PREDICTIVE MODEL TO DETECT FIRST-LINE ANTIRETROVIRAL THERAPY FAILURE AMONG HIV/AIDS PATIENTS IN ZEWDITU HOSPITAL, ADDIS ABABA

A Thesis Presented

by

Helina Assefa

to

The faculty of Informatics

of

St. Mary's University

In partial Fulfillment of the Requirements

for the Degree Of Master of Science

in

Computer Science

January, 2022

Addis Ababa

ACCEPTANCE

Prediction of Treatment Failure among Adults Receiving First-Line Antiretroviral Therapy from Zewditu Hospital in Addis Ababa

By

Helina Assefa

Accepted by the Faculty of Informatics, St. Mary's University, in partial fulfillment of the requirements for the degree of Master of Science in Computer Science

Thesis Examination Committee:

_____

Internal Examiner

Name	Signature	Date
_____	_____	_____

_____

External Examiner

Name	Signature	Date
_____	_____	_____

_____

Dean, Faculty of Informatics

Name	Signature	Date
Dr. Getahun Semeon	_____	_____

28 January 2022

DECLARATION

I, the undersigned, declare that this thesis work is my original work, has not been presented for a degree in this or any other universities, and all sources of materials used for the thesis work have been duly acknowledged.

Helina Assefa

_____

Signature

Addis Ababa, Ethiopia

This thesis has been submitted for examination with my approval as advisor.

Dr. Getahun Semeon

_____

Signature

Addis Ababa, Ethiopia

28 January 2022

Acknowledgment

Successful completion of this thesis was possible with support and sincere advice I received from many people. I am indebted to extend, here, my gratitude to their generosity.

First and foremost, I would like to express the deepest gratitude to my advisor Dr. Getahun Semeon, for his invaluable and challenging comments that constructively helped me throughout conducting this study and his exceptional patience follow-up to wait until I come up with something tangible.

I also sincerely thank Sister Alemtsehay Abebe, from Ministry of Health, for her unreserved professional support on HIV/AIDS patient care and ART administration.

Finally, and most importantly, I thank my family for their love, patience, encouragement, understanding, and unconditional support.

TABLE OF CONTENTS

List of Acronyms

ART	Anti-Retrieval Therapy
CD4	Cluster of Differentiation 4
HIV	Human Immunodeficiency Virus
AIDS	Acquired Immunodeficiency Syndrome Disease
SVM	Support Vector Machine8
UN	United Nations
WHO	World Health Organization

List of Figures

List of Tables

Abstract

This study utilizes expert consultation to develop machine learning based predictive model that detects clients who are at high-risk of treatment failure among those who are receiving first-line ARV therapy. The study uses retrospective cross-sectional data of clients who are at least 6 months on ART when data was collected from Zewditu Hospital. The study has followed the Cio data mining model. The study has conducted two main procedures for model development; cluster modeling and classification modeling. The cluster modeling was conducted by using the K-mean algorithm and classification modeling was conducted by implementing decision tree (J48), NaiveBayes, SVM and random forest algorithms The experimentation results show that all the algorithms were the same in terms of accuracy (98.998%), precision (0.990), recall (1.00), and F1-score (0.995). They differ in the time taken to build the classification model. J48 and Naïve Bayes algorithms are have better time efficiency. Accordingly, the J48 and Naïve Bayes algorithms were found the best algorithms to develop ART treatment detection model for the data considered in this study.

Key Words: First-liner ART, ART failure detection, Clustering, Classification, WEKA, Cios Model, Zewditu Hospital.

1. INTRODUCTION

1.1.Background of the Study

The Human Immunodeficiency Virus (HIV) and the Acquired Immunodeficiency Syndrome (AIDS) disease still remains a global public health issue [1]–[3]. As of 2020, HIV/AIDS has infected over 79.3 million people on all parts of the planet, leading to 36.3 million deaths globally [4]. HIV care is currently periodic acute care [5]. So effectiveness of HIV treatment and correctly diagnosing treatment failure in a timely manner is critical for preventing HIV-related disease and death and transmission of the virus [6], [7].

Antiretroviral therapy (ART) is the state-of-the-art HIV treatment [8]. ART has significantly reduced HIV-related morbidity and mortality [8] by suppressing patients' HIV viral level in their serum to undetectable levels, effectively eliminating the risk of transmitting HIV to others and live normal lifespans [6], [8]. For most people with HIV who have never taken ARVs before, first-line HIV treatment regimens include antiretroviral (ARV) drugs that are safe, effective, and convenient recommended as best [9]. For the effectiveness of ART, patients have to take medication every day and regularly see a doctor for the entirety of their life [6]. However, due to drug resistance, opportunistic infections, individual circumstances and behaviors patients [2], [3], [10], [11] treatment failure is a major obstacle faced by therapist in treating HIV infected patients [10]. Treatment failure is frequently linked to mortality, which is costly locally, and the development of drug resistant viral mutations [10], [12], [13].

When a first-line ART fails, it is necessary to conduct resistance testing in order to carry out HIV effective therapy [10], so that patients can be switched to second-line regimen in order to have a sustained viral suppression [13], [14]. Incorrect diagnosis of treatment failure can result in continued viral replication, deterioration of patient's immune system, extra clinical costs such as treatment of opportunistic infections, increased risk of HIV transmission, selection of resistant strains, and death [7].

In order to minimize these undesired consequences, clinical prediction models that utilize patient-level evidence can be applied to healthcare decision-making [2], [15]. These models can be based on statistical analysis or machine learning (ML). Despite dominant application of statistical analysis, a growing trend in application of ML in HIV research

was observed [3]. Machine learning (ML) methods have the ability to identify and discover patterns in complex datasets and predict future outcomes of HIV treatment with good predictive performance [3], [5], [16], [17]. Though machine learning methods were used to predict CD4 count changes and their predictors among patients on ART [18], [19], eligibility of patients for ART [5], [20]–[22], to predict risk of ART termination among patients on ART [6], [23], to predict ART-induced toxicity of HIV patients receiving ART based on [8], to predict patient response to ART [10], [11], [17], to predict pediatric survival among HIV/AIDS patients [11], [24], [25], these works were limited in that their reference was WHO 2010 or 2013 guidelines.

In Ethiopia where there is a significant HIV pandemic status [26] with an increasing first-line ART failure [27], [28]. Moreover, the second-line ARTs are out pacing first-line ARTs [29] in this country. Despite these facts, researchers that used data from this country employ only statistical methods to identify factors for first-line ART treatment failure [7], [27], [29], [30].

Predictive modeling can improve targeting of interventions through differentiated models of care, increasing cost-effectiveness and improving patient outcomes [31] and establish various prioritization criteria for initiating therapy based on achieving maximum survival and clinical benefits for patients [22]. This helps HIV/AIDS therapists make a proactive prediction of first-line treatment failure and make effective treatment intervention.

This study is conducted to develop a predictive modeling based on WHO 2021 consolidated guideline [32] and consultation with HIV/AIDS expert from Ministry of Health.

1.2. Statement of the Problem

UNAIDS/WHO in 2019 estimated that in Ethiopia about 690, 000 people live with HIV and 543,000 people know their status by 2019 [33]. While there is no cure for HIV/AIDS, it can be treated with antiretroviral therapy (ART) medicines [34]. First-line ART failure among ART users in Ethiopia was significant [27], [28]. In 2018 in Ethiopia, the magnitude of ART treatment failure was 15.9% and currently the number of patients receiving second line antiretroviral therapy (ART) is more increasing than those taking first line ART [29].

As the ART uptake increases, the emergence of resistant viruses resulting in treatment failure is inevitable [35] and should be anticipated proactively [13]. Therefore, early detection of treatment failure is a key to sustain first-line therapy effectiveness and to prevent HIV-drug resistance [14], [36]. Failure of treatment makes patients to follow another line of treatment. Timely detection of treatment failure with subsequent switch to second-line treatment reduces mortality among HIV infected people receiving ART [13].

Regular treatment failure detection is low because of inadequate capacity and lack of laboratory facilities in resource-limited settings including Ethiopia [37]. Moreover, monitoring and prediction of ART treatment and the failure thereof in Ethiopia is guided by human experts based on WHO protocols [30]. It is the duty of the physician to decide whether the patient should start second-line ART or not. The amount of time required to process all the records by considering the attributes for each and every record is large and the process is somewhat tiresome [38].

Different researchers assessed the existing evidence on ART treatment failure and associated factors in Ethiopia. These studies [7], [27], [29], [30] were valuable in identifying key determinants of first-line ART failure among adult patients on firs-line ART. The studies compared immunological, virological and clinical data of patients whose first-line ARTs failed against clients whose ARTs sustained by using statistical methods.

While immunological and clinical criteria are insufficient to correctly identifying treatment failure among adult patients receiving either first-line ART regimens supporting the use of viralogical criteria for predict first-line treatment failure [39], studies conducted by employing supervised learning for predicting ARV treatment failure employ varying definitions of virological failure (>50, >400, 400-1,000, 1,000-4,999, >1,000, >5,000, >10,000) suggesting inconsistency among researchers in providing viral load thresholds for prediction [40]. Moreover, extant predictive models on first-line ART failure are based on either 2010 [39].

Monitoring people on ART is important to ensure successful treatment, identify adherence problems and determine whether ART regimens should be switched in case of treatment failure [32]. However, in Ethiopia there is no application that enables ART therapists to predict risk of ART treatment failure so that they can take proactive measure.

This paper develops a machine learning based predictive model that detects which among ART clients is most at risk of experiencing first-line ART treatment failure based on 2021 WHO guideline [32] and consultation with expert from Ministry of Health.

1.3. Research Questions

This study proposes an investigation that will answer three specific questions.

- What are the main risk factors for first-line ART failure among adults on ART in Zewditu Hospital?
- What is the appropriate data mining methodology for predicting first-line ART failure among adults on ART in Zewditu Hospital?
- What data mining technique has the best predictive power to develop first-line ART failure prediction model for adults on ART in Zewditu Hospital?

1.4. Objective of the Study

1.4.1. General Objective

The main objective of this study is to build a predictive model that detects first-line ART failure among HIV/AIDS patients in Zewditu Hospital, Addis Ababa.

1.4.2. Specific Objectives

The specific objectives used to achieve the stated general objective are to:

- Identify risk factors for first-line ART failures among adults on ART in Zewditu Hospital.
- Determine appropriate data mining methodology for predicting first-line ART failure among adults on ART in Zewditu Hospital.
- Identify data mining technique has the best predictive power to develop first-line ART failure prediction model for adults on ART in Zewditu Hospital?

1.5. Research Methodology

A de-identified retrospective cross-sectional data of 6434 ART clients was collected from Zewditu Hospital, Addis Ababa for this study. Cios Model was adopted for data mining. Experimental design was employed to conduct the study. 3.8.5 Version of WEKA simulation software was used for training algorithms.

Four data classification algorithms (random forest, Bayesian classification, and support vector machines) were selected based on literature review to develop predictive model. To train algorithms, a 10-fold cross-sectional validation training method was applied.

The performance of the algorithms was evaluated by accuracy, ROC analysis, sensitivity, precision, specificity, recall and F1 score.

1.6. Significance of the Study

This study has important significance to the body of knowledge, policy making and practice of ART treatment.

Very limited studies have been conducted on predicting first-line ART failure using data mining algorithms in Ethiopia. This study will extend the knowledge of first-line ART failure detection through the application of data mining algorithms. The study will also motivate others to further enquiries into applications of machine learning in similar and other health care domains in Ethiopia.

Decision/policymakers pertinent to HIV/AIDS care will understand the importance of machine learning for improving targeting of interventions through differentiated models of care, increasing cost-effectiveness and improving patient outcomes, and this will help them design and implement appropriate data mining policy and strategies on HIV/AIDS patient data.

The findings of the proposed study will be helpful in detecting ART treatment failures and making ART diagnoses efficient as it can be used as a treatment-decision support tool in clinical practice. This will, in turn, strengthen the screening of HIV patients in risk of first-line ART failure, with a view to early detection and effective implementation of targeted interventions.

