EasyChair Preprint

№ 2279

Predictive Models of HIV/AIDS Epidemic Status using Data Mining Techniques: A Review

Daniel Mesafint and D.H. Manjaiah

December 31, 2019

Predictive Models of HIV/AIDS Epidemic Status using Data Mining Techniques: A Review

Daniel Mesafint
Department of Computer Science
Mangalore University
Mangalore, India
mesafintdaniel5@gmail.com

Manjaiah D.H
Department of Computer Science
Mangalore University
Mangalore, India
drmdhmu@gmail.com

***Abstract*—Data Mining plays an important role for uncovering new trends in health care organization which in turn helpful for all the parties associated with this field. Data mining is important for the health care sector in identification and detection of diseases, help researchers to make effective health care policies, develop recommendation systems and health profiles for patients. There are difficulties in evaluating the large data generated in the health care sector that are used to discover knowledge and find patterns for decision making. Health care data needs to be analyzed accurately in diagnosis, management and treatment of diseases. In this paper, we reviewed data mining techniques, its processes, tools, related works in HIV/AIDS and health care system. The purpose of this paper is to provide an insight towards requirements of health domain and about suitable choice of available technique and to understand about data mining and its importance in health care organizations.**

***Index Terms*—Data mining, KDD, SEMMA, CRISP-DM, Classification, Prediction, Clustering, Association, Neural networks, HIV/AIDS.**

I. Introduction

AIDS is a chronic and potentially most threatening infectious disease caused by human immunodeficiency virus in the 21st century. 78 million people were estimated to be suffering from HIV/AIDS and 35 million people have died since the start of the epidemic year but 36.7 million people were reported as HIV/AIDS infected and 1.1 million people have died in 2015 globally; 2.1 million people were found to be newly HIV/AIDS infected globally. [1]

The epidemiological estimates of HIV/AIDS infection and the mortality rate with this disease are crucial for planning and monitoring of trends at the national, regional, and worldwide level. Continued effort is mandatory to design a better way of improving the validity of the estimates and developing public health policy. [2]

Having deep knowledge about HIV epidemic status helps to develop solutions in different ways. Information system that reveals the characteristics, magnitude, distribution, causes and consequences of the status of disease is required in order to design programs, strategies and final solutions for the problems of disease. Gaining enough knowledge on the problem area is the step to develop a solution for it. The devised solutions shall be more effective since they are supported by appropriate information.

New techniques and methods are required to evaluate, analyze, search and discover new hidden patterns and relationships in large database. As a result, the discipline of knowledge discovery or data mining in data bases, which deals with the study of such tools and techniques, has evolved into an important and active area of research. [3]

The remainder of the paper is structured as follows: Section I provides the data mining analytic with methodologies. Section II provides an insight of the data mining algorithms and techniques. In other words this section provides summary of findings with discussion in some detail. Section III provides review of literature pertaining to health data mining, especially on mining HIV/AIDS status. Section IV provides conclusions.

A. OVERVIEW OF DATA MINING:

Data mining refers to extracting or mining knowledge from large amounts of data. The term is actually a misnomer. It is the computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. [4]

Data mining is a logical process that is used to search through large amount of data in order to find useful data. The goal of this technique is to find patterns that were previously unknown. Once these patterns are found they can further be used to make certain decisions for development of their businesses.

1) METHODOLOGIES OF DATA MINING::

- **KDD Process:** Refers to the broad process of finding knowledge in data, and emphasizes the ”high-level” application of particular data mining methods. It is of interest to researchers in machine learning, pattern recognition, databases, statistics, artificial intelligence, knowledge acquisition for expert systems, and data visualization. [5]

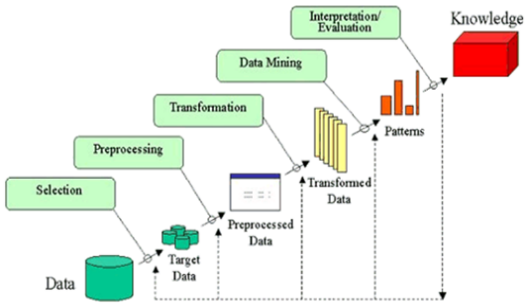

Fig. 1. Knowledge Discovery Process [5]

