

## **Application of Data Mining Techniques on Pre ART Data: The Case of Felege Hiwot Referral Hospital**

**Getaneh Berie Tarekegn**

MSC, Department of Computer Science  
Faculty of Engineering and Technology  
Assosa University, Assosa, Ethiopia  
getaneh\_berie@yahoo.com

**Dr.Vuda Sreenivasarao**

Associated Professor  
Department of ISE, School of Computing  
Adama Science and Technology University,  
Adama, Ethiopia  
vudasrinivasaro@gmail.com

---

**Abstract:** *HIV/AIDS has claimed the lives of millions and has left behind hundreds of thousands of orphans in Ethiopia. The government of Ethiopia took several steps in preventing further disease spread, and in increasing accessibility to HIV care, treatment and support for persons living with HIV. Antiretroviral therapy (ART) is one of the treatments given to HIV patients Felege Hiwot Referral Hospital (FHRH) to restore patients with severe disease to healthy.*

*The dataset for the study contains pre ART records of the year 2005 and 2006 E.C produced by the ART office of patients Felege Hiwot Referral Hospital. The dataset has been utilized for the purpose of predicting clients' eligibility for ART.*

*Before these data has been used for the purpose of classification a number of pre-processing steps such as data cleaning, data reduction and data transformation have been effectively used which helped in achieving the objective finally or to increase the speed and efficiency of mining process.*

*The final goal of this paper is to build ART eligibility predictive model that helps to deciding whether HIV positive individual should start Anti-retroviral treatment or not. For building ART eligibility predictive model, Naive Bayesian Classifier and J48 Decision Tree Classifier are used.*

*After experimenting J48 decision tree and Naive Bayesian classifier using both 10-fold cross validation and percentage split (66%) test modes, J48 classifier using 10-fold cross validation that performs well and can be used as a best predicting model algorithm than Naive Bayesian classifier in predicting clients' eligibility for ART is created.*

**Keywords:** *ART, Cluster Differentiation 4 Count, Data Mining, J48 Decision Tree, Naive Bayesian Classifier, WEKA, 10-Fold Cross Validation*

---

### **1. INTRODUCTION**

Before some years, having HIV/AIDS was almost equivalent to a “death sentence”. Since 1996, with the introduction of combined anti-retroviral treatment, HIV/AIDS is transferred from “virtual death sentence” to chronic, manageable disease [1].

Anti-retroviral treatment (ART) is the main type of treatment for HIV/AIDS by using drugs. The drugs do not kill or cure the virus, but it can stop people living with the virus from becoming ill for many years by preventing the growth of the virus [2]. ART has significantly reduced HIV-associated morbidity and mortality and has transformed the disease (HIV/AIDS) into a chronic, controllable condition. In addition, effective treatment of HIV-infected individuals with ART is highly effective at preventing transmission to sexual partners.

The treatment consists of drugs that have to be taken every day for the rest of someone's life. But anyone who is HIV positive is not going to start the ART [3]. Some aspects like CD4Count, age, WHO clinical staging, should be considered whether someone is eligible to start using the ART or not [4].

The Felege Hiwot Referral Hospital (FHRH) is located in the north western part of Ethiopia, in a city called Bahir Dar. Felege Hiwot Referral Hospital is a government hospital in Bahir Dar which is one

of the five referral hospitals in Amhara Regional State of Ethiopia. The hospital provides general outpatient service, surgical and obstetric emergency and regular services, and general medical and pediatric in-patient services, voluntary counseling and testing (VCT) etc. ART is also one of the services which are being provided by the hospital.

The remaining part of this paper is organized as follows. Section two discusses the present system, in section three we present methodology of the study and section four discusses the proposed solution. In section five and six discusses experimental results evaluation of the model. Section seven presents conclusion.

## **2. PRESENT SYSTEM**

The effort for preventing HIV/AIDS ranges from behavioral intervention to introduction of the ART program. ART has dramatically improved the livelihood of people living with HIV/AIDS. World Health Organization (WHO) recommends optimum time for initiating ART should be guided by CD4 (Cluster Differentiation 4) count, pregnancy, age and clinical staging.