1.7.Scope of the Study

The scope of this study is determined in terms of *area, subject, time and methodological* coverage.

The target population of this study is sourced among ART clients from Zewditu Hospital in Addis Ababa, Ethiopia.

The key topics that are covered in this study are first-line ART failure, Cios Data Mining Model, WEKA software, and data mining algorithms (decision tree, random forest, Bayesian classification, and SVM).

The time scope of the study will be 2013 E.C. /2021 G.C. The study employs data of first-line ART clients in 2013 E.C.

This study follows an experimental research design based on Cios model, uses retrospective cross-sectional data, employed WEKA software to develop predictive model with four data mining algorithms (decision tree, random forest, Bayesian classification, and SVM).

1.8.Ethical Consideration

While conducting this study, ethical issues were given proper due consideration before, during and after the research process.

Before data collection process, the researcher provided a brief description of the nature of the study; obtained appropriate prior consent of the respondents; discussed purpose of the study and how data will be used; and developed composite profiles to guarantee the privacy and anonymity of respondents.

During data analysis, the researcher promoted integrity.

During report writing, the researcher reported multiple and/or contrary findings (if any); reported honesty [of data, findings, and conclusions]; gave credit to other similar or related studies; and gave credit for ownership [of data] to researcher, respondents, and adviser[s].

1.9. Organization of the Study

Including **Chapter One**, the structure of the study outlines **five** chapters. **Chapter two** of the study deals with literature reviews on definition, types and causes of employee turnover; turnover rate and intention, relationship between causes of turnover and turnover intention; managing turnover; turnover related theories and empirical works; and conceptual framework.

In **Chapter three**, the study describes research methodology that goes through research approach, research design, population, sampling design and technique, sources and types of data, data collection instruments, data collection procedures, methods of data presentation and analysis, and data quality assurance.

Chapter four covers data presentation, analysis, and findings of the study. The chapter summarizes the findings of the study, and discusses the findings along with pertinent literature. The **closing chapter** of the thesis incorporates summary of the findings, conclusions, limitations of the study and the recommendations by the researcher.

Reference and data collection instrument will also be included at the end of the study report.

2. REVIEW OF LITERATURE

2.1.Introduction

This chapter presents literature review conducted to set the context of the study. The chapter provides contents on the concept of data mining, techniques of data mining, data mining algorithms, data mining tasks, data mining models, and related works.

2.2.Data Mining

Gaining knowledge from a huge data set is referred to as data mining. This means knowledge that can be veiled due to the size of data can be extracted with the help of non-human intelligence. Data mining is a computational and a logical process applied to obtain useful meaning from a sizable dataset [1] [41] [42].

Data analysts apply data mining technique in order for them to help uncover new and practically important connections among data sets of a particular target. Therefore, the purpose of data mining is to extract hidden patterns in a dataset of concern that may be employed for decision-making processes in practical arenas [41].

Where data mining is predictive, its aim will be to generate models that employ use specific target data to predict the outcome of interest. Such predictive data mining techniques are applied in "medical diagnosis, prognosis, treatment planning and also for general screening purposes" [24].

2.2.1. Application of Data mining in Health Sector

Challenges related to processing, handling, and applications of huge healthcare data can be managed by applying machine learning algorithms [5]. This implies that also helps in decision-making can be effective by applying different machine learning techniques. Such decision making which helps practitioners predict diseases and make timely diagnoses affects the health of a patient in a positive way [5].

Sever clinical outcomes such as mortality or readmissions can be predicted by applying artificial intelligence on electronically recorded health data [18]. This supports health decision making.

In such applications different machine learning methods demonstrate different strengths and weaknesses. Of multiple techniques that have been applied decision tree and random forest are common methods followed by Support Vector Machines, LASSO regression, boosting methods, novel Bayesian approach [26].

2.2.2. Application of Data Mining in HIV/AIDS Treatment

HIV/AIDs treatment data has opened various opportunities for application of machine learning algorithms for various purposes. Among these are predicting the eligibility HIV patients for ART [5], [20]–[22], predicting CD4 count changes and their predictors among patients on ART [18], [19], and predicting patient response to ART [10], [11], [17].

Similarly, Machine learning algorithms have been applied for predicting the risk of ART termination among patients on ART [6], [31] predicting ART-induced toxicity of HIV patients receiving ART based on [8] and predicting pediatric survival among HIV/AIDS patients [24], [25], [43].

2.3. Techniques of Data Mining

In order to help identify barely detectable patterns in a dataset of interest, already built tools can be used in data mining. Among these, clustering, classification and regression are common data mining models that are employed to uncover patterns in target datasets [41].

2.3.1. Clustering

When a need to identify similar object classes in a dataset of interest arises, clustering technique employed to the target data. Clustering is a supervised classification. Particularly, clustering techniques are used uncover patterns of overall distribution and correlations between data attributes [42], [44].

Moreover, clustering can be used as a pre - processing approach to selecting and classifying sub - set attributes before classification task is performed [42], [44].

2.3.2. Classification

When categorizing the whole dataset is required, classification technique is applied on the data [16]. Classification supervised machine learning approach that is the most commonly used for data mining; and it involves learning and classifying [16].

In the learning phase, a dataset is supplied to machine learning algorithm following which training ensues, and then the algorithm generates rules and patterns from the dataset supplied to it [16].

In the second phase, classification is performed. Here, the trained algorithm is fed with test dataset. Following the classification, accuracy of a the algorithm will be evaluated [16].

Bayesian classifiers, neural networks, and SVMs (Support Vector Machines) are predominantly implemented classification algorithms [16].

2.3.3. Regression

Another supervised machine learning technique for data mining regression. When a need arises to predict a continuous and numerical datasets whose attributes are already known, regression is best approach to apply for this purpose [2].

A dataset with already known values is required for regression analysis to start. Regression analysis can be used for modeling the relationship between one or more independent variables and dependent variables.

Regression analysis is performed based on training process. The analysis estimates values of target dataset by comparing already known and predicted values [2].

2.4.Data Mining Algorithms

This section outlines four data mining algorithms. Reviews of decision tree, random forest, Naïve Bayes and Support Vector machine data mining algorithms are presented, respectively.

2.4.1. Decision Tree

A decision tree is an approach using a tree data structure such as a chart or matrix of choices and its feasible results in order to forecast the ultimate choice. It is a pseudo code to approach evaluated objectives. These kinds of algorithms are very popular for interactive learning and have been used effectively for various assignments [5].

2.4.2. Random Forest

Random Forests generally are ensemble learning methods that are used for classification and regression tasks. It works by bootstrapping from the training set [10].

2.4.3. Naïve Bayes

A naive (or simple) Bayesian (NB) classifier is a probabilistic classifier which assumes that all attributes contribute equally, and independently, to the final decision [42]. NB has got wider application in health-related data mining for two reasons. It is a simple data mining technique and also enables handling a data set with many features.

2.4.4. Support Vector Machines (SVM)

Support Vector Machines (SVM) is amongst the most popular and efficient classification and regression methods currently available. These algorithms apply simple linear methods in a high-dimensional feature space that is non-linearly related to the input space [10], [42].

2.5. Data Mining Tasks

Data mining models can be classified into two categories: descriptive (or unsupervised learning) and predictive (or supervised learning) [42].

2.5.1. Descriptive Data Mining

Descriptive data mining consists of a collection of techniques aiming to discover unknown patterns or relationships in data. This exploratory analysis includes clustering, association, summarization, and sequence discovery [42].

Predictive data mining infers prediction rules from data. It includes tasks such as classification, regression, time series analysis, and prediction [42].

2.5.2. Predictive Modeling

Predictive modeling as a function of data mining allows the ability to predict and determine an unknown value of an outcome variable (target variable) based on the values of independent variables [1]. The model is made up of independent variables which are called predictors which have a high probability of influencing the outcome variable [1].

Prediction modeling has three components: target data, predictor data and a model that maps the relationship between the two [45].

2.6.Data Miming Models

2.6.1. KDD

KDD (**Knowledge Discovery in Database)** refers to the broad data knowledge finding process and emphasizes the "high-level" application of specific data mining methods. Machine learning, pattern recognition, databases, statistics, artificial intelligence, expert systems knowledge acquisition, and data visualization are of interest to researchers [1], [41].

Knowledge discovery in data bases (KDD) used for data mining has five stages [1], [5], [41], [46]:

1. **Data selection:** This stage consists on creating a target dataset, or focusing on a subset of variables or data samples, on which discovery is to be performed. The data relevant to the analysis is decided on and retrieved from the data collection.
2. **Data pre-processing:** This stage consists on the target data cleaning and preprocessing in order to obtain consistent data.
3. **Data transformation:** It is also known as data consolidation; in this phase the selected data is transformed into forms appropriate for the mining procedure. This stage consists on the transformation of the data using dimensionality reduction or transformation methods.

4. **Data mining:** It is the crucial step in which clever techniques are applied to extract potentially useful patterns. It consists on the searching for patterns of interest in a particular representational form, depending on the DM objective.
5. **Interpretation/Evaluation:** This stage consists on the interpretation and evaluation of the mined patterns.

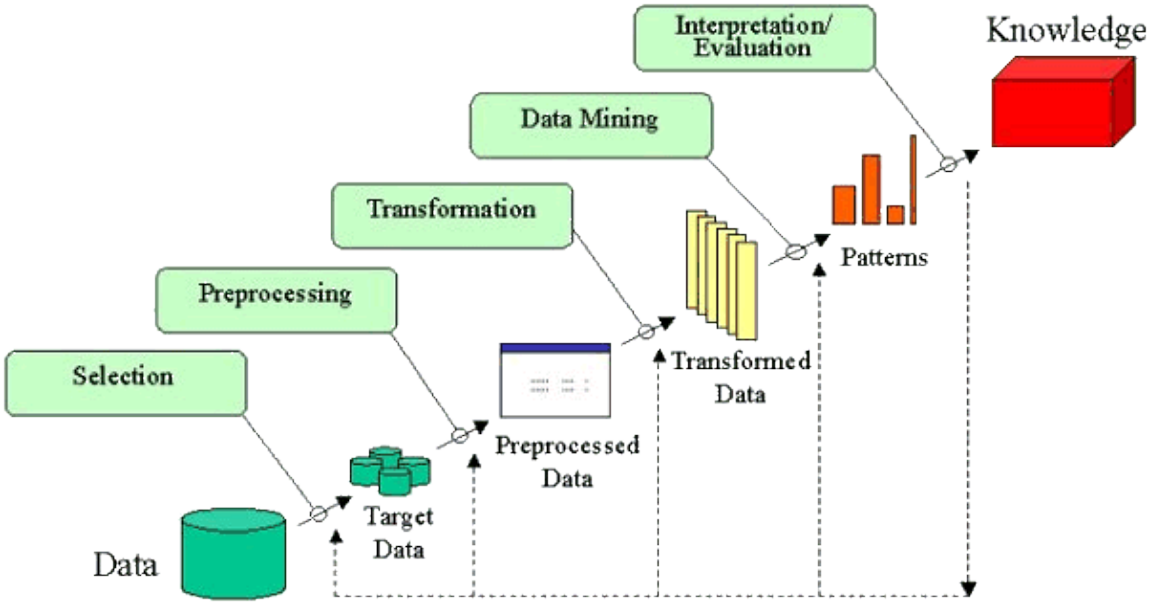

Figure 2-1: Scematic of KDD Process

2.6.2. SEMMA

The acronym SEMMA stands for Sample, Explore, Modify, Model, Assess. Beginning with a statistically representative sample of data, SEMMA intends to make it easy to apply exploratory statistical and visualization techniques, select and transform the most significant predictive variables, model the variables to predict outcomes, and finally confirm a model's accuracy. The SEMMA process has five stages [1], [41], [46]:

1. **Sample:** This is where a portion of a large data set (big enough to contain the significant information yet small enough to manipulate quickly) is extracted.
2. **Explore:** This is where the user searched for unanticipated trends and anomalies in order to gain a better understanding of the data set.