- **SEMMA:** Is an acronym used to describe the SAS data mining process. It stands for Sample, Explore, Modify, Model, and Assess. SAS Enterprise Miner nodes are arranged on tabs with the same names. [6]

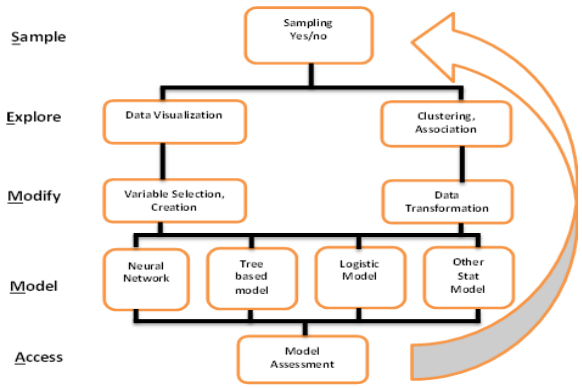

Fig. 2. SEMMA Model (Delen and Olson (2008))

- **CRISP-DM:** Provides a structured approach to planning a data mining project. It is a robust and well-proven methodology. This model is an idealized sequence of events. In practice many of the tasks can be performed in a different order and it will often be necessary to backtrack to previous tasks and repeat certain actions. The model does not try to capture all possible routes through the data mining process. [6]

Fig. 3. CRISP-DM Model [6]

TABLE I
SUMMARY TABLES OF DATA MINING METHODOLOGIES

KDD	SEMMA	CRISP-DM
Pre KDD	———	Business Understanding
Selection	Sample	Data Understanding
Pre-processing	Explore	Data Understanding
Transformation	Modify	Data Preparation
Data Mining	Model	Modeling
Interpretation/ Evaluation	Assessment	Evaluation
Post KDD	———	Deployment

II. DATA MINING TASKS AND TECHNIQUES:

Various tasks and techniques such as: Classification, Clustering, Regression, Neural Networks, Association Rules, Decision Trees, Genetic Algorithm, Nearest Neighbor method etc., are used for knowledge discovery from databases [4]. Now a days the ANN especially deep learning plays a crucial role in handling tasks of all the machine learning algorithms.

A. CLASSIFICATION:

Classification is the most commonly applied data mining technique, which employs a set of pre-classified examples to develop a model that can classify the population of records at large. This approach frequently employs decision tree or neural network-based classification algorithms. The data classification process involves learning and classification. In classification test data are used to estimate the accuracy of the classification rules. The classifier-training algorithm uses these pre-classified examples to determine the set of parameters required for proper discrimination. The algorithm then encodes these parameters into a model called a classifier.

Types of classification models:

- Classification by decision tree induction
- Bayesian Classification
- Neural Networks
- Support Vector Machines (SVM)
- Classification Based on Associations

B. CLUSTERING:

Clustering can be said as identification of similar classes of objects. By using clustering techniques we can further identify dense and sparse regions in object space and can discover overall distribution pattern and correlations among data attributes. Classification approach can also be used for effective means of distinguishing groups or classes of object but it becomes costly so clustering can be used as preprocessing approach for attribute subset selection and classification.

Types of clustering methods

- K-Means Methods
- Partitioning Methods
- Hierarchical Agglomerative (divisive) methods
- Density based methods
- Grid-based methods
- Model-based methods

C. PREDICATION:

Regression technique can be adapted for predication. Regression analysis can be used to model the relationship between one or more independent variables and dependent variables. In data mining independent variables are attributes already known and response variables are what we want to predict. Unfortunately, many real-world problems are not simply prediction. Therefore, more complex techniques (e.g., logistic regression, decision trees, or neural nets) may be necessary to forecast future values. The same model types can often be used for both regression and classification. For example, the CART (Classification and Regression Trees) decision tree algorithm can be used to build both classification trees (to classify categorical response variables) and regression trees (to forecast continuous response variables). Neural networks too can create both classification and regression models.

Types of regression methods

- Linear Regression
- Multivariate Linear Regression
- Nonlinear Regression
- Multivariate Nonlinear Regression

D. ASSOCIATION RULE:

Association and correlation is usually to find frequent item set findings among large data sets. This type of finding helps businesses to make certain decisions, such as catalogue design, cross marketing and customer shopping behavior analysis. Association Rule algorithms need to be able to generate rules with confidence values less than one. However the number of possible Association Rules for a given data set is generally very large and a high proportion of the rules are usually of little (if any) value.

