Social Science and Medicine No5, pp. 659 – 667.

Aultman, K.S., Gottlieb M., and Giovanni M.Y. (2002). Fauci, AS. Anopheles gambiae genome, completing the malaria triad. Science, 298, 13.

Balls M.J., Bodker R., Thomas C.J., Kisinza W., Msangeni H.A., and Lindsay S.W. (2004). Effect of Topography on the Risk of Malaria Infection in the Usambara Mountains, Tanzania. *Trans R Soc Trop Med Hyg*, vol. 9(8), pp. 400-408.

Bozdech Z., Llinas M., Pulliam B.L., Wong E.D., Zhu J., and DeRisi J.L. (2003). The Transcript of the Intra Erythrocytic Developmental Cycle of Plasmodium Falciparum. *PLoS Biol*, vol. 1(1).

Caruana R., and Niculescu-Mizil A. (2006). An empirical comparison of supervised learning algorithms. In Proceedings of the 23rd International Conference on Machine Learning.

Chaloner K., and Verdinelli I. (1995). Bayesian Experimental Design: a review. Statistical Science, vol 10(3), pp. 273-304.

Chong E.K., Kreucher C.M., and Hero A.O. (2009). Partially Observable Markov Decision Process Approximations for Adaptive Sensing. Discrete Event Dynamic Systems, vol. 19(3), pp. 377–422.

De Kleer J., and Williams B.C (1987) Diagnosing Multiple Faults, Artificial Intelligence pp. 97 - 130

De Kleer J., (2006). Getting the probabilities right for measurement selection.

De Kleer J., (2007). Diagnosing Intermittent Faults.

Djam X.Y., and Wajiga, G.M. (2012). A Novel Diagnostic Framework: The Application of Soft Computing Technology. The Pacific Journal of Science and Technology, vol 13. (1), pp. 261-272.

Domingos P., and Pazzani M. (1997). on the optimality of the simple Bayesian Classifier under zero-one loss. Machine Learning, vol. 29(2–3), pp. 103-130.

Eleni-Maria T., Stylianos K., Christos K., and Christina P. (2010). Innovative Data MiningApproaches for Outcome Prediction of Trauma Patients. J. Biomedical Science and Engineering, vol. 1 (3), pp. 791-798.

Feldman A., Provan G., and Van Gemund A. J. C. (2008). Computing Minimal Diagnoses by Greedy Stochastic Search. In *Proc. AAAI.*

Frank E., Trigg L., Holmes G., and Witten I.H. (2000) Technical Note: Naïve Bayes for classification. Machine Learning. Vol. 41(1), pp.5 – 25.

Gallup J., and Sachs J. (2001). The Economic Burden of Malaria. *Am. J. Trop. Med. Hyg,* vol. 64(1, 2), pp. 85 – 96

Githeko A.K., Lindsay S.W., Confalonieri U.E., and Patz J.A. (2000). Climate Change and Vector-Borne Diseases: A Regional Analysis. World Health Organization, *Bulletin of WHO*, vol. 7(8), pp. 1136–1147.

Golovin D., Krause A., and Ray D. (2010). Near-optimal Bayesian Active Learning with Noisy Observations.

Grover-Kopec E., Kawano M., Klaver R.W., Blumenthal B., Ceccato P., and Connor S.J. (2005). An Online Operational Rainfall-Monitoring Resource for Epidemic

MalariaEarly Warning Systems in Africa. *Malaria Journal*, vol. 4, pp. 6.

Guerin P.J., Olliaro P., and Nosten F. (2002). Malaria: Current Status of Control, Diagnosis, Treatment, and A Proposed Agenda for Research and Development, the Lancet Infectious Diseases, vol. 2 (9), pp. 564-573.

Hirschfeld D., (2010). Small change in forest cover can double malaria rate, Science and Development Network Series.