3. **Modify:** This is where the user creates, selects, and transforms the variables upon which to focus the model construction process.
4. **Model:** This is where the user searches for a variable combination that reliably predicts a desired outcome.
5. **Assess:** This is where the user evaluates the usefulness and the reliability of findings from the data mining process.

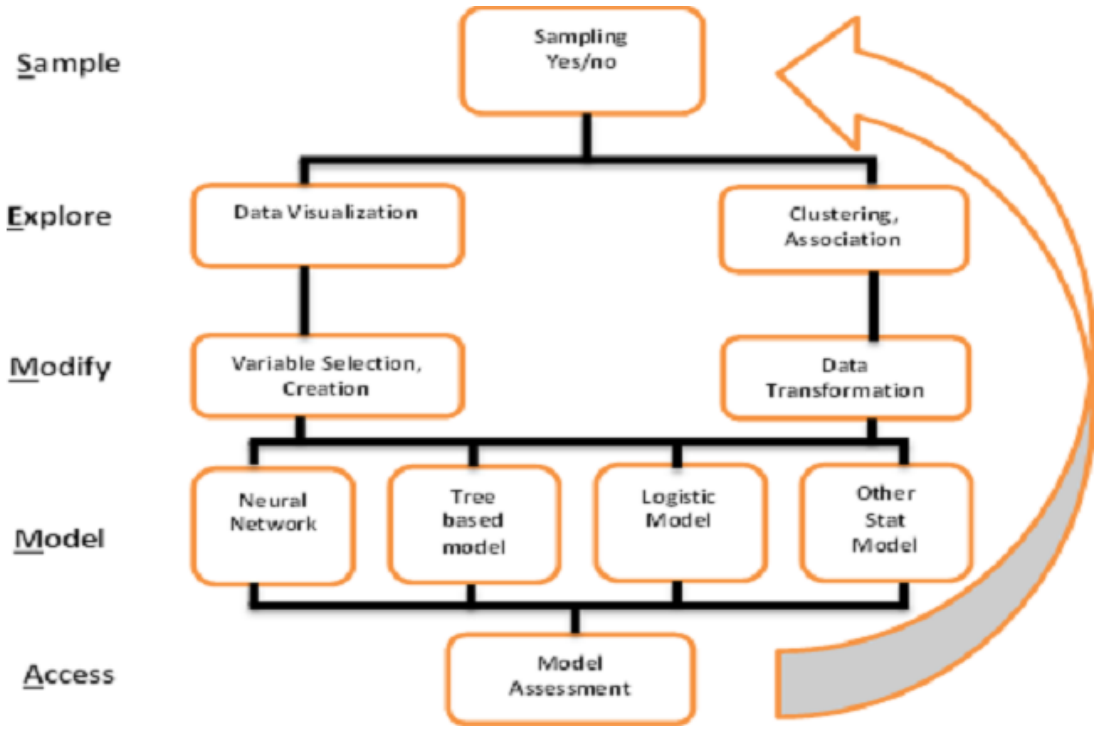

Figure 2-2: Schematic of the SEMMA Process

2.6.3. (CRISP-DM) Process

CRISP-DM (CRoss Industry Standard Process for Data Mining) is a data mining project compromises a multi-step, iterative process. It consists of the following six iterative phases [1], [18], [26], [41], [46], [47]:

1. **Business understanding:** this initial phase focuses on understanding the project objectives and requirements from a business perspective, then converting this knowledge into a DM problem definition and a preliminary plan designed to achieve the objectives.
2. **Data understanding:** the data understanding phase starts with an initial data collection and proceeds with activities in order to get familiar with the

data, to identify data quality problems, to discover first insights into the data or to detect interesting subsets to form hypotheses for hidden information.

3. **Data preparation:** the data preparation phase covers all activities to construct the final dataset from the initial raw data.
4. **Modeling:** in this phase, various modeling techniques are selected and applied and their parameters are calibrated to optimal values.
5. **Evaluation:** at this stage the model (or models) obtained are more thoroughly evaluated and the steps executed to construct the model are reviewed to be certain it properly achieves the business objectives.
6. **Deployment:** creation of the model is generally not the end of the project. Even if the purpose of the model is to increase knowledge of the data, the knowledge gained will need to be organized presented in a way that the customer can use it.

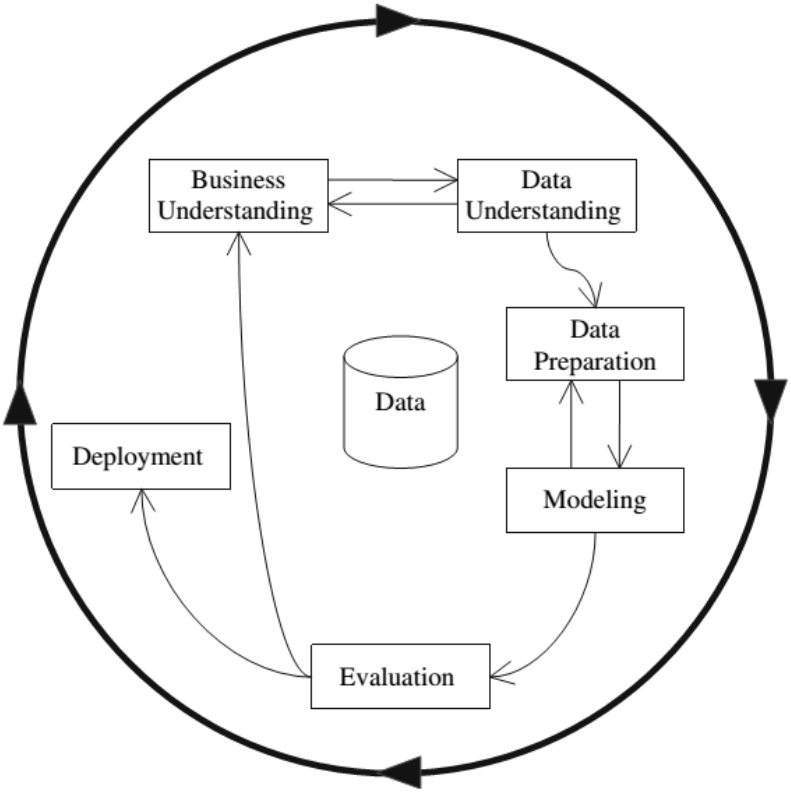

Figure 2-3: Schemtic of CRISP-DM Process

CRISP-DM provides a structured approach to data mining project planning. It's a well-proven and robust methodology [41]. Since it is not technology specific, industry-independent, and several data storage and data preparation technologies can

support the process, it is also a de facto standard for applying a process model in data mining projects [26].

2.6.4. The Hybrid Model

The Hybrid Model (aka Cios Model) consists of six steps [47]:

1. **Understanding of the problem domain:** In this step one works closely with domain experts to define the problem and determine the research goals, identifies key people, and learns about current solutions to the problem. A description of the problem including its restrictions is done. The research goals then need to be translated into the DM goals, and include initial selection of the DM tools.
2. **Understanding of the data:** This step includes collection of sample data, and deciding which data will be needed including its format and size. If background knowledge does exist some attributes may be ranked as more important. Next, we need to verify usefulness of the data in respect to the DM goals. Data needs to be checked for completeness, redundancy, missing values, plausibility of attribute values, etc.
3. **Preparation of the data:** This is the key step upon which the success of the entire knowledge discovery process depends; it usually consumes about half of the entire research effort. In this step, which data will be used as input for DM tools of step 4, is decided. It may involve sampling of data, data cleaning like checking completeness of data records, removing or correcting for noise, etc. The cleaned data can be, further processed by feature selection and extraction algorithms (to reduce dimensionality), and by derivation of new attributes (say by discretization). The result would be new data records, meeting specific input requirements for the planned to be used DM tools.
4. **Data mining:** This is another key step in the knowledge discovery process. Although it is the DM tools that discover new information, their application usually takes less time than data preparation. This step involves usage of the planned DM tools and selection of the new ones. DM tools include many types of algorithms, such as neural networks, clustering, preprocessing techniques, Bayesian methods, machine learning,

etc. This step involves the use of several DM tools on data prepared in step 3. First, the training and testing procedures are designed and the data model is constructed using one of the chosen DM tools; the generated data model is verified by using testing procedures.

5. **Evaluation of the discovered knowledge:** This step includes understanding the results, checking whether the new information is novel and interesting, interpretation of the results by domain experts, and checking the impact of the discovered knowledge. Only the approved models are retained. The entire DM process may be revisited to identify which alternative actions could have been taken to improve the results.
6. **Using the discovered knowledge:** This step is entirely in the hands of the owner of the database. It consists of planning where & how the discovered knowledge will be used. The application area in the current domain should be extended to other domains.

Cios model adopts the CRISP-DM model to satisfy the needs of academic research community, and thus a hybrid model [47].

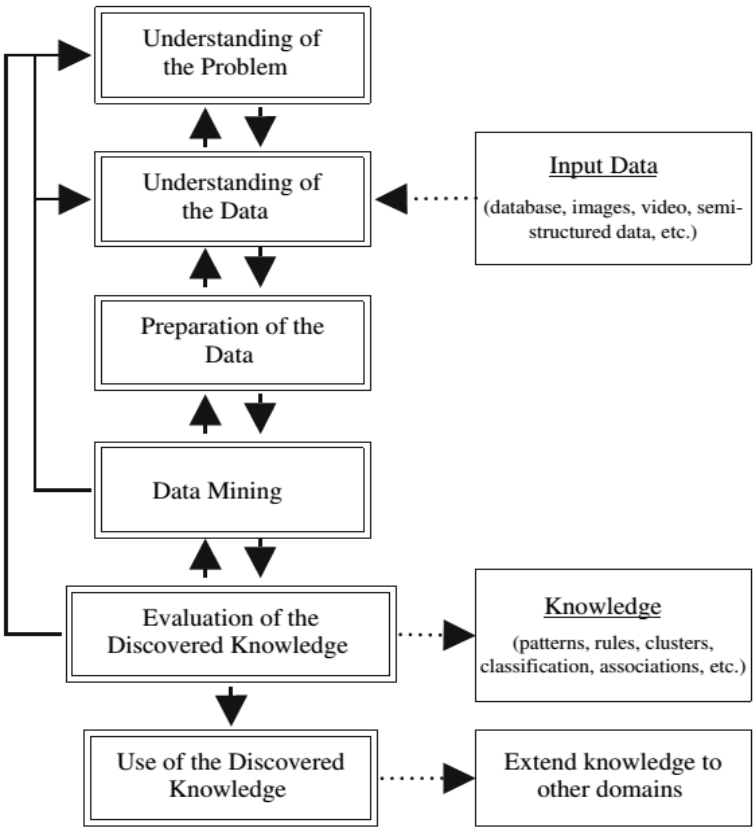

Figure 2-4: Schematic of the Cios Model

2.6.5. Comparison of Data Mining Models

The review of literature about the above four data mining models shows that the models have their uniqueness making them suitable for a specific data mining process. KDD employs a methodology that extracts knowledge from a database through the pre-processing of the database.

The unique part of the KDD process model is that it does not include business and data understanding. This model there is no involvement of domain experts. This proposed study intendeds to include opinion of domain experts. Therefore, the KDD process model is not relevant for the study.

SEMMA allows for development and maintenance of data mining projects. Since the purpose of the proposed study goes beyond data mining, SEMMA will not be adopted [1].

CRISP–DM has all its stages documented and organized hence easy to understand the flow of the process and easy to revise it. Different from the KDD model, the CRISP model includes business and data understanding.

However, CRISP-DM is limited in managing the requirements of current technologies like machine learning algorithms. Moreover, CRISP-DM does not cover the whole project lifecycle. Despite CRISP-DM defines a deployment phase explicitly, it does not adequately cover the deployment of the analysis in the productive environment, where model performance must be continuously monitored and controlled [26]. Therefore, CRSIP-DM will not also be considered for the proposed study.

