Computer Engineering and Intelligent Systems
ISSN 2222-1719 (Paper) ISSN 2222-2863 (Online)
Vol.13, No.3, 2022

Classification Model for Hepatitis B Disease Using Supervised Machine Learning Technique

ADDISALEM OBSI ASSEGIE
College of Engineering and Technology, ASSOSA-GBCTE University, Ethiopia
E-mail obsiaddisalem@gmail.com

Abstract
Hepatitis B is the most common serious liver infection in the world and caused by the Hepatitis disease. This results in many people injure and deaths, many human life lost due to this disease. The Most countries around the world, including Ethiopia, have increased the number of patients. This has led to an increase in the number of life lose. However, it is frequently challenging to determine which specific environments lead to such factor. Various studies have been conducted to classify hepatitis B disease, and others are focusing on whether the peoples will live or die because of this disease. Furthermore, most of the studies conducted so far focused on hepatitis B disease prediction with fewer number of features. The study aims to classify the factors relevant to hepatitis B disease such as chronic and acute hepatitis B disease factors based on the independent variables collected from Arba Minch. The data for this study was collected from Arba Minch General Hospital. It covers ten years hepatitis B patient data record from the year 2002-2012 E.C. the preprocessed dataset has 14 attribute and 50032 instance. This study has been conducted using an experimental approach to determine the best- performing model. This study used the WEKA tool and Asp.Net programming language for implementation and analysis purposes. For this study, the researchers trained four different models, including J48, REP Tree, Bayes Net, and PART algorithms. Those models are selected based on a comprehensive study showed to select the best First style performing model. In this study, evaluation of the model was done using percentage split (80/20), and classification performance metrics was used in order to compare the models. The finding of this study displays that the J48 classifier outclasses then the rest of the classifiers with an accuracy of 85.5% on training data and 82.7% on test data. Based on this result, a system prototype was developed and tested that is accomplished of classifying features of hepatitis B disease.
Keywords: Machine Learning, Classification Algorithm, J48, Hepatitis B Diseases
DOI: 10.7176/CEIS/13-3-01
Publication date: May 31st 2022

1. Introduction

The word hepatitis comes from the Ancient Greek word hepar (root word hepat) meaning liver and in Latin meaning inflammation. Hepatitis damages to the liver with infection of the liver cells. The liver is the major gland in the human body. It evaluates almost 3 lb (1.36 kg). It is rose-colored chocolate in shade and is divided into four portions of diverse dimensions and spans. Lifespan diagnosis of hepatitis is a challenging task in early stage due to various interdependent features.

Hepatitis B is supposed to be the main etiological agent for chronic liver disease worldwide. Over 2 billion people today have been infected with HBD and 350 million of them are chronically and acute infected, with annual death of more than 1million of HB-related CLD. Mother-to child transmission is responsible for approximately one-half of chronic hepatitis B infection worldwide [6].

Hepatitis is the frustration of liver, most frequently caused by Disease infections. Five hepatotoxic Disease A, B, C, D, and E are infamous to cause hepatitis. Of these, hepatitis B are of greater status and among the most frequent Disease infections in humans [7]. Hepatitis B are the leading causes of liver-related disease and transience throughout the world. The healthcare industry collects an enormous amount of data which is not correctly extracted and not set to the best use. Detection of these hidden patterns and connections frequently goes vacant [8] [9].

The amount of HB Disease infections in Ethiopia takes not continued affluent considered at public nearby [10]. In 2016, the World Health Assembly published the Global Health Sector Strategy, calling for thee lamination of the threat of hepatitis in humans by 2030, reducing the number of new infections by 95% and the number of hepatitis deaths by 65% [11]. It is under the diaphragm on the realistic in the thoracic area of the stomach. Lifeblood touches the liver through the hepatic route and the gateway manner. The gate way manner transmits lifeblood covering processed diet from the small intestine, whereas the hepatic route transmits oxygen-rich lifeblood from the vein.

Hepatitis, is virus-related or inflammation of disease infection, is mainly In Ethiopia, there is more than 60% of chronic hepatitis [1]. According to WHO, Ethiopians among hepatitis endemic countries in the world its methods of transmission include consuming contaminated water, living in unclean environments and eating poorly cooked meat. Considering that transmission is greatly contingent upon hygiene and safety, clean practices such as washing hands regularly and chemical purification of water should be encouraged. Therefore, the researcher

need relevant and high quality data from huge records using machine learning techniques a person may have a genetic problem, a metabolic disorder, or an immune related injury. Obesity can be a cause of liver damage which can lead to enlargement. These are known as non-infectious, because they cannot spread from person-to-person. Huge amounts of data can be obtained and stored in both technical and business connections [5].