Types of association rule

- Multilevel association rule
- Multidimensional association rule
- Quantitative association rule

E. NEURAL NETWORKS:

Neural network is a set of connected input/output units and each connection has a weight present with it. During the learning phase, network learns by adjusting weights so as to be able to predict the correct class labels of the input tuples. Neural networks have the remarkable ability to derive meaning from complicated or imprecise data and can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. These are well suited for continuous valued inputs and outputs. For example handwritten character reorganization, for training a computer to pronounce English text and many real world business problems and have already been successfully applied in many industries. Neural networks are best at identifying patterns or trends in data and well suited for prediction or forecasting needs.

- Back Propagation

III. RELATED WORKS:

It allows knowing about the concepts in the research in detail. A vast study is done under this literature study to know about the possibilities of happening. A deep understanding and determination can be achieved by means of the literature study. For this study, many papers are studied and a brief report is given on the work done previously by other authors in the same background.

This section presented a number of researches that demonstrate the application of data mining in HIV/AIDS control and prevention program in particular and health care domain in general are discussed.

Saravanan et al. [7] In this paper the Dempster Shafer (DS) theory of evidence is used to extract the hidden information by filtering the features from the large data set. Four variables are focused and one domain expert is used to find the belief and plausibility values which gives to select the variables for data mining to understand the behavior of HIV. Use of more than one expert and more than four variables for mass assignment of the variables can be done in the further research work.

Cerna P.and Abdulahi T. [8] Investigate the data mining technology to predict the Anti-Retroviral drugs usages for pharmacy based up on patient records of Harar's Jegul hospital in Ethiopia. The study uses KDD as a method which is classification of decision tree algorithms with M5P model tree. The study also recommends further work by using different methods of data mining techniques for a basis of comparison. Such will enhance the reliability and validity of

the predicted result.

Behailu G. and Tesfahun H. [9] present the application of data mining techniques to predict the CD4 status of ART on HIV patient. In this study CRISP-DM methodology used as a method and a classification algorithm used to predict the status is J48 and PART. The accuracy level of the J48 is 88.79%. The data set used in the research was Jimma and Bonga Hospital ART data base from 2003 to 2012.

Aweke G. [10] develops predictive models to identify the determinant risk factors of HIV infection using data mining techniques. The authors followed hybrid methodology for prediction of HIV infection determinants and factors. In this study, the researchers used classification techniques with Decision Tree J48, SMO and PART algorithms for experiment purpose and building the models. Best classifications were found through PART with 96.7% accuracy rate. The studies were used VCT data sets from the year 2005 to 2011 G.C. for future work, the author recommends to came up applicable solutions in this specific area.

Hailu T.G. [11] Discuss the prediction rule of the different techniques to develop the HIV testing prediction model. The study uses CRISP-DM methodology to predict and explore Association rules among selected features or attributes. And the algorithms Decision tree, Naive Bayes, Neural network, logistic regression has been used to build the model. The data set from EDHS 2011. The study indicates decision tree performed the best with accuracy of 96%, the decision tree induction method J48 came out to be the second best with a classification accuracy of 79%, followed by neural network 78%. Logistic regression has also achieved the least classification accuracy of 74%. This study was based on data from the 2011 EDHS.

Rosma et al. [12] described the feasibility of applying data mining technique to predict the survival of AIDS. An adaptive fuzzy regression classification technique, FuReA, was used to predict the length of survival of AIDS patients based on their CD4, CD8 and viral load counts. The authors discovered that neural network model was able to predict the survival of AIDS with an accuracy of 60% to 100%. In this study Data from 997 patients with HIV/AIDS at the University Hospital, Kuala Lumpur, Malaysia from 1987 to 2007 were used for predicting survivals among HIV/AIDS patients.

Zewdu T and Beshah T. [13] Develop a predictive model for HIV status; the study uses CRISP-DM methods to implement the model from VCT data sets of Addis Ababa town, Ethiopia. The data set was prepared as in REC formats. The authors used different tools in order to prepare and implements the data sets like Microsoft Excel and WEKA. In this research J48, Naïve bayes and PART were used as algorithms and the prediction accuracy is 93.95%.