The Cios model [47] that incorporates the needs of both the industry and academic community is will be adopted as a data mining model for this study. This is because the proposed study has an academic interest but also a practical relevance.

The models also share some similarities. KDD and SEMMA are similar in their first five stages. Some or most of the steps in KDD can be found in CRISP-DM and vice versa.

CRISP-DM and SEMMA are mostly industry oriented [1].

2.7.Related Works

Machine learning algorithms haven been applied for different purposes on HIV/AIDs patients' data.

Rutherford et al. [40] conducted a systematic review of the performance characteristics of the 2010 WHO immunologic and clinical criteria for virologic failure based on 18 studies that predict treatment failure in adults and children on antiretroviral therapy. By calculating unweighted sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of immunologic and clinical criteria for predicting virologic failure, the researchers found out that the 2010 WHO clinical and immunologic criteria are insensitive and have low PPV for predicting virologic failure. The limitation of this work is that there are up to 7 varying definitions of virological failure (>50, >400, 400-1,000, 1,000-4,999, >1,000, >5,000, >10,000 suggesting inconsistency among

researchers in providing viral load thresholds for prediction. In other words, the threshold viral load that enables prediction of ART failure is not clearly provided in the study.

Waruru et al. [39] identified positive predictive value of the [2010] WHO clinical and immunologic criteria to predict viral load failure among adults on first, or second-line antiretroviral therapy in Kenya. The researchers found that immunological and clinical criteria for correctly identifying treatment failure had low PPV for adult patients receiving either first-line or second-line/subsequent ART regimens suggesting inadequacy of clinical and immunologic criteria and requirement of virological criteria to correctly identify treatment failure. This has two weaknesses. In the first place it is not based on latest WHO guideline. Second, while it provided virolgical criteria to correctly identify treatment failure, it didn't provide the magnitude of threshold of recommended criteria

Idowu and Balogun [19] developed a classification model for CD4 Count of HIV Patients using a comparative analysis supervised machine learning algorithms. To predict CD4 count changes and predictors among patients on ART, this study followed a CRISP-DM data mining model. The study applied C4.5 decision trees, SVM, and MLP on WEKA software and found that the performances of the classifiers are 100% (MLP), and 91.1 %(SVM). While this study predicted CD4 changes, it didn't provide how CD4 changes relate to ART failure.

Cheng and Wnag [28] explored the important attributes of human immunodeficiency virus and generating decision rules. By employing CD4 count, clinical stage, treatment willingness, and drug abuse of first line ART client data from Taiwan, the researchers generated decision rules and proposed appropriate classifier. They proposed a rough set classifier based on adding recency (R) (i.e., the last physician visit), frequency (F) (i.e., the frequency of medical visits), and monetary (M) (i.e., medication adherence) attributes and integrated attribute selection methods to generate discriminatory rules and find the core attributes of HIV.

Liu et al. [7] analyzed optimal allocation of gold standard testing under constrained availability with an application to assessment of HIV treatment failure. The researchers provided augmenting rules of diagnosing treatment failure based on low-cost markers (such as CD4 cell count) with a selective use of VL testing.

Gunda et al. [48] studied the accuracy of 2010 WHO immunological criteria in identifying virological failure among HIV-infected adults on First line antiretroviral therapy in Mwanza, North-western Tanzania. The researchers assessing the accuracy of immunological criteria in detecting treatment failure among HIV infected Tanzanian adults receiving first line ART.

Rohr et al. [49] developed a predictive risk model for first-line antiretroviral therapy failure in South Africa. The researchers used stepwise (two consecutive viral load levels >1000 copies/mL) selection of predictors o predict virologic failure on first-line ART.

The literature review shows that decision tree [18]–[21], random forest [31], [45], [50], Bayesian classification [43], and support vector machines [19], data mining algorithms are widely used for prediction.

It is evident that decision tree, random forest, Bayesian classification, and SVM are dominant algorithms to attract the attention of researchers in this field.

Moreover, in the extant literature confirm that different versions of WEKA software [1], [5], python scikit-learn module [10], SPSS Modeler 18.0 [44] and RapidMiner 7.1.001 [42] have been employed to train the predictive models. It can be observed that WEKA is widely used software to train data mining algorithms for different purposes.

Previous researches also used 10-fold cross validation [24], percentage split [21], leave-one-out method [42], temporal cross validation [6], raining, testing and validation data [23] and 5-fold cross validation1000 Monte Carlo runs [8] methods to train predictive models. It is clear from the literature that 10-fold cross validation is suitable method to train predictive models in this field.

Further, while the reviewed literature employed accuracy [25] sensitivity [10], [22], precision [18], specificity [46], recall [18], F1-score [6], [24] and t-test [24] to evaluate a classifier's performance, they used ROC analysis [11] to evaluate the overall predictive classification performance of a model.

The review depicts that accuracy, sensitivity, precision, specificity, recall, F1-score and ROC analysis are important to evaluate predictive performance for machine learning.

Building on the literature, decision tree [6], [8], [38], random forest [31], [50], Bayesian classification [11], and support vector machines [19], algorithms were applied on Weka software [5] to train through 10-fold cross validation method [21] . The performance of the algorithms is evaluated by accuracy, ROC analysis, sensitivity, precision, specificity, recall and F1 score.

Extant literature suggests that there is inconsistency among researchers in providing viral load thresholds for prediction [40]. This means, the threshold viral load that enables prediction of ART failure is not clearly provided in the study. Therefore, in order to provide threshold viral load important for prediction of ART failure, the 2021 WHO guideline was consulted. Moreover, a discussion with expert from Ministry of Health was conducted.

According to a discussion with expert from Ministry of Health, CD4 cell count is less important for determining patients at high risk of first-line ART failure. Moreover, CD4 cell count is important for initiating ART and ART treatments are possible regardless of CD4 cell count based on the preference of individual identified as HIV positive [32].

The expert emphasizes that the main predictor of first-line ART failure is viral load. The 2021 WHO guideline reinforces this by providing that the threshold viral load for identifying clients with high risk of first-line ART failure is > 1000 copies/ml [32].

3. RESEARCH METHODOLOGY

3.1. Introduction

This chapter of the study presents research design, understanding of the domain problem, data understanding, data preparation and preprocessing, data mining model building, and evaluation of discovered knowledge.

3.2. Research Design

The proposed study will be conducted by following an experimental design. The experimentation process will employ the hybrid data mining model (aka Cios Model). Since study has both an academic interest but also a practical relevance, the Cios model [47] that incorporates the needs of both the industry and academic community is adopted as a data mining model for this study.

While Section 2.6.4 under Data Mining Models of this study, provides the details of the Cios Model, and here under is presented the same model as it is applied to the data collected for the purpose of this study.

3.3. Understanding the Problem Domain

To understand the problem domain, that is, treatment monitoring of ART, not only desk review of all critical and relevant documents related to ART but also expert consultation from Ministry of Health was made. Monitoring people on ART is important to ensure successful treatment, identify adherence problems and determine whether ART regimens should be switched in case of treatment failure [32]. Compared with clinical or immunological monitoring, viral load testing provides an early and more accurate indication of treatment failure and the need to switch from first-line to second-line drugs [32].

3.4. Data Understanding

This study is based on a retrospective cross-sectional data of ART clients from Zewditu Hospital's 2021 ART registers. The study utilizes data of adult HIV/AIDS patients

(age>15) [32], [33] who are receiving first-line and second-line antiretroviral therapy at least for six months [37], [40] when data was accessed.

A de-identified record of ART clients contains Name of Facility, Patient ID, Sex, Date of Birth, Age, Marital Status, Registration Date, WHO stage, Weight, Height, ART Start Date, ART Dose, Adherence, Follow-up Status, Functional Status, Viral load count, Date Viral load performed, Viral load status, CD4 cell Count, Transferred.

Cross-consultation of expert from Ministry of Health was conducted to check whether the data set satisfies quality issues (completeness, redundancy, missing values, plausibility of attribute values, etc.) and to select attributes based on their relevance [5].

Table 3-1: Description of Variables in the Raw Data

Variable	Type	Description
Name of Facility	Nominal	The facility where the patient is receiving treatment
Patient ID	Numeric	Identification number of the ART client in the hospital register
Sex	Binary	Binary indicator for gender; 1 is male; 0 is female.
Date of Birth	Numeric	Date patient is born
Age	Integer	Age of patient at enrolment
Marital Status	Nominal	Whether patient is married or not
Registration Date	Numeric	Date patient is enrolled at the facility
Weight	Numeric	Weight of patient at enrolment
Height	Numeric	Height of patient at enrolment
ART Start Date	Numeric	Date Patient started ART therapy
ART Dose	Numeric	Amount of ART patient is taking during the therapy
Adherence	Nominal	Poor, fair, good, Stopped

Follow-up Status	Binary	Binary indicator for follow up status: 1 is Active on ART, 0 is Null.
Functional Status	Nominal	Functional status of patient: Working, Ambulatory, or Bed-ridden
Viral load count	Numeric	Number of HIV in patients body in copies/mL
Date Viral load performed	Numeric	Date Viral load test is performed
Viral load Status	Nominal	Viral load status: Suppressed, Unsuppressed.

Source: ART clients' Data from Zewditu Hospital, 2021

3.5. Data Preparation & Preprocessing

3.5.1. Data Selection

The study used a list of 7,687 clients alive on ART from Zewditu Hospital in Addis Ababa. The list contains records of 7,470 adult (>15) ART clients, 216 child ART clients. Therefore, a list of 7,470 adult clients was filtered for prepossessing.

3.5.2. Data Cleaning

Records that have outliers and/or incomplete and/or missing values under each column were removed. 882 records that contain patient's weight of no or less than 20 kgs, 153 records with unknown viral load status, and 1 record of unknown patient record were removed. This left 6434 records for data mining. Of 6434 clients, 6372 are on first-line ART treatment and 62 are on second-line ART treatment. The researcher used MS-EXCEL application for cleaning the data [1], [21], [44].

The inconsistency of the data was identified with support of expert consultation and 2021 WHO guideline. This process has shown that there are no data represented in different ways than professionally accepted or 2021 WHO guideline recommends. Thus, there is no duplicate attribute.

After data cleaning procedure, guided by 2021 WHO guideline and expert recommendation, four attributes were selected to be used in the experiment. These attributes include Sex, Weight, and Viral load status.

3.5.3. Data Transformation and Data Reduction

In this phase variables were transformed into forms that can be handled by the data mining tools by adapting data reduction techniques[1], [23]. The focus was to eliminating extreme outliers, and excluding irrelevant variables and discretizing (binning) continuous variables [23].

In data transformation, the data are transformed or consolidated into forms appropriate for mining. In the dataset, majority of the variables include continuous attributes. Thus, attributes were discretized (binned) to reduce the distinct values of the attributes so that it suites the mining tool and to obtain meaningful patterns. Data discretization techniques can be used to reduce the number of values for a given continuous attribute by dividing the range of the attribute in to intervals. Interval labels can then be used to replace actual data values. Replacing numerous values of a continuous attribute by a small number of interval labels there by reduces and simplifies the original data. Discretization of continuous variable is presented in Table 3.2 below.

Table 3-2: Attribute Descrtization Labels

Attribute				
Age	16-20	21-35	36-60	61+
	Adolescent	Young	Adult	Old
Viral load count	<1000		>1000	
	Low		High	
Sex	Female		Male	
Weight	<50	>=50-<=75		>75
	Underweight	Normal		Overweight

Figure 3-1: Attributes Descrritization Labels

Source: Expert Consultation and 2021 WHO guideline, 2021

3.5.4. Attribute selection

The dataset includes number of attributes that include relevant and irrelevant variables for the study. Some important variables have large number of missing data and thus they were removed from the dataset. In addition, there were irrelevant variables for the study. Consequently, they are removed from the study data. Since the presence of irrelevant attributes tend to affect the efficiency of data mining algorithms and may create poor classifiers, appropriate and relevant attributes which help address the research objective were selected. Finally, the study has identified four important variables. The list of attributes finally selected included in the dataset are Sex, Weight, and Viral load count. The selection of attributes was based on relevance to the subject matter and general facts from reviewing literature.