According to [2]. It is a very challenging task to the researchers to classify the diseases from the huge medical databases. The major challenge that the healthcare industry faces now-a-days is dominance of facility as the researcher recommended in health center there is shortage of materials. Diagnosing the disease correctly & providing the effective treatment to patients will define the quality of health service. Poor diagnosis causes tragic significances that are not accepted. In medical, hepatitis disease means injury to the liver with inflammation of the liver cells [3]. This study should make contributions in the life classification and prediction of hepatitis disease for medical research and introduce a detailed and broad evaluation of standard machine learning methods.

The term machine learning refers to the automated detection of meaningful patterns in the collected data. In the past few decades, it's become a standard tool in almost any task that needs information extraction from large data sets. It also widely used in scientific applications such as bio-informatics, medicine, and astronomy. One common feature of all of those applications is that, in contrast to more traditional uses of computers, in these cases, recognitions to the complexity of the patterns that need to be detected, a person's programmer cannot provide a particular, fine detailed specification of how such tasks should be executed. As the review in different research papers in recent years in healthcare sectors, machine learning became a fortune of use for disease classification and prediction. The term machine learning can be defined in many different ways by different scholars. Machine learning – is Algorithms that generate Algorithms are a sequence of instructions used to solve a problem. Depending on the scholars that define by Arthur Lee Samuel, a pioneer in the AI field, in 1959 about machine learning means Machine learning is the process of dig up information from the massive datasets Machine learning is a branch of artificial intelligence (AI) focused on building applications that learn from data and improve their accuracy over time without being programmed to do. so Machine learning tools are concerned with generous programs with the ability to "learn" and adapt new things from the specified data sets in the determined application domain [12], [13].Machine learning could be a branch of AI (AI) that aims at enabling machines to perform their businesses skillfully by using intelligent code. The statistical learning ways represent the backbone of intelligent code that's wished to develop machine intelligence.

As a result of machine learning algorithms need the knowledge to search out, the discipline should have reference to the discipline of the database. One wonders the way to view the large picture during which such a connection is clarified [14]. Poor diagnosing causes disastrous consequences that don't appear to be accepted. Records or knowledge of medical history is extremely large; however, these are from several dissimilar foundations. The interpretations that are done by physicians are essential elements of that knowledge. The information within the real-world may be noisy, incomplete, and inconsistent, therefore knowledge preprocessing is needed in directive to fill the omitted values within the database [15].

The use of machine learning as a tool to aid medical diagnostics is becoming increasingly prevalent. The different analysis models help researchers, data scientists, engineers, and analysts to produce proper results and decisions, has numerous applications such as building life classification models which can be extremely beneficial in the healthcare industry [16].

Hepatitis is a liver disease caused by the hepatitis virus and some cancers such as liver cancer and lymphomas in humans. Currently, machine learning technology is an important branch of artificial intelligence and widely used for analyzing medical data. Machine learning can automatically discover and execution the interactions and nonlinear relationships between variables and improve the accuracy of disease in classification model.

In order to help in the diagnosis of hepatitis, classification techniques such as Logistic Regression, Decision Tree, Linear Support Vector and Naive Bayes can be applied to classify real time hepatitis data based on an establishing training and testing set. But based on the performance and efficiency of the algorithm the researcher used J48, REP Tree, Bayes Net and PART rule based algorithms. The exact developments in which machine learning methodologies are applied are very varied. However, they all have one thing in common: rapidly escalating access to large quantities of data. Machine learning ML algorithms provide effective ways to build classification models using WEKA machine learning tool based on different scholars WEKA is more important than another for disease classification based on their recommendations the researcher prefer to use WEKA machines learning software.

II. Related Research Study

The healthcare environment is generally perceived as being 'information-rich' yet 'knowledge poor'. There is a huge amount of data available within the healthcare systems. However, there is a lack of effective analysis tools to discover hidden relationships and trend`s in data.

K. Santosh Bhargav et al [45] titles as "Application of Machine Learning Classification Algorithms on Hepatitis Dataset "at GITAM Institute of Technology, GITAM Visakhapatnam, India 2018. In this research paper

researchers are focusing on the classify weather the person will live or die using the accuracies and performance measures of the different machine learning algorithms. From the results, the conclusion obtained is as follows: the Logistic Regression algorithm gives with optimum accuracy of 83.17% which is closely followed by Decision Tree Algorithm with the optimal accuracy of 82.05%. Following the Decision Tree Algorithm is the Linear Support Vector Machine with an optimal accuracy of 76.92%, and lastly the Naïve Bayes Algorithm which has the optimal accuracy of 76.92%. Finally, these algorithms can help in classifying whether a person lives or dies.