But at the Felege Hiwot Referral Hospital currently, it is the duty of the physician to decide whether the patient should start ART or not. The amount of time required to process all the records by considering the attributes listed above for each and every record is large and the process is somewhat tiresome. Thus there comes a need to have some help to reduce the work overload of the physician by using electronic devices to make the decisions by referring to the values of the attributes stated above in context with the decision made as the values of the variable meets some condition.

The main objective of this study is to develop ART eligibility predictive model in support of deciding whether HIV positive individual should start Anti-retroviral treatment or not.

## **3. METHODOLOGY**

In order to realize the objectives of this project, the following methodology is used.

### **3.1. Literature Review**

It is vital to review other related literatures to have a deep knowledge on the domain area and the problem solving methods. Different books, journal articles, conference proceedings and Internet resources are reviewed in detail to get better understanding of the application domain and to have detail knowledge on the various techniques of data mining.

### **3.2. The KDD Process**

According to [6], KDD is the nontrivial process of identifying valid, novel, potentially useful and ultimately understandable patterns in data. For conducting this project KDD model was used. **KDD process** is commonly defined with the following stages:

- Selection: It involves collecting data from possible sources to get the target data
- Preprocessing: It concerns with cleaning the target data so as to get preprocessed data.
- Transformation: It is all about any sort of rearrangements of the preprocessed data so that it would be easy to apply the data mining tool.
- Data mining: It involves applying appropriate data mining techniques and algorithms onto the transformed data to get interesting patterns of the data.
- Interpretation/Evaluation: At this stage of the KDD process, the discovered pattern is evaluated to give recommendations [6].

### **3.3. Data Collection**

The dataset used for conducting this experiment is collected from Felege Hiwot Referral Hospital database. Around 1512 records were taken from the MS Access database which is collected in the year 2005 and 2006 E.C for their own purpose, and changed into the excel spread sheet format. This dataset contains HIV patient information who visited the Hospital on those mentioned years for the purpose of ART initiation.

### **3.4. Implementation Tool**

Before doing a project, selecting the best tool which fits the intended goal is an essential task.

According to [7], WEKA has achieved widespread acceptance within academia and business circles, and has become a widely used tool for data mining research. For this specific project, WEKA 3.6 was

used to build a model and Excel 2010 was used for some of preprocessing tasks such as data cleaning (e.g. for missing value identification).

The main strengths WEKA is freely available under the GNU General Public License, very portable because it is fully implemented in the Java programming language and 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 [8].

#### **4. PROPOSED SOLUTION**

##### **4.1. Data Preprocessing**

Today’s real-world data are highly susceptible to noisy, missing, and inconsistent data due to their typically huge size and their likely origin from multiple, heterogenous sources. Low-quality data will lead to low-quality mining results [9]. Therefore, improvement of the data quality is necessary in order to enhance the efficiency and ease of the mining process.

After collecting the ART data from the Felege Hiwot Referral Hospital I tried to get the general property of the data, missing values, noisy data and the degree of duplicates through descriptive data summarization. And then, in order to improve the quality of the data and, consequently, of the mining results we applied the following data preprocessing tasks.

##### **4.2. Data Cleaning**

Data cleaning refers to identifying incomplete, incorrect, inaccurate, irrelevant, etc. parts of the data and then replacing, modifying, or deleting this dirty data or coarse data[10]. During data cleaning phase, I identified some incomplete and noisy attribute values in the original dataset. Out of 1512 records, there are 86 records with missing class label values and those records are totally removed from the original dataset. By consulting domain expert, 12 tuples with so many missing values are also removed. WHOStage, CD4Count and CD4Percent are among the attributes with missing value. As the domain expert recommended us, we filled the missing value of those attributes by zero (0). There are also some attributes in the original dataset containing values that are incorrect or invalid or doesn’t give any sense. For example, there is a patient whose sex is “male” and pregnant is “true” and there a patient, whose sex is “female”, age is between “2 & 5” years and pregnant is “true”. I modified such records. Finally, at the end of data cleaning process, I select 1414 records for further processing.