3.5.5. Data formatting

The datasets provided to WEKA software was prepared in a format that is acceptable for the software. To feed the final dataset into the WEKA, the file was changed into other file format. The excel file was first changed into a comma-separated values (CSV) file format. After changing the dataset into a CSV format the next step was opening the file with the Weka data mining software.

3.6.Data Mining Model Building

In this phase, data mining techniques were selected with their respective parameters to obtain optimal values and applied on prepared data. Additionally, under this phase descriptive (clustering) and predictive (classification) data mining tasks were performed.

Training and testing procedures were designed; predictive model was constructed and verified using test dataset. Since this study intends to build a predictive model, the models intended to be addressed are classification models. This includes data mining tool selection and the algorithms used for modeling technique.

3.6.1. Data mining techniques

a. Clustering

Clustering allocates patients similar in their attributes to the same group, thus providing natural groupings of the patients based on similar health utilization behavior [23]. The clustering experiments were conducted by using K-mean clustering algorithms.

The experiments were conducted by altering distance function and seed size. Since the study intendeds to classify the behavior of the ART clients into two as sustainable and failure, the value of k will be 2 [42], [44].

b. Classification

Classification as a learning function links a data item into one of several predefined classes. Patient risk scores of two outcomes [31] will be developed for this study. This study therefore classifies ART into categories of failure and/or sustainable treatment.

For this purpose, the study employed four classifiers that have wide application in the subject area: decision tree [1], [2], [5], [6], [8], [18]–[21], [38], [44], [46], random forest [1], [6], [8], [10], [18], [23], [31], [45], [50], Bayesian classification [1], [2], [11], [38], [42], [43], [45], [46], and support vector machines [8], [10], [19], [25], [42], [45].

3.6.2. Training and testing procedure

The four classifiers were trained with the dataset before they can be used. Training on the preprocessed dataset was performed by using stratified 10-fold cross-validation test parameter [1]. This was followed by using the cross validation test parameter to divide the instances randomly into 10 parts.

WEKA 3.8.5 Stable Version was used to test and train the classifiers [1], [5], [18], [19], [21], [24], [25], [43], [46] by using training datasets and the proposed models were tested using test sets of data. WEKA was chosen as a simulation environment because it runs on any modern computing platform, contains a comprehensive

collection of data preprocessing and modeling techniques. WEKA supports several standard data mining tasks like data clustering, classification, regression, preprocessing, visualization and feature selection [21].

3.7.Evaluation of the Discovered Knowledge

This stage is where understanding and interpretation of the results will be made. Where necessary, the results will be interpreted by domain experts. Confusion matrix are used to demonstrate classification problems [10], [22], [25], [43].

The classifiers' performance was evaluated in terms of accuracy (proportion of observations correctly classified by the algorithm among all observations in the unseen test set) [5], [8], [10], [18], [19], [21], [22], [24], [42], [43], [46], sensitivity (the proportion of known positive outcomes in the unseen test set that are correctly identified as such by the algorithm) [10], [22], [25], [43], [46], and positive predictive value also known as precision (the proportion of positive outcomes predicted by the algorithm that correspond to known positive outcomes in the unseen test set) [5], [18], [24], [25], [42].

In addition, specificity (the proportion of known negative outcomes in the unseen test set that are correctly identified as such) [10], [43], [46], recall (the number of correct positive results divided by the number of positive results that should have been returned) [18], [24], [42] and F1 score (the harmonic mean of classification precision and recall) [6], [24] were also employed to evaluate classifiers' performance.

Further, the area under the curve (AUC) of a receiver operating characteristic (ROC) curve will be utilized to evaluate the broad predictive classification performance of the model [10], [11], [18], [22], [24], [25], [43], [46].

4. EXPERIMENTATION

4.1. Introduction

This study was conducted with an objective of developing predictive model for detection of First-line ART treatment failure among adult ART clients in Zewditu Hospital, Addis Ababa. To achieve this objective the study followed data mining strategy. In line with the research objective, this chapter presents procedure followed during experiment for the data mining. The chapter discusses the result of experiment on clustering, classification and model evaluation. Finally, discussion about the discovered knowledge is presented.

The study has used dataset of 6434 adult ART clients that indicates ART treatment follow-up by ART therapists in Zewditu Hospital. Following scholarly practice that has better training testing [18], among these instances, 90% (5,791) instances were used for the model training and the 10% (643) instances were used for testing the model. The instances for the testing were randomly selected.

The data mining process was conducted by using WEKA data mining tool Version 3.8.5 for Windows. For clustering the study used simple K-means clustering algorithm and 3 experiments were conducted. On the other hand for the classification, the study implemented different algorithms such as decision tree (J48), NaïveBayes, random forest tree, and support vector machine algorithms. The best algorithm was selected based on classification accuracy.

4.2. Model Building

Model building process followed two steps: cluster modeling and classification modeling. Model building process of the study is conducted by using 5,791 training dataset. The cluster modeling aims to form segment of ART clients with risk of treatment failure and treatment sustenance. Clustered dataset is developed and used for training of classification model. The selection of clusters is decided based on judgment of domain experts.

The classification modeling is conducted to build ART treatment failure detection model by forming association from the attributes suggested. The model training for classification model development was conducted by using 10-fold cross-validation and percentage split.

To test the prediction performance of the classification model developed, separately prepared testing dataset was used. The testing used 643 instances that are randomly selected from the original dataset.

The preprocessing dataset in Weka is presented in figure 4.1 below for sample snapshot.

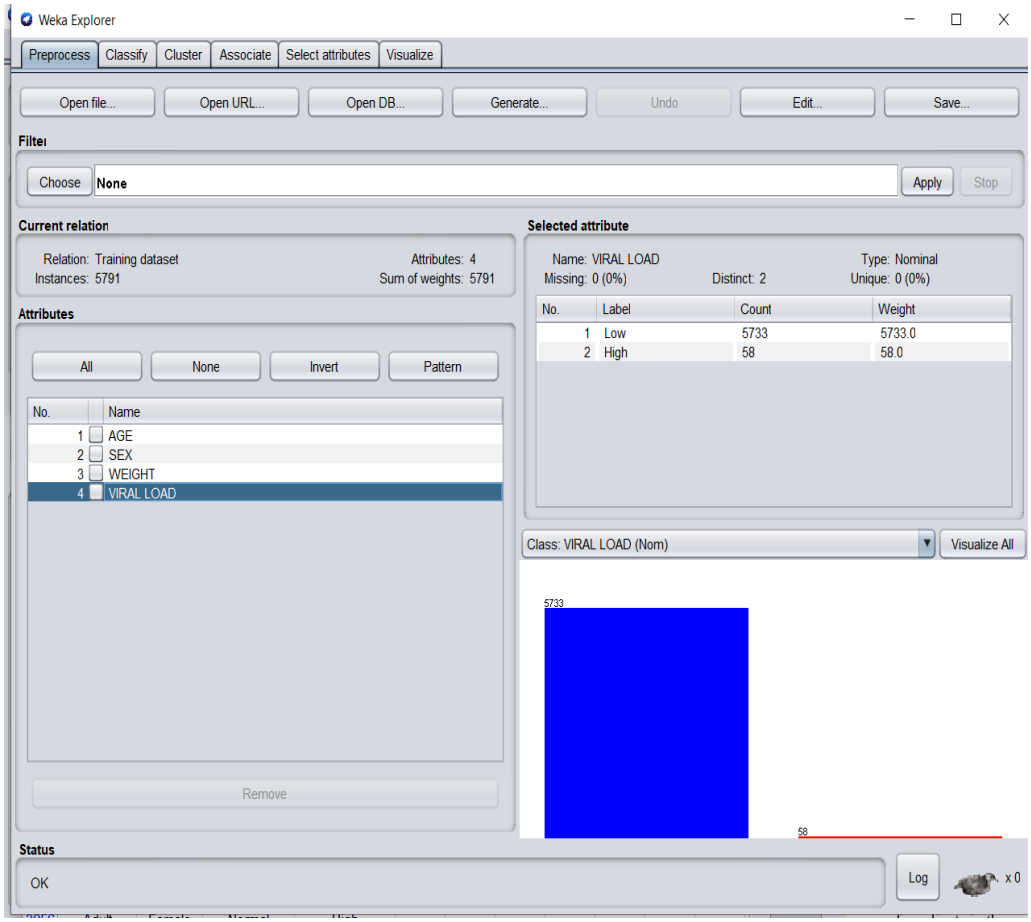

Figure 4-1: Preprocessing Data in Weka 3.8.5

(Source: Result Weka, 2021)

The study identified 2 labels for three attributes and non for one attribute. The last attribute is viral load and the attribute comprises 5733 cases in the first category and 58 cases in the second category. This implies the number of cases in the second category suggest that there are few clients in the second-line of ART treatment (viral load >1000).

4.3. Cluster Modeling

This study has conducted experiments to develop a clustering model and the model is used to develop the classification model. The clustering experiments were conducted by using K-mean clustering algorithms [42], [44]. The experiments were conducted by altering distance function and seed size.

The study has intended to classify the ART clients into two: Sustaining and Failed. Thus, the K-value set to be 2 suggesting sustaining and failed treatments. Although the line of treatment goes beyond two, for the purpose of simplicity, the study has used only two clusters. In addition to computational simplicity, domain experts were consulted in the authority and suggested that it is better to classify clients alive on the first-line treatment into two.

For the clustering result decision, the study has used three criterion; intra cluster similarity measure, number of iteration to conduct a convergence, and judgment of domain expert. The intra cluster similarity is measured by within cluster sum of squared error (the lower is better). As presented in previous chapter, the study has finally selected 4 attributes. In addition, the attribute VIRAL LOAD is represented in 2 labels.

According to the domain expert and 2021 WHO guideline, the ART treatment failure mainly depends on this attribute.

4.3.1. Experiment I

The first experiment was conducted by using the default values of the data mining tool; K = 2, EuclideanDistance as distance function and seed value of seed = 10 on training dataset. The result of experiment I is summarized in Table 4-1 below.

Table 4-1: Cluster Modeling Experiment I Result Summary

	CLUSTER INSTANCES	
Attribute	**1**	**2**
	3298(57%)	2493(43%)
AGE		
Young	464.0974	217.9026

Old	198.0494	389.9506
Adult	2523.8677	1789.1323
Adolescent	116.3503	99.6497
[Total]	3302.3648	2496.6352
SEX		
Female	3288.9041	11.0959
Male	11.4607	2483.5393
[Total]	3300.3648	2494.6352
WEIGHT		
Normal	2332.7897	1626.2103
Overweight	561.6424	757.3576
Underweight	406.9328	112.0672
[Total]	3301.3648	2495.6352
VIRAL LOAD		
Low	3268.2789	2466.7211
High	32.0859	27.9141
[Total]	3300.3648	2494.6352
RANK	**2**	**1**

Source: Weka Clustering Result, 2021

Table 4-1 above depicts the result of first experiment conducted to develop clustering model and presents instances in each cluster, attributes in the clusters and ranks of the clusters in explaining the risk of ART treatment failure.

As shown in the Table 4-1 above, 3,298 (57%) of the instances are grouped in Cluster 1 and 2,493 (43%) of the instances are grouped in Cluster 2. The result about attributes in the clusters suggests some attributes commonly exist in both clusters.

The cluster modeling is conducted to form clusters that indicate sustained or failed ART treatments. The clusters created are used as dependent variable while the attributes are used as independent variables.

Since ART treatment failure is associated with High VIRAL LOAD, Cluster 2 with High-to-Low viral load ratio of 1.09 % is better than Cluster1 with 0.98% High-to-

Low viral load ratio. The quality of the experiment is in agreement with the actual data where High-to-Low viral load ratio is 1.01%.