V.Shankarsowmien et al. [46] Titled as "Diagnosis of Hepatitis using Decision tree algorithm" Many researchers investigated on liver disease, in which they have used algorithms such as C5.0, PCL, J48, and fuzzy rule. In this paper, C4.5 algorithm is used which is an efficient one than the existing algorithms. Since, the number of attributes are lesser, the complexity of the decision tree is reduced.

Avinash Golande, Pavan Kumar T [47] titled as "Heart Disease Prediction Using Effective Machine Learning Techniques" The training data is trained by using four different machine learning algorithms i.e. Decision Tree, KNN, K-mean clustering and Adaboost. This investigation tells us about dissimilar technologies that are used in dissimilar papers with dissimilar count of attributes with different accuracies depending on the tools designed for execution.

Joel Jacob et al. [48]Titled "A Diagnosis of Liver Disease Using Machine Learning Techniques "the researcher focused to predict liver disease using various machine learning techniques. It predicted using Support Vector Machine (SVM), Logistic Regression, K-Nearest Neighbor (K-NN) and Neural Network. All of them predicted with better results. With Each algorithm, it have observed Accuracy, Precision, Sensitivity and Specificity. Based on it the researcher conclude that the maximum accuracy for the test data was the artificial neural network.

Muktevi Srivenkatesh [49]titled as" Performance Evolution of Different Machine Learning Algorithms for Prediction of Liver "in 2019.in this research the researcher was made based on five calculations positioned K-Nearest Neighbour, Support Vector Machines, Logistic Regression, Navi Bayes, Random Forest With respect to, which speaks to the level of occasions ordered accurately, in this a variety calculated relapse having 76.27 % pursued by different calculations Concerning estimation of indicators, the estimations of Mean total error(MAE), Root Mean Square error(RMSE), Relative Absolute error(RAE), Root relative square error (RRSR) demonstrated that Logistic Regression indicators scored the most reduced qualities (MAE = 0.23) (RMSE = 0.48, RAE =60.02%, RRSE = 38.83) trailed by different calculations.

Ebru Aydindag Bayrak1 et al [50] titled as "Performance Analysis of Machine Learning Algorithms and Feature Selection Methods on Hepatitis Disease" in 2019. In this research it focuses on the most popular machine learning classification methods for Hepatitis. Naïve Bayes Classifier, Logistic Regression and J48 Decision Tree classification methods were used as machine learning classifier. Based on this A comparative analysis was performed on the basis of filter-based feature selection algorithms to classify hepatitis disease. A then it was apply feature selection method, Naïve Bayes Classifier shows the highest accuracy value % 84.51 compared with other classification algorithms.

Sara H. Muhi et al [51]titled as "Modeling for Predicting the Severity of Hepatitis Based on Artificial Neural Networks" in 2020. In this research work, three artificial neural networks (ANNs) methods, are used, Back Propagation neural network (BPNN), Radial Basis Function Network (RBFN) and, K-nearest neighbor (KNN), are used to forecast the level of hepatitis intensity.

Henok Yared Agizew [52]titled "as adaptive learning expert system for diagnosis and management of viral hepatitis" in 2019. In this research work the researcher used decision tree to simulate procedures involved in diagnosis and management of viral hepatitis, a rule-based knowledge representation methods and also SWI-prolog editor tool is used for codification of the represented knowledge.

Justin Monsi et al. [53] "Titled as XRAY AI: Lung Disease Prediction Using Machine Learning "in 2019.in this research work the researcher explored a system for detecting lung diseases from NIH X-ray dataset it filtered the data, prepared into input to the network and train the model several times making it learn and finally predict the disease by image detecting system and recommend to attaching attributes like rate of smoking and, family history to the data list entry.

Generally, the above-related work is conducted around the globe. They are focused on developing a prediction model for disease, and some researchers have developed a model that predicts whether the disease leads to live or die. The main aim of this research paper is to develop a classification model that classifies Hepatitis B diseases in to chronic and acute. The availability of this machine learning model in Hepatitis B disease's factors (classification) is essential to provide, advantage response time for both patient and health practitioner, because they get an alternative solution method with less time, and on-time services based on classified method of disease classification weather chronic and acute hepatitis B disease type. The machine learning models that were developed in other countries can not directly be used in our country. As the researcher mentioned above, most researchers develop prediction models only for single Hepatitis B disease(chronic or acute) and this is debut full without knowing one of the hepatitis Disease type most researchers do not describe the types of Hepatitis B

disease's.

