DEVELOPMENT OF PREDICTIVE MODEL FOR SURVIVAL OF PAEDIATRIC HIV/AIDS PATIENTS IN SOUTH WESTERN NIGERIA USING DATA MINING TECHNIQUES

BY

OLUTOLA AGBELUSI

B.Sc. (Hons), Computer Science (EKSU); M.Tech. Computer Science (FUTA)

A THESIS SUBMITTED TO THE DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING, FACULTY OF TECHNOLOGY, OBAFEMI AWOLOWO UNIVERSITY, ILE-IFE, NIGERIA

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF MASTER OF PHILOSOPHY IN COMPUTER SCIENCE

2014

ABSTRACT

The study identified survival variables for HIV/AIDS paediatric patients, developed predictive models for determining the survival of the patients who were receiving antiretroviral drug in the Southwestern Nigeria based on identified variables, compared the developed models and validate the models with the historical data. This was with a view to choosing the most efficient model for the prediction of the survival of Peadiatric HIV/AIDS patients

Interviews were conducted with the virologists and Peadiatrician at two health institutions from the study area in order to identify survival variables for HIV/AIDS Paediatric patients. Paediatric HIV/AIDS patients' data (216) were also collected from two health institutions, preprocessed and the 10-fold cross validation technique was used to partition the datasets into training and testing data. Predictive models were developed using three (3) supervised learning techniques (Naïve Bayes' classifiers, decision trees and the multi-layer perception (MLP)) and the Waikato Environment for Knowledge Analysis (WEKA) was used to simulate the models in which CD4 count, Viral Load, Opportunistic infections and Nutritional status were used as the independent variables for the prediction.

The result showed that all the three techniques (Naïve Bayes' classifiers, decision trees and the multi-layer perception (MLP)) were suitable in carrying out the task of forecasting the survival of Paediatric HIV/AIDS patients so that each patient can know their status at every point in time. The decision trees model has an accuracy of 99.07% (214 correct classifications out of 216), 0.0183 mean absolute error rate, 0.0962 root mean square error and 3.69% relative absolute error. The Receiver Operating Characteristics (ROC) area for the model was also 0.993 showing that the level of bias was very low (0.007), Naïve Baye's model has an accuracy of 81.02% (175 correct classifications out of 216), the mean absolute error rate was 0.2025, 0.2920 for the root

mean square error and 40.92% for the relative absolute error. The ROC area for the model was also 0.993 showing that the level of bias was very low (0.007) and multilayer perception model has an accuracy of 99.07% (214 correct classifications out of 216), the mean absolute error rate was 0.022, 0.0962 for the root mean square error and 4.48% for the relative absolute error. The ROC area for the model was also 0.992 showing that the level of bias was also very low (0.008). The result of the three models showed that the decision tree model was the most efficient of all the three models with an accuracy of 99.07% (214 correct classifications out of 216). The validation was done by comparing the three developed models with the historical data from the two selected health institutions which has a catchment of patients from the other parts of the South western Nigeria.

The study concluded that the decision trees technique is viable in predicting survival among HIV/AIDS patients in Southwestern Nigeria and that the prediction of poor prognosis for survival may mean that patients, relations and care givers will need to work assiduously and see how the negative survival can be changed.

CHAPTER ONE

INTRODUCTION

1.1   Synopsis

Epidemic diseases have highly destructive effects around the world and these diseases have affected both developed and developing nations. Disease epidemics are common in developing nations especially in Sub Saharan Africa in which Human Imunodeficiency Virus /Aquired Imunodeficiency Disease Syndrone (HIV/AIDS) is the most serious of all   (Idowu *et al.*, 2009).

HIV is one of the world's most serious health and development challenges (Henry, 2013). It is a type of virus called a <u>retrovirus</u> which infects human when it comes in contact with tissues such as those that line the vagina, anal area, mouth, eyes or through a break in the skin (Eric *et al.*, 2012), while Acquired Immunodeficiency Syndrome (AIDS) is the advanced stage of the retroviral infection that swept through sub-Saharan Africa with venom (Hoa, 2011; Joyce, 2014; Idowu, 2012).

Globally, HIV continues to be a very serious health issue facing the world (Ojunga *et al.*, 2014). About 34 million people were living with HIV at the end of 2011 and an estimated 0.8% of adults aged 15-49 years worldwide are living with the virus, although the burden of the epidemic continues to vary considerably between countries and regions (UNAIDS, 2012a).

In Sub-Saharan Africa, roughly 25 million people were living with HIV in 2012, accounting for nearly 70 percent of the global total of infected patients. The epidemic has both social and economic consequences, not only in the health sector but also in education, industry and the wider economy (WHO, 2012b; Shearer, 2000).

The epidemic has heavy impact on education, school attendance drops as children become sick or return home to look after affected family members (Picat, 2013). Moreover, Sub-Saharan

OBAFEMI AWOLOWO

Africa remains most severely affected with nearly 1 in every 20 adults (4.9%) living with HIV which accounts for 69% of the people living with HIV worldwide. Although, the regional prevalence of HIV infection is nearly 25 times higher in sub-Saharan Africa than in Asia (Shearer, 2000).

In 2012, about 5 million people were living with HIV in South, South-East and East Asia combined and sub-Saharan Africa region is the most heavily affected region followed by the Caribbean, Eastern Europe and Central Asia, where 1.0% of adults were living with HIV as at 2011 (Henry, 2013). In addition, Nigeria is the most populous nation in Africa with an estimated population of over 160 million people (National Population Commission, 2014). It is estimated that over 4% people die yearly in Nigeria as a result of HIV/AIDS (Nigerian Bulletin, 2014).

At present, there is no cure for HIV but it is being managed with antiretroviral drugs (ARV). Highly Active Antiretroviral drugs (HAART) is the optimal combination of ARV (Rosma *et al*., 2012; Kama and Prem, 2013). ARV do not kill the virus but slow down the growth of the virus (Ojunga *et al*., 2014 ; WHO, 2012a). Antiretroviral therapy (ART) and highly antiretroviral therapy(HAART) are the mechanisms for treating retroviral infections with drugs.

For more information, please contact **ir-help@oauife.edu.ng**

© Obafemi Awolowo University, Ile-Ife, Nigeria
For more information contact **ir-help@oauife.edu.ng**   Page 2