##### **4.3. Data Integration**

Data mining often requires data integration - the merging of data from multiple data sources. In my case we didn’t apply data integration. Because the dataset for this project is collected from a single data source which is from Felege Hiwot Referral Hospital pre ART database.

##### **4.4. Data Reduction**

Complex data analysis and mining on huge amounts of data can take a long time, making such analysis impractical or infeasible. Therefore, apply data reduction techniques to obtain a reduced representation of the dataset that is much smaller in volume, yet produce the same (or almost the same) analytical results.

The collected dataset contains 1414 instances, which is not as such a large data. Therefore, it is not necessary to apply numerosity or size reduction on this dataset. The original dataset has 30 attributes. However, the dataset contains some attributes which are irrelevant for the intended data mining task. To remove those irrelevant attributes, relevance analysis has been performed on the dataset by involving domain expert. According to [11][12] and domain expert recommendation, ChronicYear, Gender, Age, Pregnant, WHOStage, CD4Count, EligibilityForART and CD4Percent are important attribute to build a model that predicts HIV patient ART eligibility. The rest of the attributes are removed from the data set. At this stage a dataset containing 1414 instances and 8 useful attributes is selected for further processing. Table 1 shows the description of the selected attributes.

**Table1.** Description of the selected attributes.

Attribute Name	Description	Attribute Type	Attribute Value
ChronicYear	Year the patient visited ART	Nominal	Two values:

	clinic for first time		<ul style="list-style-type: none"> <li>• 2005 and 2006</li> </ul>
Age	Patient age	Nominal	Four Values:- <ul style="list-style-type: none"> <li>• &lt;2 years</li> <li>• [2,5) years</li> <li>• [5,15) years</li> <li>• &gt;=15 years</li> </ul>
Gender	Patient gender	Nominal	Two values:- <ul style="list-style-type: none"> <li>• Male and Female</li> </ul>
Pregnant	Patient having young developing inside womb	Nominal	Two values:- <ul style="list-style-type: none"> <li>• True and False</li> </ul>
WHOStage	Stages of HIV infection	Nominal	Five values:- <ul style="list-style-type: none"> <li>• Default 0</li> <li>• WHOStage 1</li> <li>• WHOStage 2</li> <li>• WHOStage 3</li> <li>• WHOStage 4</li> </ul>
CD4Count	The number of CD4 cells	Numeric	Non-zero positive number
CD4Percent	The percentage CD4 cells	Numeric	Non-zero positive number
EligibilityForART	Antiretroviral Therapy initiation	Nominal	Two values:- <ul style="list-style-type: none"> <li>• Yes and No</li> </ul>

#### 4.5. Data Transformation

Data transformation plays an important role in modifying data into forms more appropriate for mining. Based on WHO Antiretroviral Therapy recommendation for HIV infection in adults and adolescents [11], and infants and children [12], This paper discretized age value into four groups manually.

- Patient less than two years grouped into < 2
- Patient between two and five years grouped into [2,5)
- Patient between five and fifteen years grouped into [5,15)
- Patient greater than or equal to fifteen years grouped into >=15

ChronicYear and WHOStage which is detected as numeric attribute types by WEKA are transformed into nominal attribute type.

### 5. EXPERIMENTATION

In this study an attempt is made to explore ART services data to identify regular patterns in order to determine clients' eligibility for Antiretroviral Therapy. Experiments have been carried to identify classification model. The purpose of experiments is to find model that is able to predict the eligibility of the clients' for ART as Yes or No taking selected attributes as inputs. Identifying the determinant factors of ART initiation can support the physician to decide whether the client is eligible for ART or not.

In order to train the classifiers 1414 data are used for training and testing. For building ART eligibility predictive model, Naive Bayesian Classifier and J48 Decision Tree Classifier are used.

A Naive Bayesian classifier [13] is a simple probabilistic classifier based on applying Bayesian theorem (from Bayesian statistics) with strong (naive) independence assumptions. J48 is a simple C4.5 decision tree, it creates a binary tree. C4.5 builds decision trees from a set of training data which is like an ID3, using the concept of information entropy [14].