The main difference between this study and the previously reviewed study is that this research paper classifies the factors that affect Hepatitis B also classifies in to chronic or acute based on the data that feed to the tool with its attribute, symptom's which is taken from the Arba Minch General hospital recorded data set and it uses different machine learning algorithms to classify the Hepatitis B disease's types and develop a model for classification purposes. It is more significant to reduce injure, disability and death of patients by identifying the major causes of Hepatitis B disease 'type and raising awareness to patients. Furthermore, no research on Hepatitis B disease's type classification using machine learning has been conducted in Ethiopia. As a result, in this study, a machine-learning model capable of classifying the factors contributing to Hepatitis B diseases was developed using a coexisting Hospital records dataset.

III. Methodology of the Study

Research Methodology is systematic approach for the purpose of investigation .The methodology is different from one problem to another problem .It must analysis properly to bring out the important feature and study for finding to solve method which are available .the search knowledge, investigation and discover knowledge .The purpose of research methodology is to find the truth and to understand the unknown. It should identify and select appropriate and relevant method to achieve the required outcome from the study .Research methodology select what tool should be used for the analysis and taking into account all the underlying assumption and all criteria under consideration.

The Proposed Architecture for Classification Model

The proposed model is used to classify the hepatitis B disease Type such as chronic and acute. The proposed solution is based on the model shown in Figure below. After collecting the data, models are developed using j48, REP Tree, Bayes net and part machine learning algorithms and trained on a training set of the whole dataset with k-fold and percentage split. The models are evaluated using accuracy, precision, recall, and ROC area. The evaluation results were used to select the best classification model. The outcome of this proposed study is a classification model used for hepatitis B disease type classification using Arba Minch general hospital recorded dataset. Finally, the classification model is evaluated and selected based on the results found using the model evaluation.

IV. Experimental Results

In this study, many experiments have been conducted using four supervised machine learning algorithms with the most commonly used percentage split option of 80/20, which means that 80% of the datasets were for training and 20% of the datasets were for testing. The experiments were performed using the J48, REP Tree, Bayes Net, and PART classifiers. All the results are presented in the subsequent sections. Different performance metrics such as TP rate, Precision, Recall, ROC area, and accuracy are considered for all selected classification algorisms which are implemented with the help of WEKA open source tool. the Selection criteria of this tool is based on: which is free open-source more interactive more visualizations in nature and also another research recommendation for health care dataset WEKA can more visualize based on Roc area.

Shows the Accuracy Result of Model Training

Algorithm implemented	Evaluation metrics					
	Accuracy	Average TP Rate	Average Recall	Average Precision	Average F-Measure	Average ROC Area
J48	85.58%	0.85%	0.86%	0.86%	0.85%	0.91%
REP Tree	85.38%	0.85%	0.85%	0.86%	0.85%	0.90%
Bayes Net	80.11%	0.80%	0.80%	0.80%	0.80%	0.87%
PART	79.22%	0.79%	0.79%	0.79%	0.79%	0.85%

As we observe from the above Table the training result of the four models is described and it shows that the accuracy of J48 (85.58), REP Tree (85.38), Bayes Net (80.11), and PART (79.22). From those models, the J48 classification algorithm performs better in all evaluation metrics than the other remaining algorithms such as REP Tree, Bayes Net, and PART classifiers with an overall accuracy performance 85.58 the mode for this research is developed by this training accuracy.

Shows the Accuracy Result of Model Testing

Algorithm implemented	Evaluation metrics					
	Accuracy	Average TP Rate	Average Recall	Average Precision	Average F-Measure	Average ROC Area
J48	82.7%	0.82	0.82	0.86	0.82	0.87
REP Tree	82.9%	0.83	0.83	0.84	0.83	0.86
Bayes Net	78.5%	0.78	0.78	0.78	0.78	0.85
PART	80.%	0.80	0.80	0.80	0.80	0.81

As we observe from the above Table the testing result of the four models is described and it shows that the accuracy of J48 (82.70), REP Tree (82.98), Bayes Net (78.58), and PART (80.02).