4.3.2. Experiment II

This experiment is conducted to develop comparable clustering model in relation to model identified in Experiment I. similar to Experiment I, this experiment uses K = 2 and used EuclideanDistance as distance function. But the seed value is changed and implemented seed = 100. The result of Experiment II is summarized in Table 4-2.

Table 4-2: Cluster Modeling Result of Experiment II Result Summary

	CLUSTER INSTANCES	
Attribute	**1**	**2**
	2493 (43%)	3298(57%)
AGE		
Young	190.9295	491.0705
Old	398.3752	189.6248
Adult	1750.4796	2562.5204
Adolescent	89.337	126.663
[Total]	2429.1213	3369.8787
SEX		
Female	125.6288	3174.3712
Male	2301.4925	193.5075
[Total]	2427.1213	3367.8787
WEIGHT		
Normal	1572.3581	2386.6419
Overweight	765.7357	553.2643
Underweight	90.0275	428.9725
[Total]	2428.1213	3368.8787
VIRAL LOAD		
Low	2401.7549	3333.2451
High	25.3664	34.6336
[Total]	2427.1213	3367.8787
RANK	**1**	**2**

Source: Weka Result, 2021

As shown in Table 4-2 above, 2493 (43%) of the instances are clustered in Cluster 1 and the remaining 3298(57%) of the instances are clustered in Cluster 2. Regarding the classification of the attributes, the result of this experiment shows attributes are not well differentiated within the clusters.

However, considering High VIRAL LOAD as main predictor of ART treatment failure, Cluster 1 having High-to-Low viral load ratio of 1.04 % is better than Cluster 2 with 1.02% of High-to-Low viral load ratio. The quality of the experiment is in agreement with the actual data where High-to-Low viral load ratio is 1.01%.

4.3.3. Experiment III

This experiment is conducted with K = 2, seed = 1000 and by using EuclideanDistance as distance function holding defaults of simple K means clustering method in the data mining tool. Result of the experiment is summarized in Table 4-3 below.

Table 4-3: Cluster Modeling Experiment III Result Summary

	CLUSTER INSTANCES	
Attribute	**1**	**2**
	3293(57%)	2493(43%)
AGE		
Young	464.0974	217.9026
Old	198.0494	389.9506
Adult	2523.8677	1789.1323
Adolescent	116.3503	99.6497
[Total]	3302.3648	2496.6352
SEX		
Female	3288.9041	11.0959
Male	11.4607	2483.5393
[Total]	3300.3648	2494.6352
WEIGHT		
Normal	2332.7897	1626.2103

Overweight	561.6424	757.3576
Underweight	406.9328	112.0672
[Total]	3301.3648	2495.6352
VIRAL LOAD		
Low	3268.2789	2466.7211
High	32.0859	27.9141
[Total]	3300.3648	2494.6352
RANK	**2**	**1**

Source: Weka Result, 2021

As depicted in Table 4-3 above, 3293 (57%) of the cases are grouped in Cluster 1 and 2493 (43%) of the cases are grouped in Cluster 2. Regarding the classification of the attributes, still the result of this experiment shows attributes are not well differentiated within the clusters.

However, considering High VIRAL LOAD as main predictor of ART treatment failure, Cluster 2 having High-to-Low viral load ratio of 1.13% is better than Cluster 1 with 0.98 % of High-to-Low viral load ratio. The quality of the experiment is in agreement with the actual data where High-to-Low viral load ratio is 1.01%.

As a result, Cluster 2 better suggests ART treatment failure. Based on the values of attributes, in the experiment III, Cluster 2 is ranked 1st and Cluster 1 is ranked 2nd.

4.3.4. Comparison of Clustering Models

Three experiments were conducted to develop the cluster models by using simple K mean clustering algorithms. The experiments were conducted by using K = 2, EuclideanDistance as distance function and changing seed values. In the previous sections, the results of experiments were evaluated by domain experts and suggestions were provided.

This section presents evaluation of the experiments and suggestion on best clustering model based on the clustering algorithm procedures. The performance of the best model is suggested based on number of iterations and within cluster sum of squared errors. The performance measurements of within cluster sum of squared errors

indicate intra and inter cluster similarity and it is main indicator of goodness of the clustering model. The lower values of within cluster sum of squared errors suggests good model than the higher values. Similarly, smaller number of iteration suggests better model and it indicates the algorithm has converged very soon.

The comparison of the models generated from the experiments conducted is summarized in Table 4-4 below based on the selected parameters.

Table 4-4: Comparision of Clustering Models

Experimentation	Number of Iterations	Within cluster sum of squared errors
I	5	18274
II	3	18297
III	2	18217

Source: Weka Result, 2021

As shown in the Table 4-4 above, the three experiments depict relatively similar within cluster sum of squared errors. This indicates the model from experiment III is best clustering model. Therefore, this study reveals the model developed in the Experiment III is best clustering model and it is selected as final clustering model.

4.4. Classification Modeling

In the previous section, development procedure of clustering model is presented. Following development of the clustering model, the classification modeling is developed since the clustering model cannot classify new instances. The classification model analyzes accuracy of classifiers while categorizing the tax reporting into specified classes. This section of the study presents the result of classification modeling.

The classification modeling is conducted by using different algorithms that enable to choose the best classification model. This study has used decision tree (J48), Bayes algorithms (NaiveBayes), random forest and SVM. To test performance of the classification models, separate testing dataset was used. The classification modeling has

used attributes selected for the cluster model building as independent variables and the clusters built by clustering algorithms are used as dependent variable.

4.4.1. Decision Tree (J48) algorithm

Different experimentations were conducted to identify best classification modeling. The first experiment was conducted by using J48 algorithm that build decision tree model. The experiments were conducted by default values for pruning (confidence factor) and minimum number of instances per leaf (minNumObj(MNO)) with 10-folds cross validation. The summary of result of this experiment is presented in Table 4-5 below.

A snapshot of the classifier performance and confusion matrix is provided in Appendix 1.

Table 4-5: Summary of Confusion Matrix J48 algorithm

Test Option	Time taken	Classified Instances	
		Correctly classified	Incorrectly classified
Cross validation	0.08	5733 (98.9984%)	58 (1.0016 %)

Source: Weka Result, 2021

As shown in Table 4-5 above, under the cross validation test option, 98.998% of the instances are correctly classified but 1.0016% of the instances are incorrectly classified. To build this model it took only 0.08 seconds.

The threshold curve of the model is presented in Figure 4-2 below for Cluster 3.

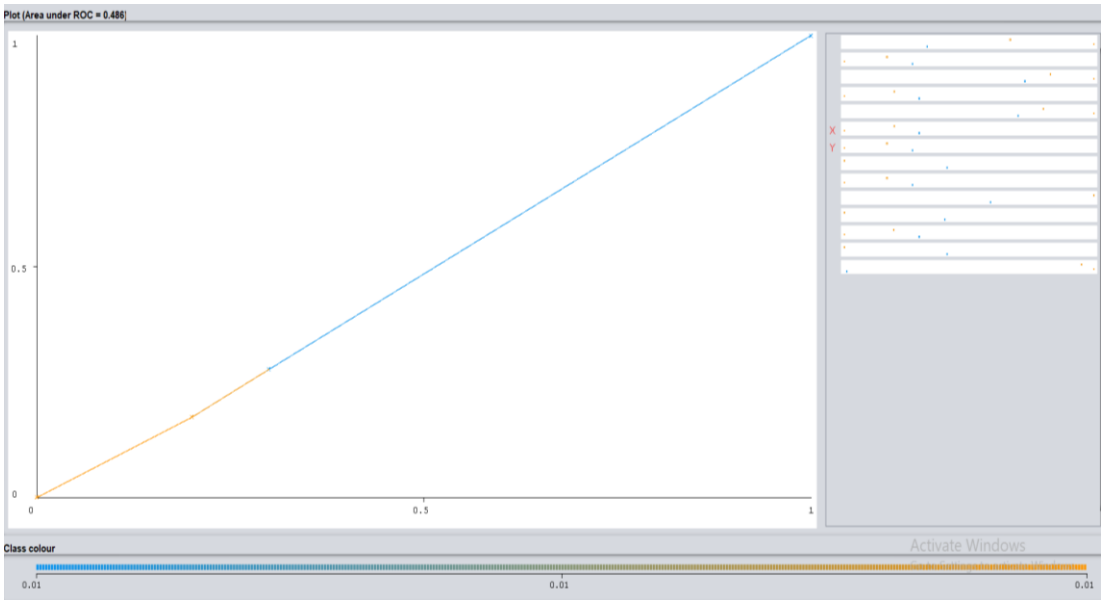

Figure 4-2: Threshold Curve of Decision Tree algorithm

Source: Weka Result, 2021

4.4.2. Naïve Bayes Classifier

To explore classification model another set of experiment is conducted by using Naïve Bayes classifier. The experiment by using Naïve Bayes classifier algorithm was conducted by using the default values of the data mining tool. Similar to previous classification modeling experiments, this experiment was conducted by using 10-fold cross validation. The summary of result of this experiment is presented in Table 4-6 below.

A snapshot of the classifier performance and confusion matrix are provided in Appendix 2.

Table 4-6: Result of Naive Bayer Classifier Experiment

Test Option	Time taken	Classified Instances	
		Correctly classified	Incorrectly classified
Cross validation	0.01 seconds	5733 (98.9984 %)	58 (1.0016 %)

Source: Weka Result, 2021

As shown in Table 4-6 above, under the cross validation test option, 98.9984% of the instances are correctly classified but 1.0016% of the instances are incorrectly classified. To build this model it took only 0.01 seconds. The threshold curve of the model is presented in Figure 4-3 below for Cluster 3.

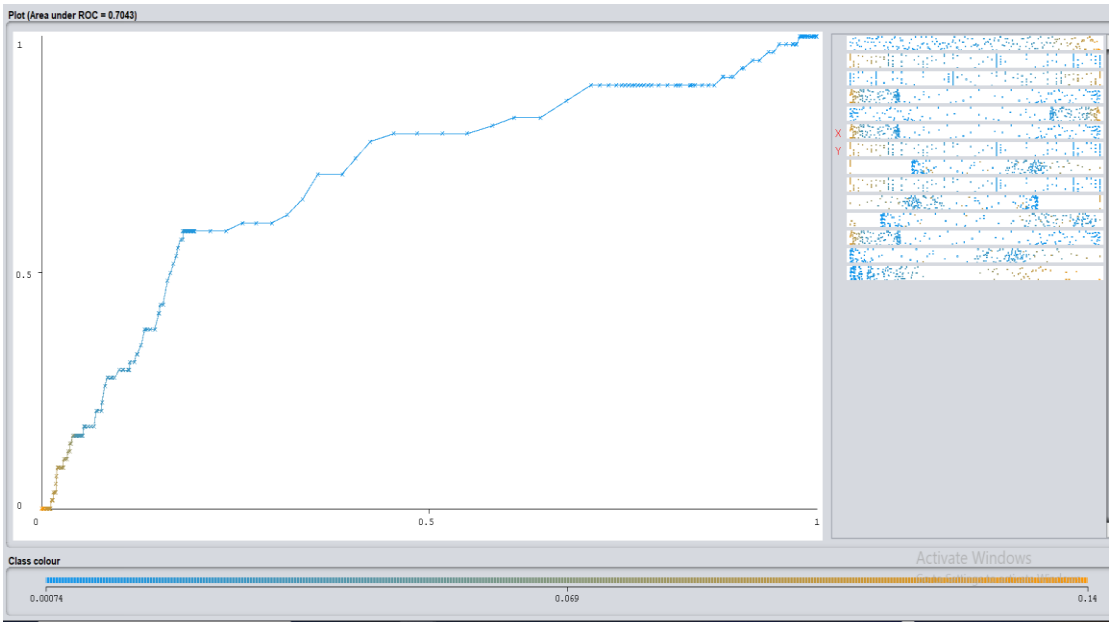

Figure 4-3 Threshold Curve of Naive algorithm:

Source: Weka Result, 2021

4.4.3. Support Vector Machine

The third experiment was conducted by using 10-fold cross-validation. The results of this experiment is summarized and presented in Table 4-7 below.