A total of four experiments are done. To evaluate the performance of the model; 10-fold cross validation and Percentage split (66%) are used. In the case 10-fold cross validation, the data are randomly partitioned equally into ten parts. The learning scheme is trained ten times using nine-tenths of the total data and the remaining is used for testing. Therefore the learning procedure is executed a total of 10 times on different training and testing sets. In the case percentage split, the learning scheme is trained using 66% of the total data and the remaining is used for testing.

In all experiments, 8 selected attributes (ChronicYear, Gender, AgeRange, Pregnant, WHOStage, CD4CountAdult, CD4PercentChild and EligibleForART) are used. Out of these attributes, the class or

dependent attribute is EligibleForART .The experiment is done using WEKA mining tool version 3.6.11. The tool takes the data in arff format in a single table. Before this the prepared data in excel format is changed to csv format.

In order to build a model this arff file is given to the classifiers. The frontend of this file is shown in figure 1.

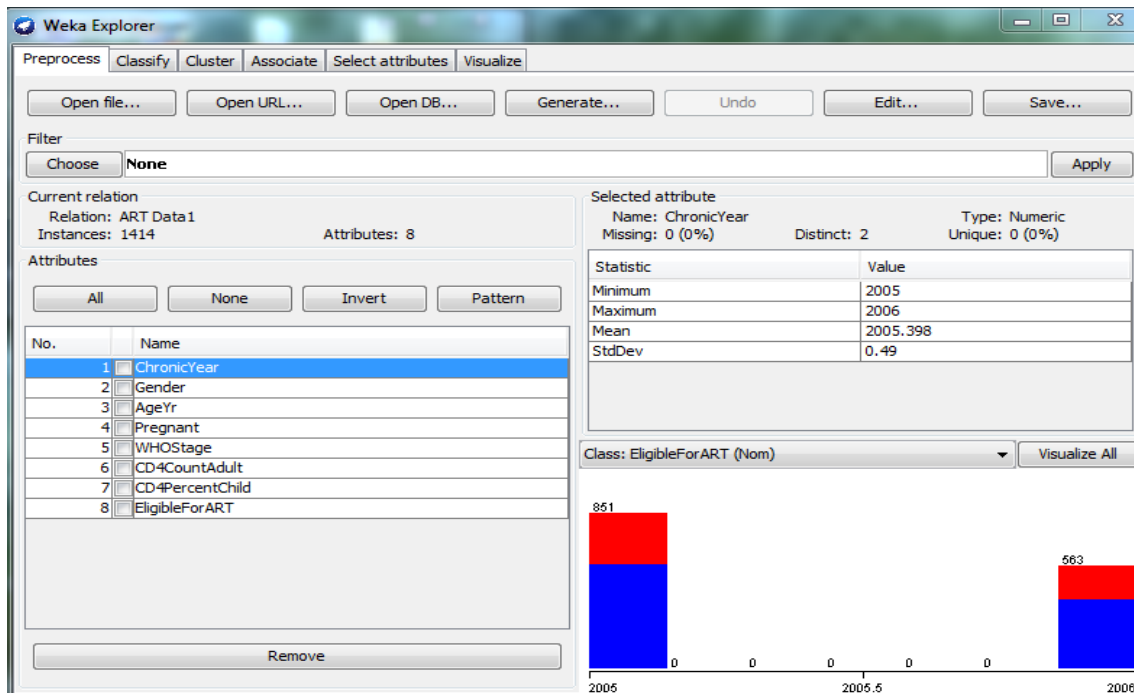


Figure1. WEKA Explorer Windows Showing the Number of Attribute and Instances.