User Acceptance Testing

User acceptance testing is the process of ensuring that whether the developed prototype satisfies the requirements of its end-users of the system by solving the domain problems. This allows users of the system to test the proposed prototypical system by actually using it and evaluating the benefits received from it. For this study the researcher has prepared questionnaire and collected response from domain expert five (physician) and four IT Expert to evaluate the user acceptance of a system, and the evaluators fill out the questionnaire after they have used the hepatitis B disease prediction system prototype. After the refinement and verification of the system, user acceptance testing is conducted and for this system the end users are physician of hepatitis B disease. The objective of the test is to determine if the system meets its desired goals which is identifying the hepatitis B disease The evaluators assess the accuracy of the developed prototype system by using the following standards, these are: simplicity of use and interact with the prototype system, the attractiveness of the prototype system, efficiency in time, the accuracy of the prototype system in success a decision to identify the hepatitis B disease. For this study the researcher has used totally nine domain expert (physician) and IT expert respondents and they were conducted on the evaluation of the proposed system. All the evolution participant respondent (physician) has knowledge about the **hepatitis B** disease and has been working in the specified hospitals. The questionnaires used to test the performance of the prototype system by end-users after they seen it and the questionnaire establish.

The following table shows that the outcomes after evaluation by selected evaluators

Shows the prototype evaluators score based on the evaluation criteria

No	Criteria for evaluation	Poor	Fair	Good	V. Good	Excellent
1.	Functionality of the prototype system for hepatitis B disease prediction	0(0.00%)	0(0.00%)	2(13.33%)	3(20.00%)	10(66.67%)
2.	Reliability of the prototype system.	0(00.00%)	0(00.00%)	0(00.00%)	4 (26.67%)	11(73.33%)
3.	Usability of the prototype system.	0(00.00%)	0(00.00%)	2(13.33%)	4(26.67%)	9(60.00%)
4.	Efficiency of the prototype system in time.	0(00.00%)	1 (6.67%)	2 (13.33%)	4(26.67%)	8(53.33%)
5.	Maintainability of the prototype system.	0(00.00%)	0(00.00%)	0(00.00%)	3(20.00%)	12(80.00%)
6.	Portability of prototype system.	0(00.00%)	0(00.00%)	2 (13.33%)	5(33.33%)	8 (53.33%)

As shown in the above table 66.67% of the evaluators scored the functionality of prototype system criteria that accurately identify the hepatitis B disease after taking the history of the hepatitis B disease patient with the evaluation as excellent, 20% as very good, 13.33% as good. The second evaluation criteria reliability of the prototype system showed a greater rate of reliability by the evaluators the majority is scored 73.33% as excellent, and 26.67% as very good. The third evaluation criteria usability of the prototype system showed a greater rate of usability by the evaluators the majority is scored 60% as excellent, 26.67% as very good, and 13.33% as good. In the efficiency of the prototype system with respect to time criteria of evaluation, 53.33% of the evaluators scored as excellent, 26.67% as very good, 13.33% as good, and 6.67% fair.

The prototype system maintainability evaluation results show the, 80% as excellent, and 20% as very good. And the portability of the prototype system shows as 53.33% of the evaluators rated the prototype system as excellent, 33.33% as very good and 13.33% as good.

Generally, the performance evaluation result showed that designed hepatitis B disease classification prototype is acceptable by the domain and IT experts. For all the evaluation criteria all the assessment result scored more than 50% and the respondents were agreed to use the prototype for the classification of hepatitis B disease for the future program.

CONCLUSION AND RECOMMENDATION

In this research, a machine learning tool and algorithm model was trained to determine that patient has chronic or acute hepatitis B disease. This machine learning tool and algorithm model used a dataset with 14 selected attributes and 5,032 Datasets was selected to bearing experiments using different classifiers. After experimentation the classification model verified that it is better than outdated data. This study applied four machine learning techniques and algorithms on the dataset and out of these classifiers/algorithms the accuracy of J48 85.58% is get better than other algorithms. Based on this model it can be decided that the proposed model can classify the chronic and acute hepatitis B disease. This model can help doctors and medical workers to provide better treatments for patients.

In this study, data preprocessing was used to make the data clean and suitable for machine learning models using machine learning tools such as WEKA Missing values in this study were handled using the mean imputation method for nominal values and the mode method (replacing missing values with the most frequent value) for categorical values. The hepatitis B disease the target class dataset is replaced with C, A, and C for chronic and A for acute hepatitis B disease respectively

J48, Rep Tree, Bayes net and part algorithms are used and trained. The evaluation model was performed using a percentage split of 80/20, which means 80% of the data was used for training and 20% of the data was used for testing. To evaluate the performance of those models, many appropriate performance metrics such as TP rate, precision, recall, ROC area, f_mesure, and accuracy and confusion matrix were used. The study looked into choosing the best machine learning techniques for classifying hepatitis B disease, as well as developing the model with appropriate data and selecting the best-fitting model.