A snapshot of the classifier performance and confusion matrix are provided in Appendix 3.

Table 4-7: Result of Random Forest algorithm

Test Option	Time taken	Classified Instances	
		Correctly classified	Incorrectly classified
Cross validation	0.54 seconds	5733 (98.9984 %)	58 (1.0016 %)

Source: Weka Result, 2021

As depicted in Table 4-7 above, under the cross validation test option, 98.9984% of the instances are correctly classified but 1.0016% of the instances are incorrectly classified. To build this model it took only 0.54 seconds. The threshold curve of the cluster 3 is presented in Figure 4-4below.

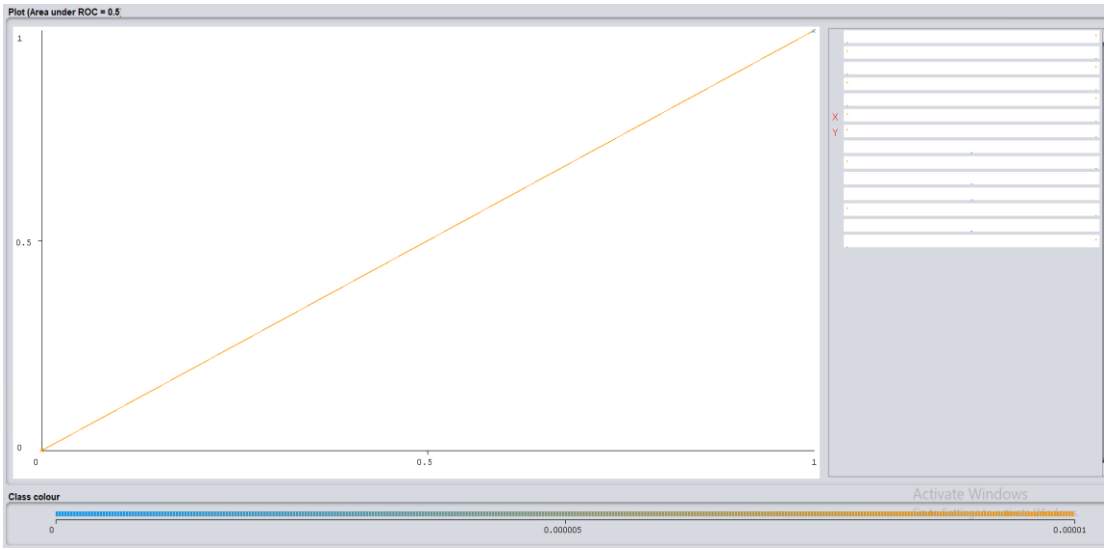

Figure 4-4: Threshold Curve for Support Vector Machine

Source: Weka Result, 2021

4.4.4. Random Forest

The last set of the experiment was conducted to develop classification model with random forest algorithm. The experiment was conducted by using random forest algorithm with default values of parameters of the data mining tool by using the 10-fold cross validation. The result of the experiment is summarized in Table 4-8.

A snapshot of the classifier performance and confusion matrix is provided in Appendix 4.

Table 4-8: Summary of Random Forest Experiment :

Test Option	Time taken	Classified Instances	
		Correctly classified	Incorrectly classified
Cross validation	0.23 seconds	5733 (98.9984 %)	58 (1.0016 %)

Source: Weka Result, 2021

As shown in Table 4-8 above, the model building took 0.23 seconds, and predicted 98.9984% of the instances accurately and 1.0016% of instances inaccurately. The threshold curve of the cluster 3 is presented in Figure 4-5 below.

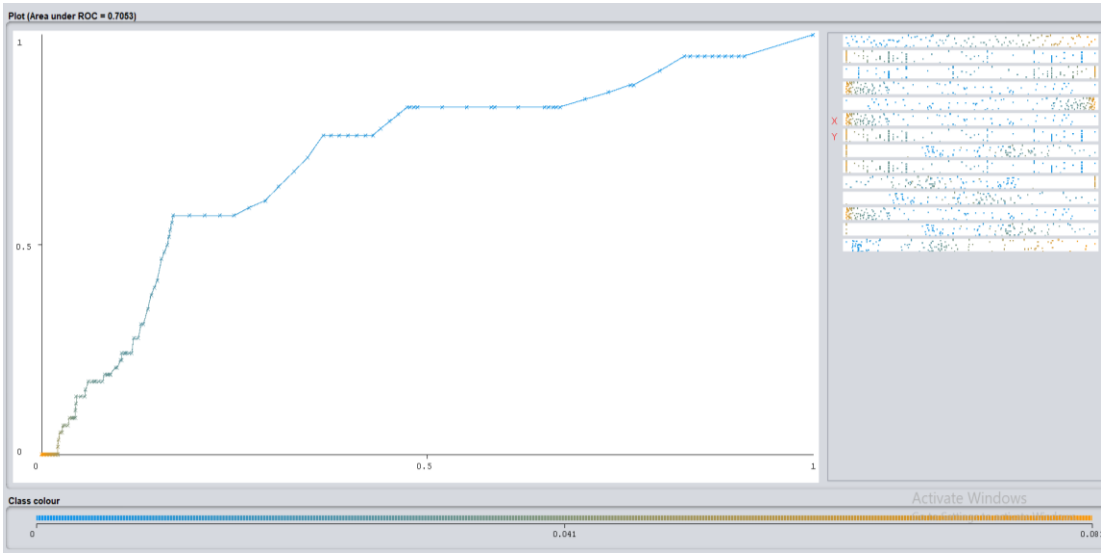

Figure 4-5: Threshold Curve for Random Forest

Source: Weka Result, 2021

4.4.5. Comparison of the results

This study was conducted mainly to develop model to handle ART treatment failure in Zewditu Hospital of Addis Ababa. In line with this general objective study intends to identify best classification algorithm for detection of the tax evasion. In the previous section, different classification techniques were used while conducting experiments to develop best classification model. This section of the study presents comparison of classifiers used to build classification model.

As presented in Section 4.2.2 above, this study was conducted by using 4 classification algorithms. These algorithms include decision tree (J48), Bayes, SVM and Random Forest. Experiments for classification model development were conducted based on 10-fold cross-validation. Best model was selected based on prediction accuracy. Summary of performance of each best classifier is presented in Table 4-9 for comparison and selection of best classification model.

Table 4-9: Sumary of Performance of Classifiers

Classifier	Test option	Time (sec)	Accuracy	Precision	Recall	ROC	F1-score
J48	10-fold	0.08	98.998%	0.990	1.000	0.486	0.995
Naïvebayes	10-fold	0.01	98.998%	0.990	1.000	0.704	0.995
SVM	10-fold	0.54	98.998%	0.990	1.000	0.500	0.995
Random Forest	10-fold	0.23	98.998%	0.990	1.000	0.705	0.995

Source: Weka Result, 2021

As depicted in Table 4-9 above, all the algorithms are the same in terms of accuracy (98.998%), precision (0.990), recall (1.00), and F1-score (0.995). They differ in the time taken to build the classification model. J48 and Naïve Bayes algorithms are have better time efficiency. Therefore, this study selected the J48 and Naïve Bayes algorithms as best algorithms to develop ART treatment detection model.

4.5.Evaluation

To achieve the objective of this research, cluster modeling and classification modeling experiments were conducted. The cluster modeling was conducted with simple K-mean algorithm (K = 2) and varying the seed values (10, 100, 1000). At the end, the clustering model was built at seed value of 1000 by using Euclidean distance function. This model segmented 98.998% instances to cluster 2. Classification experimentations were conducted following clustering experimentations.

The classification modeling was conducted by using four classifiers; J48, NaiveBayes, SVM and random forest. The experimentation followed 10 fold cross-validation training method.

The best experiments were selected based on time taken to build the classification model.

5. SUMMARY, CONCLUSION AND RECOMMENDATION

5.1. Summary

This study attempted to develop machine learning based predictive model for first-line ART treatment based on retrospective cross-sectional data from Zewditu Hospital, Addis Ababa. The study used decision tree (J48), NaiveBayes, SVM and random forest algorithms for developing predictive model. The experiment followed 10 fold cross-validation method to train the algorithms on WEKA 3.85 stable version based on Cios Data Mining Model.

All the algorithms were the same in terms of accuracy (98.998%), precision (0.990), recall (1.00), and F1-score (0.995). They differ in the time taken to build the classification model and ROC values. J48 and Naïve Bayes algorithms are have better time efficiency. Accordingly, the J48 and Naïve Bayes algorithms were found the best algorithms to develop ART treatment detection model for the data considered in this study.

5.2. Conclusion

This study was conducted by implementing the data mining techniques to detect and predict risk of treatment failure of first-line ART clients in Zewditu Hospital, Addis Ababa. Cios data mining model was applied for understanding the problem domain, understanding the data, preparing the data, data mining, evaluation of the discovered knowledge, and using the discovered knowledge. Based on the aforementioned objective the model development was conducted in two phases; cluster modeling and classification modeling.

The cluster modeling was conducted by using simple K-mean algorithm to segment data in tax evasion and no evasion. Different clustering models were experimented by the study and the best clustering model was developed by using k = 2, seed = 1000 and Euclidean distance function.

After clustering, the classification modeling was developed by using four classification algorithms; J48, Naïvebayes, SVM and Random Forest. All the algorithms were the same in terms of accuracy (98.998%), precision (0.990), recall (1.00), and F1-score (0.995). They differ in the time taken to build the classification model and ROC values.

The ROC values are 0.486, 0.74, 0.5, and 0.75 respectively for J48, Naïvebayes, SVM, and Random Forest algorithms respectively. The study identified that the J48 and Naïve Bayes algorithms were found the best algorithms to develop ART treatment detection model for the data considered in this study.

5.3.Recommendations

This study puts forth the following two recommendations. The first recommendation is the development of deployment systems, such as the automated evaluation of ART treatment to aid ART therapists in clinical settings in the future. The second recommendation is the addition of other attributes or the use of other models to develop predictive model for ART treatment failure.

5.4.Limitations of the Study

This study has two main limitations. Though the study considers expert consultation and the latest WHO guideline, it considers only one feature, VIRAL LOAD. This is the first limitation of the study. The second main limitation of this study its inability to deploy its result.

REFERENCES

[1] D. A. Annang, "Performance comparison of data mining techniques for predicting hiv status among female sex workers in Ghana," PhD Thesis, University of Ghana, 2018.

[2] S. Jayanthi and M. Revathi, "Major disease diagnosis and treatment suggestion system using data maining techniques".

[3] K. R. Bisaso, G. T. Anguzu, S. A. Karungi, A. Kiragga, and B. Castelnuovo, "A survey of machine learning applications in HIV clinical research and care," *Computers in biology and medicine*, vol. 91, pp. 366–371, 2017.

[4] USAID, "Global HIV & AIDS statistics — Fact sheet," *https://www.unaids.org*, 2021. https://www.unaids.org/en/resources/fact-sheet (accessed Mar. 15, 2021).

[5] T. D. Chala, "Data mining technology enabled anti retroviral therapy (ART) for HIV positive patients in gondar university hospital, Ethiopia," *Bioinformation*, vol. 15, no. 11, p. 790, 2019.

[6] A. Kumar *et al.*, "30 A machine learning system for retaining patients in HIV care," *arXiv preprint arXiv:2006.04944*, 2020.

[7] T. Liu, J. W. Hogan, L. Wang, S. Zhang, and R. Kantor, "Optimal allocation of gold standard testing under constrained availability: application to assessment of HIV treatment failure," *Journal of the American Statistical Association*, vol. 108, no. 504, pp. 1173–1188, 2013.