### 5.1. Experiment I

To build the predictive model, J48 decision tree algorithm is trained and evaluated using 10-fold cross validation. For this experiment the default parameter values of J48 decision tree algorithm are used except the minimum number of instances per leaf which is modified to one (1) and unpruned which is modified to True. Figure 2 presents the summarized output of this model.

```

==== Run information ====
Scheme: weka.classifiers.trees.J48 -U -M 1
Attributes: 8
Test mode: 10-fold cross-validation
J48 unpruned tree
Number of Leaves : 11
Size of the tree : 17
Time taken to build model: 0.02 seconds

--- Stratified cross-validation ---
=== Summary ===
Correctly Classified Instances      1355      95.8274 %
Incorrectly Classified Instances     59        4.1726 %
Kappa statistic                    0.9052
Mean absolute error                 0.078
Root mean squared error             0.2025
Relative absolute error             17.5904 %
Root relative squared error        43.0194 %
Total Number of Instances          1414

==== Detailed Accuracy By Class ====
                TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
Weighted Avg.   0.976    0.077    0.962     0.976    0.969     0.943    Yes
                 0.923    0.024    0.95     0.923    0.936     0.943    No

==== Confusion Matrix ====
  a  b  | <-- Classified as
922 23 | a = Yes
 36 433 | b = No
    
```

Figure2. Summary of the outputs of unpruned J48 Decision using 10-fold cross-validation test mode.

### 5.2. Experiment II

To build the predictive model, J48 decision tree algorithm is trained and evaluated using Percentage Split: 66% of the dataset is used for training set and the remaining dataset is used for test set. For this experiment the default parameter values of J48 decision tree algorithm are used except the minimum number of instances per leaf which is modified to one (1) and unpruned which is modified to True. Figure 3 presents the summarized output of this model.

```

=== Run information ===

Scheme:weka.classifiers.trees.J48 -U -M 1
Instances: 1414
Attributes: 8
Test mode: split 66.0% train, remainder test
J48 unpruned tree
Number of Leaves : 11
Size of the tree : 17
Time taken to build model: 0.02 seconds

=== Evaluation on test split ===
=== Summary ===

Correctly Classified Instances 456 94.8025 %
Incorrectly Classified Instances 25 5.1975 %
Kappa statistic 0.8855
Mean absolute error 0.0839
Root mean squared error 0.2251
Relative absolute error 18.7826 %
Root relative squared error 47.1466 %
Total Number of Instances 481

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
          0.962    0.077    0.959    0.962    0.96       0.933    Yes
Weighted Avg.  0.923    0.038    0.928    0.923    0.925    0.933    No
          0.948    0.064    0.948    0.948    0.948    0.933

=== Confusion Matrix ===

  a  b  <-- classified as
301 12 | a = Yes
 13 155 | b = No
    
```

**Figure3.** Summary of the outputs of unpruned J48 Decision tree using Percentage split (66%) test mode.

### 5.3. Experiment III

To build the predictive model, Naive Bayesian classifier is trained and evaluated using 10-fold cross validation. For this experiment the default parameter values of Naive Bayesian classifier are used. Figure 4 presents the summarized output of this model.

```

=== Run information ===

Scheme:weka.classifiers.bayes.NaiveBayes
Attributes: 8
Test mode: 10-fold cross-validation
Naive Bayes Classifier
Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances 1324 93.6351 %
Incorrectly Classified Instances 90 6.3649 %
Kappa statistic 0.8566
Mean absolute error 0.1179
Root mean squared error 0.2578
Relative absolute error 26.5907 %
Root relative squared error 54.7498 %
Total Number of Instances 1414

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
          0.951    0.094    0.953    0.951    0.952    0.938    Yes
Weighted Avg.  0.906    0.049    0.902    0.906    0.904    0.938    No
          0.936    0.079    0.936    0.936    0.936    0.938

=== Confusion Matrix ===

  a  b  <-- classified as
899 46 | a = Yes
 44 425 | b = No
    
```

**Figure4.** Summary of the outputs of Naive Bayesian classifier using 10-fold cross-validation test mode.

### 5.4. Experiment IV

To build the predictive model, Naive Bayesian classifier is trained and evaluated using Percentage Split: 66% of the dataset is used for training set and the remaining dataset is used for test set. For this experiment the default parameter values of Naive Bayesian classifier are used. Figure 5 presents the summarized output of this model.

```

=== Run information ===

Scheme:weka.classifiers.bayes.NaiveBayes
Instances: 1414
Attributes: 8
Test mode: split 66.0% train, remainder test
Naive Bayes Classifier
Time taken to build model: 0.02 seconds
=== Evaluation on test split ===
=== Summary ===

Correctly Classified Instances      435          90.4366 %
Incorrectly Classified Instances    46           9.5634 %
Kappa statistic                    0.7936
Mean absolute error                 0.1389
Root mean squared error             0.2902
Relative absolute error             31.1048 %
Root relative squared error         60.7758 %
Total Number of Instances          481

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
                0.904   0.095   0.946     0.904   0.925     0.924   Yes
                0.905   0.096   0.835     0.905   0.869     0.924   No
Weighted Avg.   0.904   0.095   0.908     0.904   0.905     0.924

=== Confusion Matrix ===

  a  b  <-- classified as
283 30 |  a = Yes
 16 152 |  b = No
    
```