Leke B. et al. [14] The research work focuses on the classification of HIV/AIDS status of individuals based on a given demographic factors. Radial basis function, Bayesian and Neural Network Architecture were used as a methodology and these gives accuracy rate of 84.24%. This study uses demographic and medical data sets of 2001 survey of South African Antenatal Seroprevalence. Supervised learning method used for train neural networks in order to classify the status of HIV between individuals.

Gosain A. et al. [15] The paper present decision tree and association rule techniques to predict the occurrences of transmission HIV based on treatment history. The authors focused on the demonstration of data mining methods for new k knowledge and pattern related and demographic, socioeconomic and medical histories of patients. Classification based techniques like decision tree and association rule is applied for huge volume of health care data sets. The data used in the study is ART Clinic, All India Institute of Medical Sciences (AIIMS), New Delhi. Data set from HIV Database (HIVDB) was derived.

Markos Z. et al. [16] The study applies KDD methodology to build predictive model from 2011 Ethiopian Demographic and Health Survey (EDHS) dataset. The algorithms were used is J48, PART and Naïve bayes. From those rules PART pruned rule induction found best performance 92.6% and accuracy rate is 97.8%. Finally the study recommend for future research, it must be add large DHS datasets and employ other classification algorithms and tools to get better accuracy result.

Singh Y. et al. [17] The author's works on machine learning application to predict HIV patients CD4 cell count using genome sequences. In this research work a regression model is used to count the CD4, and the data were used is patient data sets of containing protease or reverse transcriptase genome sequence and CD4 count from Stanford HIV drug resistance databases. Result shows that predicting using machine learning is successful. Algorithms used are neural networks.

Annang D. [18] This study employed the use of supervised learning as a classification technique to classify the HIV status of a female sex worker. The study showed that the predictive models: random tree, neural networks, J48 algorithms, logistic regression and naïve bayes were able to predict whether a female sex worker will test negative or positive for HIV given certain socio-demographic and behavioral factors as inputs with an accuracy of 98.9%, 97.41%, 93.18%, 91.12% and 89.97% respectively. The study follows CRISP–DM methodology. And it used Ghana's 2015 FSW IBBSS data set. The study uses as a tool for data preprocessing and preparation is Microsoft Excel and WEKA 3.6.9. And implement the experimentations using 5 algorithms: Random Tree, J48, Naïve Bayes, Logistic

regression and Neural Network. The research recommended for Future work feature selection algorithms must be applied for comparison the selected attributes.

Srinivas K. et al. [19] The study examine data mining techniques of classification based Naïve Bayes, decision tree and ANN. Dependency Augmented Naïve Bayes classifier and Naive creedal classifier 2 are used for this research to prepare the data. The data set used in this study is massive number of ARFF data sets, including the data sets from the UCI repository.

Anbarasi M. et al. [20] The study applies genetic algorithms to determine attributes from the diagnosis of heart disease to enhance the prediction heart disease. Three algorithms such as classification by clustering, decision tree and naïve bayes are used for prediction. Data set of 909 records with 13 features is used. This study needs further applying fuzzy learning models to evaluate the intensity of cardiac disease.

Osareh A. et al. [21] The paper uses neural network, k-nearest and SVM together with sequence forward selection, signal to noise ration feature ranking and component feature extraction to diagnose breast cancer. The general accuracy rate of these machine learning achieved 98.80% and 96.33% using SVM classifiers. The data sets utilized in the study is fine needle aspirate breast lesions with negative samples and positive samples (235 and 457 respectively).

Bellaachia A. and Guven E. [22] Breast cancer survivability prediction is made in this study using data mining techniques. The author investigate three algorithms the C4.5 decision tree, Naïve Bayes and the back-propagated neural network and conduct several experiments. The study achieved good performance comparable with others prediction works, which is done before by other researchers. And much better performance is gained by C4.5 decision tree algorithms rather than the two. In this experiment SEER public data are used.