Imianvan A.A., Anosike U.F., and Obi J. C. (2011). An Expert System for the Intelligent Diagnosis of HIV Using Fuzzy Cluster Means Algorithm. Global Journal of Computer Science and Technology Vol. 11(12), [Online ISSN: 0975-4172 & Print ISSN: 0975-4350].

Jackson C.M., Laura J., Cathy F., Abigail C., and Kimberly F.S. (2010). Modelling The Effect of Climate Change on Prevalence of Malaria in Western Africa .*Statistical Neerlandica* Vol. 64(4), pp. 388–400.

Kathleen W., (2002). Activity Report 108: A review of control methods for African Malaria Vectors by Environmental Health Project, Washington, USA.

Karpagavalli S., Jamuna K.S., and Vijaya M.S. (2009). Machine learning approach for preoperative anaesthetic risk prediction. International Journal of Recent Trends in Engineering.

Langley P., Iba W., and Thompson K. (1992). An analysis of Bayesian classifiers. Proceedings of the 10th National Conference on Artificial Intelligence, San Jose, CA Menlo Park, CA: AAAI Press. pp. 223–228.

Lindsay S.W., Bodker R., Malima R., Msangeni H.A., and Kisinza W. (2000). Effect of 1997-98 El Nino on highland Malaria in Tanzania. Lancet, vol. 3(55), pp. 989-90.

Lindsay S.W., and Martens W.J.M. (1998). Malaria in the African highlands: past, present and future. Bulletin of the World Health Organization vol. 76 (1), pp. 33-45.

MARA/ARMA. (1998). Towards an Atlas of Malaria risk in Africa. First technical report of the MARA/ARMA collaboration. MARA/ARMA initiative, Durban, South Africa (available at www.arma.org.za).

Martens W.J., Niessen L.W., Rotmans J., Jetten T.H., and McMichael A.J. (1995). Potential Impact of Global Climate Change on Malaria Risk. Environ Health Perspect, vol. 103(5), pp. 458-64.

Mendis K., Sina B.J., Marchesini P., and Carter R. (2001). The neglected burden of Plasmodium Vivax Malaria. Am J Trop Med Hyg, vol. 64(1-2), pp. 97-106.

Minakawa N., Sonye G., Mogi M., Githeko A., and Yan G. (2002). The Effects of Climatic Factors on the Distribution and Abundance of Malaria Vectors in Kenya. *J Med Entomol,* vol. *3*(9), pp. 833–841.

Ministry of Health Report (MOH). (2003). UN. United Nations Technical Health and Housing Report.

Munga S., Laith Y., Emmanuel M., Guofa Z., Tom O., Noboru M., Andrew G., and Guiyun Y. (2009). Land Use and Land Cover Changes and Spatiotemporal Dynamics of Anopheline Larval Habitats during a Four-Year Period in a Highland Community of Africa, Am. J. Trop. Med. Hyg, vol. 81(6), pp. 1079-1084.

Nicholas E.R., Charles J.P., David M.R., and Adriano G.D (2010). Automated image processing method for the Diagnosis and Classification of Malaria on Thin Blood, http://www.springerlink.com/content/wgh485h218379005, retrieved on 29/07/2015.

Obot O.U., and Uzoka F-M.E. (2008), Fuzzy-rule based Framework for the Management of Tropical Disease, Int. J. Medical Engineering and Informatics. Vol.1. No. 1. http://inderscience.metapress.com/media/acecbdaysq7uyvbwdp4u/contributions15/6/4/0/ 564023q746568m75_html/fulltext. Retrieved 20/06/2015.

Olabiyisi S.O., and Omidiora E.O. (2011). A Decision Support System Model for Diagnosing Tropical Diseases Using Fuzzy Logic. African Journal of Computing and ICT.

Patz J.A., Graczyk T.K., Geller N., and Vittor A.Y. (2000). Effects of Environment Change on Emerging Parasitic Diseases. Int J Parasitol, vol. 30 (1), pp.395-405.

Reid C., (2000). Implication of Climate Change on Malaria in Karnataka-India. Senior Honors Thesis in Environmental Science - Center for Environmental Studies, Brown University.