Figure 5. Summary of the outputs of Naïve Bayesian classifier using Percentage split (66%) test mode.

## 6. EVALUATION OF THE MODEL

The outputs of Experiment I, II, III and IV are analyzed and evaluated in terms of the details of the confusion matrix of the model.

From the output of Experiment I, as shown in figure, the selected classifier registers 95.8% accuracy. The confusion matrix shows that, out of the total 945 actually eligible for ART clients, only 922 (97.6%) clients are classified as Eligible for ART and the rest are misclassified as not eligible for ART. And out of the total 469 actually not eligible for ART clients, only 433 (92.3%) clients are classified as not eligible for ART and the rest are misclassified as eligible for ART negative. This means the model has better performance in terms of correctly classifying eligible for ART than not eligible for ART.

From the output of Experiment II, as shown in figure, the selected classifier registers 94.8% accuracy. The confusion matrix shows that, out of the total 313 actually eligible for ART clients, only 301 (96.5%) clients are classified as Eligible for ART and the rest are misclassified as not eligible for ART. And out of the total 168 actually not eligible for ART clients, only 155 (92.3%) clients are classified as not eligible for ART and the rest are misclassified as eligible for ART negative. This means the model has better performance in terms of correctly classifying eligible for ART than not eligible for ART.

From the output of Experiment III, as shown in figure, the selected classifier registers 93.6% accuracy. The confusion matrix shows that, out of the total 945 actually eligible for ART clients, only 899 (95.1%) clients are classified as Eligible for ART and the rest are misclassified as not eligible for ART. And out of the total 469 actually not eligible for ART clients, only 425 (90.6%) clients are classified as not eligible for ART and the rest are misclassified as eligible for ART negative. This means the model has better performance in terms of correctly classifying eligible for ART than not eligible for ART.

From the output of Experiment IV, as shown in figure, the selected classifier registers 90.4% accuracy. The confusion matrix shows that, out of the total 313 actually eligible for ART clients, only 283 (90.4%) clients are classified as Eligible for ART and the rest are misclassified as not eligible for ART. And out of the total 168 actually not eligible for ART clients, only 152 (90.5%) clients are classified as not eligible for ART and the rest are misclassified as eligible for ART negative. This means the model has better performance in terms of correctly classifying not eligible for ART than eligible for ART.

## 7. CONCLUSION AND RECOMMENDATION

### 7.1. Conclusion

HIV/AIDS has claimed the lives of millions and has left behind hundreds of thousands of orphans in Ethiopia. The government of Ethiopia took several steps in preventing further disease spread, and in increasing accessibility to HIV care, treatment and support for persons living with HIV. Antiretroviral therapy (ART) is one of the treatments given to HIV patients by Felege Hiwot Referral Hospital to restore patients with severe disease to healthy.

The dataset for the study contains pre ART records of the year 2005 and 2006 E.C produced by the ART office of Felege Hiwot Referral Hospital. The dataset has been utilized for the purpose of predicting clients' eligibility for ART. Raw data which were simply accumulated as result of day to day activities have been utilized for this project to show important patterns in the data.

Before these data has been used for the purpose of classification a number of pre-processing steps such as data cleaning, data reduction and data transformation have been effectively used which helped in achieving the objective finally.

After experimenting J48 decision tree and Naive Bayesian classifier using both 10-fold cross validation and percentage split (66%) test modes, unpruned J48 classifier using 10-fold cross validation that performs classification with 95.8% accuracy in predicting clients' eligibility for ART is created.