Ilayaraja M. and Meyyappan T. [23] Propose association rule based on Apriori techniques to get the frequency of diseases. The study is done by on-patients with distinct geographical location and with different time periods. The authors uses the existing EMR records for train data sets, in this record they select 29 diseases on one year (2012) from total 1216 patient records, WEKA is employed. The analysis shown the information those four different diseases affected the patients frequently at various geographical locations during the year 2012.

Ji Y. et al. [24] The study proposes innovative techniques to mine associations of infrequent diseases from electronic patient data sets. The researcher develops a novel measure of exclusive casual leverage and fuzzy recognition-primed decision model. Based on these two models they try to mine the casual relationships between drug and their associate reactions. The experiment data were from Veterans Affairs Medical Center. The retrieved data included 16,206 patients from this data 96.3% were male, 3.7% were female.

Dagdia Z. [25] Presented a case study for using a Rough Set theory approach as a data mining technique in the context of epidemiology and Cancer Incidence Prediction. The research uses a previously introduced distributed method called RST and considers a data set provided by the Open Cancer organization. After some data preprocessing the research perform feature (risk factor) selection with RST and analyze the results from two different angles: insights epidemiologist can gain from the selected risk factors and the quality of the regression model.

Desikan, P. et al. [26] In this paper the author's develop a new data mining techniques and tools to compute medical data analysis and make solutions. They propose a hybrid tool integrates ANN and RST to make capable data analysis and predictions. The data set used is spermatological animal semen for prediction. The accuracy rate of the prediction is observed and sharping.

Undrajavarapu S. [27] A Review on Data Mining Process in Health care Department to Identify the Frequently Occurring Diseases. The study recommends association rule based Apriori for generating frequency of diseases. The study is made based on different geographical area with various time periods. The data set is EMR from hospitals. The output of the research is four different diseases are affected frequently with different geographical area in a particular year.

Chaurasia V. and Saurabh P. [28] The researchers' works on innovative methods for detection of breast cancer using classification data mining techniques. The classification techniques and WEKA were used as data mining tools. The research examine the accuracy of various classification data mining techniques. A total of six hundred and eighty three data of ten datasets features were used to determine the accuracy of the data mining techniques used. Three data mining techniques were compared using WEKA and the result shows the Sequential Minimal optimization has better accuracy than other techniques. The data breast cancer data set has been used to test, for classification. The data set is from Wisconsin data set from UCI machine learning. In this experiment, the researcher compare three classification techniques with WEKA and comparison of SVM result is has better accuracy i.e. 96.2% than IBK and BF Tree methods.

Kunwar V. et al. [29] Focus on using data mining classification techniques to analyzed chronic diseases. The work predicts chronic kidney diseases using data mining classification techniques such as ANN and Naives Bayes. Rapid miner tool was used to compare the two classification techniques and the results shows that Naives Bayes performs better than ANN. The data set used in this study is taken

from UCI Machine Learning Repository. The achieved results showed that Naïve Bayes, ANN, 100% and 72.73% accuracy respectively. Further research can be accompanied using other classifiers like KNN and Fuzzy logic.

Shakil K. et al. [30] Research on the prediction of Dengue disease using WEKA as data mining tool. Some data sets features were used based on the symptoms of the diseases. They compared different data mining techniques with their own classifier algorithm. The result shows that Naïve Bayes have an accuracy of 100% and J48 has accuracy of 99.70%. Naives Bayes give a better prediction of Dengue diseases survival. The study used dengue data set for prediction and classification.

Masethe H. D. and Masethe M. A. [31] The researcher's uses patient's data sets to predict heart diseases using classification data mining techniques. It determines the best framework that can give better accuracy in terms of performance for the diagnoses of the data set feature input. Classification data mining techniques was used with WEKA to predict and interpret the result.

The study uses different algorithms like Naive Bayes, J48, CART, REPTREE and Bayes Net for predict heart attacking disease. and it's accuracy rate of prediction is 99%. Medical Practitioners in South Africa data set were used to predict heart attack.

Chaudhary A. and Garg P. [32] The researcher uses A-priori and k-means algorithm to predict heart diseases and kidney failure. Predictive system is developed to evaluate data from the user which consists of 42 data set features. Data are analyzed using data mining tools and the results were evaluated with operating characteristics (ROC) and calibration plots. Using these algorithms the model gives simple and efficient way for finding the stage of the diseases. For future work the author's recommend this to implement for more diseases effective to using Artificial Intelligent.