The result of the experiment accuracy on hold-out (test) data shows j48 (85.58%), Rep tree (82.7%), Bayes net (80.11) and part classifier (79.2%). Based on this, the J48 classifier is found to be the best classification model with an overall accuracy of 85.58% for the training set and 82.7% for the testing set, which is taken as an appropriate model to classify hepatitis B disease classification. Finally, the classification model prototype is then developed and tested using J48 classifiers, and it is found to be comprehensive and accomplished of being used in classification of hepatitis B disease type.

Recommendations

In this study, in order to develop a machine learning model, getting organized data was more difficult, and the data was converted from manual-based into electronic format, so this was a more time-consuming task. Thus, the Health office should try to register patient history data on the computer that enables the domain experts and other concerned bodies such us health practitioner to take and access the data simply when it is needed. In this research, determinations have been prepared to apply machine learning tool and algorithm to classify Hepatitis B disease.

Thus, based on the results of the research the following recommendations are forwarded as they are important issues for further research directions in classification modeling and critical of Hepatitis B disease. So that there are many machine learning techniques, this research used only the classification technique with three classification and selected four algorithm's based on their classification performances (Decision tree, Bayes and Rule based) classifier's which is combined with WEKA 3.9 tool.

This study the researcher has used WEKA 3.9 tool, but there are many other machine learning tools like Mat Lab and other programing which may be more sophisticated to develop model and analysis the data. It is better to develop the model using more instances of datasets from different sources of this study is capable, there capability be other ways to obtain more accurate and better performing results using other various models for Hepatitis B disease classification, especially for chronic and acute hepatitis B disease classification. Even if the accuracy to expand this classification model, testing further machine learning approaches with (increasing sample size if it is possible) such as unsupervised learning might be reflected, and comparing the results with this study is recommended even other such as machine learning method like function, and vector machine were not considered for testing and classification in this research.

References

[1] W. E. Adugna Endale Woldegiorgis, "Community-based sero-prevalence of hepatitis B and C infections in South Omo Zone, Southern Ethiopia," 2019.

[2] M. Dr. S. Vijayarani, "Liver Disease Prediction using SVM and Naïve Bayes Algorithms," *International Journal of Science, Engineering and Technology Research,* vol. 4, no. 4, April 2015.

[3] H. M. H. Fadl Mutaher Ba-Alwi, "Comparative Study for Analysis the Prognostic in Hepatitis Data: Data Mining Approach," *International Journal of Scientific & Engineering Research,,* vol. Volume 4, no. Issue 8, August 2013.

[4] P. T.Karthikeyan, "Analysis of Classification Algorithms Applied to Hepatitis Patients," *International Journal of Computer Applications,* vol. Volume 62, p. 0975 – 8887, January 2013.

[5] P. Yildirim, "Filter Based Feature Selection Methods for Prediction of Risks in Hepatitis Disease,"