Sharma V.P., Dhiman R.C., Ansari M.A., Nagpal BN., Srivastava A., Manavalan P., Adiga S., Radhakrishnan K., and Chandrashekhar M.G. (1996). Study on the feasibility of delineating mosquitogenic conditions in and around Delhi using remote sensing satellite data. Indian J Malariol, vol. 33, pp. 107.

Sharma V.P., Prasittisuk C., and Kondrashin A.V. (1991). Magnitude of Forest Related Malaria in the WHO Southeast Region. In: Sharma VP, Kondrashin AV. editors. Forest Malaria in Southeast Asia. Proceedings of an informal consultative meeting WHO/MRC 1991 Feb 18-22, pp. 29-53.

Shillu J., Tewolde G., Solomon M., Helen F., Mehari Z., Charles M., John G., Robert N., Eugene B., and John C.B. (2003). High Seasonal Variation in Entomological Inoculation Rates in Eritheria, A Semi-Arid Region Of Unstable Malaria In Africa., *Am. J. Trop. Med. Hyg*, vol. 69(6), pp. 607-613.

Smith T., Charlwood J.D., Takken W., Tanner M., and Spiegelhalter D.J. (1995). Mapping densities of malaria vectors within a single village. Acta Tropica, vol. 5(9), pp. 1–18. Snow R.W., and Marsh K. (2002). The Consequences of Reducing Transmission of Plasmodium Falciparum in Africa. Adv Parasitol, 52, pp. 235-264.

Tonnang E.Z., Richard Y.M.K., and Pius Z.Y. (2010). Predicting and Mapping Malaria under Climate Change Scenarios: The Potential Redistribution of Malaria Vectors in Africa, *Malar J.* vol. 9, pp. 111.

Uzoka F-M.E., Obot O., and Barker K (2009), A performance Comparison of Fuzzy Logic and AHP as Engines for the Development of Intelligent Medical Diagnosis Systems, Nigeria Computer Society 23rd National Conference pp 135- 150.

Wald A., (1947) Sequential Analysis. John Wiley and Sons.

Weidong G.U., and Robert J. (2009). Novak, Agent-Based Modelling of Mosquito Foraging Behaviour for Malaria Control, Trans R Soc Trop Med Hyg. November, vol. 103(11), pp.1105.

WHO/AFRO (2001). Progress in Rolling Back Malaria in the African Region. Malaria, Liaison Bulletin of the Malaria Programme.

WEKA - University of Waikato: Department of Computer Science, Documentation:http://www.cs.waikato.ac.nz/ml/WEKA/index_documentation.html.

WHO/UNICEF: World Malaria Report (2005), World Health Organization/UNICEF. Report series: WHO/HTM/MAL/2005.1102.

Whitty C.J., and Allan R. (2004). Averting a Malaria Disaster in Africa –Where does the Buck-up stop? in the Bulletin of the World Health Organisation, pp.3-4.

Worrall E.S., Basu X.X., and Hanson K. (2003). The Relationship Between Socio-Economic Status And Malaria: A Review Of Literature: Paper for the WHO/TDR ScientificWorking Group on Malaria, Geneva, Switzerland, pp. 24–27.

World Health Organization Malaria Fact sheet N°94 April (2010).

World Health Organization. (2011). Guidelines for the Treatment of Malaria, 2nd Edition, http://whqlibdoc.who.int/publications/2010, Retrieved on 06/07/2015.

Zeon T.F., Priyank T., Abhinandan P., and Priyal T. (2013). DOCAID: Predictive healthcare analytics using naive Bayes classification. Second Student Research Symposium (SRS), International Conference on Advances in Computing, Communications and Informatics (ICACCI'13).

Zhou X., Liu K.Y., and Wong S.T.C (2004). Cancer Classification and Prediction using logistic regression with Bayesian gene selection. Journal of Biomedical Informatics. Elsevier, vol. 3(7), pp. 249- 259.