[8] J. S. Lee, E. Paintsil, V. Gopalakrishnan, and M. Ghebremichael, "A comparison of machine learning techniques for classification of HIV patients with antiretroviral therapy-induced mitochondrial toxicity from those without mitochondrial toxicity," *BMC medical research methodology*, vol. 19, no. 1, pp. 1–10, 2019.

[9] NIH, "First-Line Therapy," *www.clinicalinfo.hiv.gov*, 2021. https://clinicalinfo.hiv.gov/en/glossary/first-line-therapy (accessed Mar. 15, 2021).

[10] Z. Khalid and O. U. Sezerman, "Prediction of HIV drug resistance by combining sequence and structural properties," *IEEE/ACM transactions on computational biology and bioinformatics*, vol. 15, no. 3, pp. 966–973, 2016.

[11] O. Tarasova *et al.*, "A computational approach for the prediction of treatment history and the effectiveness or failure of antiretroviral therapy," *International journal of molecular sciences*, vol. 21, no. 3, p. 748, 2020.

[12] M. L. Petersen, M. J. Van Der Laan, S. Napravnik, J. J. Eron, R. D. Moore, and S. G. Deeks, "Long term consequences of the delay between virologic failure of highly active antiretroviral therapy and regimen modification," *AIDS (London, England)*, vol. 22, no. 16, p. 2097, 2008.

[13] W. Teshome and A. Tefera, "Detection of immunological treatment failure among HIV infected patients in Ethiopia: a retrospective cohort study," *BMC immunology*, vol. 16, no. 1, pp. 1–7, 2015.

[14] WHO and Unicef, "Towards universal access: Scaling up priority HIV/AIDS interventions in the health sector," in *Towards universal access: scaling up priority HIV/AIDS interventions in the health sector*, 2009, pp. 162–162.

[15] A. Brnabic and L. M. Hess, “Systematic literature review of machine learning methods used in the analysis of real-world data for patient-provider decision making,” *BMC medical informatics and decision making*, vol. 21, no. 1, pp. 1–19, 2021.

[16] M. Supriya and A. J. Deepa, “Machine learning approach on healthcare big data: A review,” *Big Data and Information Analytics*, vol. 5, no. 1, pp. 58–75, 2020.

[17] M. E. Ekpenyong, P. I. Etebong, and T. C. Jackson, “Fuzzy-multidimensional deep learning for efficient prediction of patient response to antiretroviral therapy,” *Heliyon*, vol. 5, no. 7, p. e02080, 2019.

[18] M. Kebede, D. T. Zegeye, and B. M. Zeleke, “Predicting CD4 count changes among patients on antiretroviral treatment: Application of data mining techniques,” *Computer methods and programs in biomedicine*, vol. 152, pp. 149–157, 2017.

[19] P. A. Idowu and J. A. Balogun, “Development of a Classification Model for CD4 Count of HIV Patients Using Supervised Machine learning algorithms: A Comparative Analysis,” in *Computational Models for Biomedical Reasoning and Problem Solving*, IGI Global, 2019, pp. 149–176.

[20] G. S. Lakshmi and P. I. Devi, “Prediction of anti-retro viral for HIV and STD patients using data mining technique,” in *2017 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS)*, 2017, pp. 1–6.

[21] G. B. Tarekegn and V. Sreenivasarao, “Application of data mining techniques on pre art data: The case of Felege Hiwot Referral Hospital,” *International Journal of Research Studies in Computer Science and Engineering*, vol. 3, no. 2, pp. 1–9, 2016.

[22] A. Rivero-Juárez *et al.*, “Using machine learning methods to determine a typology of patients with HIV-HCV infection to be treated with antivirals,” *Plos one*, vol. 15, no. 1, p. e0227188, 2020.

[23] B. Olatosi, J. Zhang, S. Weissman, J. Hu, M. R. Haider, and X. Li, “Using big data analytics to improve HIV medical care utilisation in South Carolina: A study protocol,” *BMJ open*, vol. 9, no. 7, p. e027688, 2019.

[24] I. P. Adebayo, A. Olutola, and T. A. Aladekomo, “The prediction of paediatric hiv/aids patient survival: A data mining approach,” *Asian Journal of Computer and Information Systems*, vol. 4, no. 3, 2016.

[25] C. Olayemi Olufunke, O. Olasehinde Olayemi, and O. Agbelusi, “Predictive model of pediatric hiv/aids survival in nigeria using support vector machine,” *Communications*, vol. 5, pp. 29–36, 2016.

[26] C. Schröer, F. Kruse, and J. M. Gómez, “A systematic literature review on applying CRISP-DM process model,” *Procedia Computer Science*, vol. 181, pp. 526–534, 2021.

[27] J. Nega, S. Taye, Y. Million, C. Rodrigo, and S. Eshetie, “Antiretroviral treatment failure and associated factors among HIV patients on first-line antiretroviral treatment in Sekota, northeast Ethiopia,” *AIDS Research and Therapy*, vol. 17, no. 1, pp. 1–9, 2020.

[28] C.-H. Cheng and Y.-C. Wang, “Exploring the important attributes of Human Immunodeficiency Virus and generating decision rules,” *Symmetry*, vol. 12, no. 1, p. 67, 2020.

[29] D. Mulisa, M. Tesfa, G. Mullu Kassa, and T. Tolossa, “Determinants of first line antiretroviral therapy treatment failure among adult patients on ART at central Ethiopia: un-matched case control study,” *BMC Infectious Diseases*, vol. 19, no. 1, pp. 1–13, 2019.

[30] M. A. Assemie, M. Alene, D. B. Ketema, and S. Mulatu, “Treatment failure and associated factors among first line patients on highly active antiretroviral therapy in Ethiopia: A systematic review and meta-analysis,” *Global health research and policy*, vol. 4, no. 1, pp. 1–10, 2019.

[31] M. Maskew *et al.*, “Machine learning to predict retention and viral suppression in South African HIV treatment cohorts,” *medRxiv*, 2021.

[32] WHO, Consolidated guidelines on HIV prevention, testing, treatment, service delivery and monitoring: Recommendations for a public health approach. World Health Organization, 2021.

[33] WHO, “HIV Country Profiles,” *cvshivci.org*, 2021. https://cfs.hivci.org/country-factsheet.html (accessed Mar. 15, 2021).

[34] NIH, “The Stages of HIV Infection,” *www.clinicalinfo.hiv.gov*, 2021. https://hivinfo.nih.gov/understanding-hiv/fact-sheets/stages-hiv-infection (accessed Mar. 15, 2021).

[35] A. Abdissa *et al.*, “Drug resistance in HIV patients with virological failure or slow virological response to antiretroviral therapy in Ethiopia,” *BMC infectious diseases*, vol. 14, no. 1, pp. 1–7, 2014.

[36] Z. El-Khatib *et al.*, “Adherence to drug-refill is a useful early warning indicator of virologic and immunologic failure among HIV patients on first-line ART in South Africa,” *PloS one*, vol. 6, no. 3, p. e17518, 2011.

[37] J. Estill *et al.*, “The estimated need of second-line antiretroviral therapy in Sub-Saharan Africa 2015-2030: Mathematical modelling study,” in *JOURNAL OF THE INTERNATIONAL AIDS SOCIETY*, 2015, vol. 18.

[38] M. A. Selialia and L. B. Poulo, “Applications of data mining in the health sector in Lesotho: A preliminary study,” in *2015 IST-Africa Conference*, 2015, pp. 1–8.

[39] A. Waruru *et al.*, “Positive predictive value of the WHO clinical and immunologic criteria to predict viral load failure among adults on first, or second-line antiretroviral therapy in Kenya,” *PloS one*, vol. 11, no. 7, p. e0158881, 2016.

[40] G. W. Rutherford *et al.*, “Predicting treatment failure in adults and children on antiretroviral therapy: a systematic review of the performance characteristics of the 2010 WHO immunologic and clinical criteria for virologic failure,” *Aids*, vol. 28, pp. S161–S169, 2014.

[41] D. Mesafint and D. H. Manjaiah, “Predictive models of hiv/aids epidemic status using data mining techniques: A review,” 2019.

[42] A. Oliveira, B. M. Faria, A. R. Gaio, and L. P. Reis, “Data mining in HIV-AIDS surveillance system,” *Journal of medical systems*, vol. 41, no. 4, pp. 1–12, 2017.

[43] P. A. Idowu, T. A. Aladekomo, O. Agbelusi, and C. O. Akanbi, “Prediction of pediatric HIV/Aids survival in Nigeria using naïve Bayes’ approach,” *Journal of Computer Science and Its Application*, vol. 25, no. 1, pp. 1–15, 2018.

[44] M. Li, “study on the grouping of patients with chronic infectious diseases based on data mining,” *Journal of Biosciences and Medicines*, vol. 7, no. 11, pp. 119–135, 2019.

[45] A. Ashfaq, “Predicting clinical outcomes via machine learning on electronic health records,” PhD Thesis, Halmstad University Press, 2019.

[46] T. G. Hailu, “Comparing data mining techniques in HIV testing prediction,” *Intelligent Information Management*, vol. 7, no. 03, p. 153, 2015.

[47] K. J. Cios, W. Pedrycz, R. Swiniarski, and L. Kurgan, “Data Mining: A Knowledge Discovery Approach.” Springer Publishing Company, Incorporated.

[48] D. W. Gunda, B. R. Kidenya, S. E. Mshana, S. B. Kilonzo, and B. C. Mpondo, “Accuracy of WHO immunological criteria in identifying virological failure among HIV-infected adults on first-line antiretroviral therapy in Mwanza, North-western Tanzania,” *BMC research notes*, vol. 10, no. 1, pp. 1–5, 2017.

[49] J. K. Rohr *et al.*, “Developing a predictive risk model for first-line antiretroviral therapy failure in South Africa,” *Journal of the International AIDS Society*, vol. 19, no. 1, p. 20987, 2016.

[50] S. Yang, X. Zhengy, and F. Yuan, “A patient outcome prediction based on Random Forest,” in *Proceedings of the 4th International Conference on Crowd Science and Engineering*, 2019, pp. 220–227.

APPENDIX

Appendix 1: Confusion Matrix of J48 Algorithm with test option of Cross-validation at (default values)

```
=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.990      1.000    0.995      ?        0.486     0.990     Low
                 0.000    0.000    ?          0.000    ?          ?        0.486     0.010     High
Weighted Avg.    0.990    0.990    ?          0.990    ?          ?        0.486     0.980

=== Confusion Matrix ===

    a    b   <-- classified as
 5733    0 |    a = Low
   58    0 |    b = High
```

Appendix 2: Confusion Matrix of Naïve Bayes Algorithm with test option of Cross-Validation

```
=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.990      1.000    0.995      ?        0.704     0.995     Low
                 0.000    0.000    ?          0.000    ?          ?        0.704     0.024     High
Weighted Avg.    0.990    0.990    ?          0.990    ?          ?        0.704     0.985

=== Confusion Matrix ===

    a    b   <-- classified as
 5733    0 |    a = Low
   58    0 |    b = High
```

Appendix 3: Confusion Matrix of Support Vector Machine with test option of Cross-Validation

```
=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                1.000    1.000    0.990      1.000    0.995      ?        0.500     0.990     Low
                0.000    0.000    ?          0.000    ?          ?        0.500     0.010     High
Weighted Avg.   0.990    0.990    ?          0.990    ?          ?        0.500     0.980

=== Confusion Matrix ===

   a    b   <-- classified as
5733    0 |    a = Low
  58    0 |    b = High
```

Appendix 4: Confusion Matrix of Random Forest Algorithm with test option of Cross-Validation

```
=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.990      1.000    0.995      ?        0.705     0.995     Low
                 0.000    0.000    ?          0.000    ?          ?        0.705     0.021     High
Weighted Avg.    0.990    0.990    ?          0.990    ?          ?        0.705     0.985

=== Confusion Matrix ===

    a    b   <-- classified as
 5733    0 |    a = Low
   58    0 |    b = High
```