International Journal of Machine Learning and Computing, vol. Vol. 5, August 2015.
[6] A. A. e. al., "Prevalence, Infectivity, and Associated Risk Factors of Hepatitis B Virus among Pregnant Women in Yirgalem Hospital, Ethiopia: Implication of Screening toControl Mother-to-Child Transmission," *Hindawi Journal of Pregnancy,* 5 August 2018.
[7] S. M. e. al., "Seroprevalence of hepatitis B surface antigen and anti HCV antibody and its associated risk factors among pregnant women attending maternity ward of FelegeHiwot Referral Hospital, northwest Ethiopia," *Virology Journal,* 2015.
[8] T. R. B. e. al., "Analysis of Data Mining Techniques For Healthcare Decision Support System Using Liver Disorder Dataset," in *International Conference on Computational Modeling and Security*, India, 2016.
[9] D. A. E. H. O. e. at., "Application of Data Mining Techniques to Explore Predictors of HCC in Egyptian Patients with HCV-related Chronic Liver Disease," *Asian Pacific Journal of Cancer Prevention,* vol. 16, 2015.
[10] A. E. W. e. al., "Community-based sero-prevalence of hepatitis B and C infections in South Omo Zone, Southern Ethiopia," *RESEARCH ARTICLE,* 30 December 2019.
[11] Z. D. e. a. Ying Wang, "Predicting Hepatitis B Virus Infection Based on Health Examination Data of Community Population," *International Journal of Environmental Research and Public Health,* 2019.
[12] M. A. e. al., "Artificial intelligence and machine learning in healthcare," *ehealthresearch.no,* 2018.
[13] K. A. e. al., "Predicting Infectious State of Hepatitis C Virus Affected Patient's Applying Machine Learning Methods," *2020 IEEE Region 10 Symposium,* June 2020.
[14] M. e. a. Mohammed, Machine Learning: Algorithms and Applications, Mishawaka, IN, U.S.A: CRC Press (2016).
[15] A. S. R. e. al., "A Comparative Study On Liver Disease Prediction Using Supervised Machine Learning Algorithms," *INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH,* vol. 8, no. 1, NOVEMBER 2019.
[16] T. D. K. e. a. K. Santosh Bhargav, "Application of Machine Learning Classification Algorithms on Hepatitis Dataset," *International Journal of Applied Engineering Research,* vol. 13, 2018.
[17] A. M. a. E. N. Sefinew Molla, "Seroprevalence of hepatitis B surface antigen and anti HCV antibody and its associated risk factors among pregnant women attending maternity ward of Felege Hiwot Referral Hospital, northwest Ethiopia:a cross-sectional study," *Virology Journal,* 2015.
[18] A. T. T. e. a. Aynishet Adane Gebremariam, "Seroprevalence of Hepatitis B Virus and Associated Factors among Health Professionals in University of Gondar Hospital, Northwest Ethiopia," *Advances in Preventive Medicine,* p. 5, 2019.
[19] e. a. Mohammed Abdullah Al-Hagery, "Knowledge Discovery in the Data Sets of Hepatitis Disease for Diagnosis and Prediction to Support and Serve Community," *International Journal of Computer and Electronics Research,* vol. 4, no. 6, 2015.
[20] T. A. J. D. Huda Yasin, "Hepatitis-C Classification using Data Mining Techniques," *International Journal of Computer Applications,* vol. Volume 24, p. (0975 – 8887), June 2011.
[21] Vibhuti Gupta et al, "A Systematic Review of Machine Learning Techniques in Hematopoietic Stem Cell Transplantation (HSCT)," *Sensors 2020, 20, 6100,* 27 October 2020.
[22] Adnan Qayyum et al, "Secure and Robust Machine Learning for Healthcare: A Survey," 21 Jan 2020.
[23] S. Raschka, Machine LearningLecture Notes, 2018.
[24] Internetsociety.org, "Artificial Intelligence and Machine Learning: Policy Paper," *Internetsociety.org,* April 2017.
[25] S. S. A. M.A.Jabbar1, "The Future of Health care: Machine Learning," *International Journal of Engineering & Technolog,* 29 November 2018.
[26] Diksha Sharma et al, "A Review on Machine Learning Algorithms Tasks and Applications," *International Journal of Advanced Research in Computer Engineering & Technology,* vol. 6, no. 10, October 2017.
[27] L. e. a. Amin Ul Haq, "A Hybrid Intelligent System Framework for the Prediction of Heart Disease Using Machine Learning Algorithms," *Mobile Information Systems,* p. 21, 2 December 2018.
[28] J. H. a. D. Kirsch, Machine Learning For Dummies®, IBM Limited Edition, 2018.
[29] L. Karimi, "Introduction to Machine Learning & Its Application in Healthcare," 2018.
[30] D. K. Judith Hurwitz, Machine Learning, C. A. B. Rev Mengle, Ed., United States of America: John Wiley & Sons, Inc., 2018, pp. 6-10.
[31] a. Diksha Sharma et, "A Review on Machine Learning Algorithms,Tasks and Applications," *International Journal of Advanced Research in Computer Engineering & Technology,* vol. 6 , no. 10, p. 2278 – 1323, October 2017.
[32] M. A. a. D. Sridharan, "Analysis of Clustering Algorithms in Machine Learning for Healthcare Data," Springer Nature Singapore, 2020.
[33] D. S. Anto, "Supervised Machine Learning Approaches for Medical Data Set Classification - A Review," *International Journal of Computer Science & Technology ,* vol. 2, no. 4, pp. 2229-4333, Oct. - Dec 2011.