IV. CONCLUSION:

This paper reviewed previous works on data mining in HIV/AIDS status and health care system. The paper were discussed the background, definition, methodologies of data mining, data mining techniques and algorithms used in health care and tries to compare different models. This review shows data mining tools used in health care to predict future outcome from information generated that assist health organizations to make decision. Related works of previous research were reviewed with their techniques and gaps and specific application areas of data mining in health care were mentioned. The review article will be benefiting the health care academicians, practitioners, researchers who are engaged in the areas of health care Management.

ACKNOWLEDGMENT

The authors are grateful to the editor and anonymous reviewers for those constructive comments and recommendations.

REFERENCES

[1] Alkema, Leontine, et al. "Global, regional, and national levels and trends in maternal mortality between 1990 and 2015, with scenario-based projections to 2030: a systematic analysis by the UN Maternal Mortality Estimation Inter-Agency Group." The Lancet 387.10017 (2016): 462-474.

[2] Walker, Neff, et al. "Estimating the global burden of HIV/AIDS: what do we really know about the HIV pandemic?." The Lancet 363.9427 (2004): 2180-2185.

[3] Raghavan V. et al. "Introduction. Journal of the American Society for Information Science" (1998): 49(5), 397-402.

[4] Han, Jiawei, and Micheline Kamber. "Data Mining Concepts and Techniques, published by Morgan Kauffman." (2006).

[5] Fayyad, Usama, Gregory Piatetsky-Shapiro, and Padhraic Smyth. "From data mining to knowledge discovery in databases." AI magazine 17.3 (1996): 37.

[6] Azevedo, Ana Isabel Rojão Lourenço, and Manuel Filipe Santos. "KDD, SEMMA and CRISP-DM: a parallel overview." IADS-DM (2008).

[7] Saravanan, A. M., R. Vijaya, and C. Jothi Venkateswaran. "Feature Selection for Prediction of HIV/AIDS using Data Mining Technique by applying the concept of Theory of Evidence." IJCSNS 11.5 (2011): 285.

[8] Cerna, Patrick D., and Thomas Jemal Abdulahi. "Prediction of Anti-Retroviral Drug Consumption for HIV Patient in Hospital Pharmacy using Data Mining Technique." International Journal of Information Technology and Computer Science (IJITCS) 8.2 (2016): 52-59.

[9] Haile Mariam, T. "Application of Data Mining Techniques for Predicting CD4 Status of Patients on ART in Jimma and Bonga Hospitals, Ethiopia." Journal of Health and Medical Informatics 6.06 (2015).

[10] Girma Awoke. "Predicting hiv infection risk factors using voluntary counseling and testing (vct) data." Ethiopian Journal of Health Development, 2012.

[11] Hailu, Tesfay Gidey. "Comparing data mining techniques in HIV testing prediction." Intelligent Information Management 7.03 (2015): 153.

[12] Dom, Rosma Mohd, et al. "The prediction of AIDS survival: A data mining approach." Proceedings of the 2nd WSEAS international conference on Multivariate analysis and its application in science and engineering. World Scientific and Engineering Academy and Society (WSEAS), 2009.

[13] Zewdu, Tewodros, and Tibebe Beshah. "Prediction of HIV Status in Addis Ababa using Data Mining Technology." HiLCoE Journal of Computer Science and Technology 2.2 (1998): 65-71.

[14] Leke-Betechuoh, Brain, et al. "Prediction of HIV status from demographic data using neural networks." 2006 IEEE International Conference on Systems, Man and Cybernetics. Vol. 3. IEEE, 2006.

[15] Gosain, Anjana, and Amit Kumar. "Analysis of health care data using different data mining techniques." 2009 International Conference on Intelligent Agent and Multi-Agent Systems. IEEE, 2009.

[16] Markos, Z., et al. "Predicting Under nutrition status of under-five children using data mining techniques: The Case of 2011 Ethiopian Demographic and Health Survey." J Health Med Inform 5 (2014): 152.

[17] Singh, Yashik, Nitesh Narsai, and Maurice Mars. "Applying machine learning to predict patient-specific current CD 4 cell count in order to determine the progression of human immunodeficiency virus (HIV) infection." African Journal of Biotechnology 12.23 (2013).