Finally, the classification rules obtained from the model reveal that: pregnant women are eligible for ART; if patient CD4 count is less than or equal to 350 and age is greater than or equal to five years then the patient is eligible for ART; all infants (age less than 2 years) are eligible for ART; if age is between two and five years and CD4 percent is less than or equal to 25 then the patient is eligible for ART; if patient age is between two and five years and CD4 count is less than or equal to 750 then the patient is eligible for ART.

### ACKNOWLEDGMENT

The Authors would like to thank Felege Hiwot Referral Hospital by giving the required data set to do this research and development. And also would like to thank the reviewers for their constructive comments.

### REFERENCES

- [1] Degu Jerene Dare. "HIV antiretroviral therapy in Ethiopia: Overcoming implementation challenges" PhD dissertation, Centre for International Health, University of Bergen, Norway, 2006.
- [2] International Association of Providers of AIDS care. "What Is Antiretroviral Therapy (ART)?" Internet: [http://www.aidsinfonet.org/fact\\_sheets/view/403](http://www.aidsinfonet.org/fact_sheets/view/403)
- [3] National AIDS control organization. "Antiretroviral Therapy Guidelines for HIV-Infected Adults and Adolescents Including Post-exposure Prophylaxis", New Delhi, India: Ministry of Health & Family Welfare Government of India, 2006.
- [4] WHO. "Anti-retroviral therapy for HIV infections in adults: Recommendations for public health approach." Geneva, Switzerland: Who press, 2006, pp 13-24.
- [5] Federal HIV/AIDS Prevention and Control Office. "Guidelines for implementation of the antiretroviral therapy program in Ethiopia." Addis Ababa, Ethiopia: Federal Ministry of Health, 2006.
- [6] Fayyad, Usama; Piatetsky-Shapiro, Gregory; Smyth, Padhraic (1996). "From Data Mining to Knowledge Discovery in Databases". Retrieved 17 December 2008.
- [7] M. Hall, et al., "The WEKA Data Mining Software: An Update" *SIGKDD Explorations*, vol. 11, no. 1, pp. 10-18.
- [8] S. Krishnaveni and M. Hemalatha, "A Perspective Analysis of Traffic Accident using Data Mining Techniques" *International Journal of Computer Applications*, vol. 23, no. 7, pp. 40-48, Jun 2011.
- [9] J. Han and M. Kamber, *Data Mining: Concepts and Techniques, 2ed.* San Francisco: Morgan Kaufmann, 2006.



- [10] S. Wu (2013), "A review on coarse warranty data and analysis", *Reliability Engineering and System v.114*: 1–11.
- [11] WHO, Antiretroviral therapy for HIV infection in adults and adolescents: recommendations for a public health approach.2010 rev.
- [12] WHO, Antiretroviral therapy of HIV infection in infants and children: towards universal access: recommendations for a public health approach - 2010 rev.
- [13] Rish, Irina. (2001). "An empirical study of the naive Bayes classifier". IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence.
- [14] Quinlan, J. R. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, 1993.

**AUTHORS' BIBLIOGRAPHY**



**Getaneh Berie Tarekegn**, received his B.Sc. Degree in Computer Science from University of Gondar (UoG). Currently perusing.M.Sc. in Computer Science, School of Computing and Electrical Engineering, IOT, Bahir Dar University, Ethiopia. His main research interest is Neural Network, Data Mining and Big Data.



**Dr. Vuda Sreenivasarao**, received his M.Tech degree in computer science and engineering from Sathyabama University from 2007.He received PhD degree in computer science and engineering from Singhania University, Rajasthan, India from 2010. Currently working as Associated Professor in School of Computing, Adama Science and Technology University, Adama, Ethiopia.

His main research interests are Data mining, Fuzzy logic, Mobile communication and Network Security. He has got 16 years of teaching experience. He has published more than 35 research papers in various international journals and one Springer international conference paper. He has 115 Editorial Board / Reviewers memberships in various international journals. He is a life member of various professional societies like IEEE, ACM, MAIRCC, MCSI, SMIACSIT, MIAENG, MCSTA, MAPSMS, MSDIWC, SMSCIEI, SNMUACEE and MISTE.