[34] D. S. H. e. a. Sweety Bakyarani. E, "A Survey Of Machine Learning Algorithms In Health Care," *INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH,* vol. 8, no. 11, NOVEMBER 2019.
[35] D. L. O. e. al, Advanced Data Mining Techniques, USA, 2008.
[36] R. W. e. al., "a Standard Process Model for Data Mining," *DaimlerChrysler Research & Technology FT3/KL.*
[37] J. H. Rüdiger Wirth, "CRISP-DM: Towards a Standard Process Model for Data Mining," *DaimlerChrysler Research & Technology FT3/KL.*
[38] Maryam Daneshmandi et at, "A Hybrid Data Mining Model to Improve Customer Response Modeling in Direct Marketing," *Indian Journal of Computer Science and Engineering,* vol. 3, pp. 0976-5166, Dec-Jan 2012 -2013.
[39] Iqbal Muhammad et at, "SUPERVISED MACHINE LEARNING APPROACHES: A SURVEY," *ICTACT JOURNAL ON SOFT COMPUTING,* vol. 05, no. 03, APRIL 2015.
[40] G. D. M. a. A. Prentza2, Machine Learning in Medical Applications, Verlag Berlin Heidelberg, 2001, pp. 300-307.
[41] Brett Beaulieu-Jones et al, Trends and Focus of Machine Learning Applications for Health Research, October 25,2019.
[42] D. M. Bardia□Abbasi, Machine learning applications in epilepsy, 03 September 2019.
[43] B. P. N. A. L. L. Makhlysheva A, "Artificial intelligence and machine learning in healthcare," 2018.
[44] D. M. Chan, "GLOBAL HEPATITIS REPORT,2017," World Health Organization 2017.
[45] K. Santosh Bhargav et al, "Application of Machine Learning Classification Algorithms on Hepatitis Dataset," *International Journal of Applied Engineering Research,* vol. 13, pp. 12732-12737, Number 16 2018.
[46] V.Shankar sowmien et al, "Diagnosis of Hepatitis using Decision tree algorithm," vol. Vol 8, pp. 1414-1419, Jun-Jul 2016.
[47] P. K. T. Avinash Golande, "Heart Disease Prediction Using Effective Machine Learning Techniques," *International Journal of Recent Technology and Engineering,* vol. 8, no. 1S4, June 2019.
[48] Joel Jacob et al, "Diagnosis of Liver Disease Using Machine Learning Techniques," *International Research Journal of Engineering and Technology,* vol. 05, no. 04, |Apr 2018.
[49] M. Srivenkatesh, "Performance Evolution of Different Machine Learning Algorithms for Prediction of Liver Disease," *International Journal of Innovative Technology and Exploring Engineering,* vol. 9, no. 2, December 2019.
[50] Ebru Aydindag Bayrak et al, "Performance Analysis of Machine Learning Algorithms and Feature Selection Methods on Hepatitis Disease," *International Journal of Multidisciplinary Studies and Innovative Technologies,* vol. 3, pp. 135-138, 2019.
[51] Sara H. Muhi et al, "Modeling for Predicting the Severity of Hepatitis Based on Artificial Neural," *International Journal of Intelligent Engineering and Systems,,* vol. 13, 2020.
[52] H. Y. Agizew, "ADAPTIVE LEARNING EXPERT SYSTEM FOR DIAGNOSIS AND MANAGEMENT OF VIRAL HEPATITIS," *International Journal of Artificial Intelligence and Applications,* vol. 10, March 2019.
[53] Justin Monsi et al, "XRAY AI:Lung Disease Prediction Using Machine Learning," *international jurnal of information system and computer sciences,* vol. 8, March - Apri 2019.
[54] A. N. J. K. Sivaranjani, "Survey on Disease Diagnostic using Data Mining Techniques," *International Journal of Computer Sciences and Engineering,* vol. 6, no. -2, March 2018.
[55] M. J. C. P. e. al., "Survey on Asthma Prediction Using Classification Technique," *International Journal of Computer Science and Mobile Computing,* vol. 5, no. 7, pp. 515-518, , July 2016.
[56] S. A. e. al., "Application model of k-means clustering: insights into promotion strategy of vocational high school," *International Journal of Engineering & Technology,* pp. 182-187, 2018.
[57] D. K. D. Jai Ruby, "Predicting the Performance of Students in Higher Education Using Data Mining Classification Algorithms - A Case Study," *International Journal for Research in Applied Science & Engineering Technology,* vol. 2 November, no. XI, November 2014.
[58] P. P. e. al., "Preprocessing and Classification in WEKA Using Different Classifiers," *Journal of Engineering Research and Applications,* vol. 4, no. 8, pp. 91-93, August 2014.
[59] M. J. Swasti Singhal, "A Study on WEKA Tool for Data Preprocessing, Classification and Clustering," *International Journal of Innovative Technology and Exploring Engineering,* vol. 2, no. 6, May 2013.
[60] G. H. e. a. Eibe Frank, "The WEKA Data Mining Software: An Update," *SIGKDD Explorations,* vol. 11, no. 1.
[61] H. M. H. Fadl Mutaher Ba-Alwi, "Comparative Study for Analysis the Prognostic in Hepatitis Data:Data Mining Approach," *International Journal of Scientific & Engineering Research,* vol. Volume 4, no. Issue 8, August 2013.
[62] Machine Learning,Lecture Notes.
[63] w. h. o. 90, " hepatitis," 1 September 2019.