[18] Annang, Dennis Adjei. "Performance Comparison of Data Mining Techniques for Predicting Hiv Status Among Female Sex Workers in Ghana." Diss. University Of Ghana, 2018.

[19] Srinivas, K., B. Kavihta Rani, and A. Govrdhan. "Applications of data mining techniques in healthcare and prediction of heart attacks." International Journal on Computer Science and Engineering (IJCSE) 2.02 (2010): 250-255.

[20] Anbarasi, M., E. Anupriya, and N. C. S. N. Iyengar. "Enhanced prediction of heart disease with feature subset selection using genetic algorithm." International Journal of Engineering Science and Technology 2.10 (2010): 5370-5376.

[21] Osareh, Alireza, and Bita Shadgar. "Machine learning techniques to diagnose breast cancer." 2010 5th International Symposium on Health Informatics and Bioinformatics. IEEE, 2010.
[22] Abdelghani, Bellaachia, and Erhan Guven. "Predicting breast cancer survivability using data mining techniques." SIAM INTERNATIONAL CONFERENCE ON DATA MINING, 2006.
[23] Ilayaraja, M., and T. Meyyappan. "Mining medical data to identify frequent diseases using Apriori algorithm." 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering. IEEE, 2013.
[24] Ji, Yanqing, et al. "Mining Infrequent Causal Associations in Electronic Health Databases." 2011 IEEE 11th International Conference on Data Mining Workshops. IEEE, 2011.
[25] Dagdia, Zaineb Chelly. "A scalable and distributed dendritic cell algorithm for big data classification." Swarm and Evolutionary Computation (2018).
[26] Desikan, Prasanna, Kuo-Wei Hsu, and Jaideep Srivastava. "Data mining for healthcare management." (2011).
[27] Subhashini Undrajavarapu. "A review on data mining process in healthcare department to identify the frequently occurring diseases." IJRCCT, 4(4):315-318, 2015.
[28] Chaurasia, Vikas, and Saurabh Pal. "A novel approach for breast cancer detection using data mining techniques." (2017).
[29] Veenita Kunwar, Khushboo Chandel, A Sai Sabitha, and Abhay Bansal. "Chronic kidney disease analysis using data mining classi
cation techniques." In Cloud System and Big Data Engineering (Conuence), 2016 6th International Conference, pages 300-305. IEEE, 2016.
[30] Kashish Ara Shakil, Shadma Anis, and Mansaf Alam. "Dengue disease prediction using weka data mining tool." arXiv preprint arXiv:1502.05167, 2015.
[31] Hlaudi Daniel Masethe and Mosima Anna Masethe. "Prediction of heart disease using classification algorithms." In Proceedings of the world Congress on Engineering and computer Science, volume 2, pages 22-24, 2014.
[32] Anu Chaudhary and Puneet Garg. "Detecting and diagnosing a disease by patient monitoring system." International Journal of Mechanical Engineering And Information Technology, 2(6):493-499, 2014.

AUTHOR'S

Daniel Mesafint is research scholar in Computer Science, Mangalore University, India, under the supervision of Dr. Manjaiah D.H. He held the Bachelor of Education (B.Ed.) degree in Business Education Emphasis in Information Technology, Mekele University, Ethiopia and Masters of Science (M.Sc.) degree in Information Technology, Adama Science and Technology University (ASTU), Adama, Ethiopia. His primary research interests are in the areas of Data Mining, Data Warehousing, Machine Learning.

Manjaiah D.H. is Professor and Chairman in Post Graduate Studies and Research in Computer Science, Mangalore University, India. His Qualification is B.E.,M.Tech., Ph.D., in Computer Science. His primary interests of researches are in the areas of Quality of Services in reliable Data Communication System on AD-HOC Networks and Wireless Sensor Networks, Design issues of IPv4 to IPv6 mobility, authentication, tunneling, transition and handoff specifications for heterogeneous networks, Significant Analysis of Novel routing algorithms for AD-HOC networks and energy efficient algorithms for wireless sensor networks and Design specifications of Localization, Virtualization and Data Centre using Cloud Computing / Grid Computing.