IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

CHRONIC LIVER DISEASE PREDICTION USING MACHINE-LEARNING ALGORITHMS

Princy.A,

Student, M.Sc. Computer Science, Nirmala College for Women, Coimbatore.

Uma Maheswari.R M.C.A., M.Phil.,

Assistant Professor, Nirmala College for Women, Coimbatore.

Abstract

Machine learning has a lot of potential for automated disease diagnosis. The liver is vital to life and promotes the body's ability to rid itself of poisons. Given the increasing increase in numerous chronic liver diseases, a preliminary diagnosis of chronic liver disease is required. With a population of 1.33 billion, India is the second most populous nation in the world, and millions of people are diagnosed with chronic liver disease each year. Jaundice, hepatitis (A, B, C), and non-alcoholic liver disease are a few examples of chronic liver diseases (NAFLD). They are brought on by drinking alcohol, eating tainted food, and other things including obesity. As a result, we would want a system that is trustworthy and capable of forecasting the signs of chronic liver disease. By utilising machine learning algorithms and patient data, this method forecasts Chronic Liver Diseases. Experiments and comparison analyses show that it improves classification accuracy, reduces classification time, and hence contributes in the more accurate prognosis of chronic liver disease. Accuracy, score, precision, recall, and f-measure are used to evaluate performance. To categorise liver patients, a variety of classification algorithms are used, and the best model is selected and applied based on the classification report and performance (Liver patient or not liver patient).

Keywords: Liver disease, Machine learning algorithm, NAFLD, disease diagnose.

I.INTRODUCTION

The liver, the body's largest organ, is crucial for food digestion and for expelling the body's toxins. Alcohol consumption, infections, and parasites damage the liver, creating potentially fatal diseases. There are many different types of chronic liver diseases[1], including cirrhosis, liver cancer, liver tumours, and hepatitis. Among these are cirrhosis, which is the leading cause of death, and chronic liver diseases. As a result, one of the greatest health issues in the globe is chronic liver disease. Almost 2 million people worldwide pass away each year as a result of chronic liver disease.

According to the worldwide Burden of Disease (GBD) study, which was published in BMC Medicine, million individuals have liver cancer, and million have lost their lives to cirrhosis in 2010. As a component of artificial intelligence (AI), machine learning (ML) enables a system to learn by analysing patterns in data using statistical models and algorithms. For training purposes and prediction accuracy, supervised algorithms use human inputs and outputs, and are therefore used in a variety of classification applications. Hence, the use of ML in healthcare has also increased. The growing number of people with chronic liver disease is one of the biggest issues in healthcare. The liver is an essential organ with tasks like bile generation, chemical detoxification, creation of crucial proteins for blood clotting, and other life-sustaining activities. Long-term drinking patterns have been directly related to an increased chance of developing several chronic liver diseases, which can be treated sooner to prevent death.

II.LITERATURE SURVEY

We present machine learning techniques for predicting chronic kidney disease using clinical data in this paper[2]. K-nearest neighbours (KNN), support vector machine (SVM), logistic regression (LR), and decision tree classifiers are four machine learning methods investigated. These predictive models are built from chronic kidney disease datasets, and their performance is compared to determine the best classifier for predicting chronic kidney disease. Based on several music features, this paper [3] predicts whether a song will be a super hit or a chart buster. It is critical to predict the popularity of a song, especially in a competitive business world. In our model, we will use Logistic Regression to predict the likelihood of reaching the Top 10 songs in popularity. This paper [4] investigated and analysed the classification algorithms Nave Bayes, Decision Tree, Multi-Layer Perceptron, and k-NN used in a previous study that developed our data set, as well as Random forest and Logistic that we proposed. These algorithms were evaluated using various criteria such as precision, recall, sensitivity, specificity, and so on. We learned from the experiments that, in terms of precision, Nave Bayes is superior to others; however, in terms of recall and sensitivity, Logistic and Random Forest outperformed other algorithms in the performance of prediction tests when considering the algorithmic characteristics to a liver patient data set.

Data mining provides methodological and technical solutions for dealing with medical data analysis and the development of prediction models. Patients with liver disease have been steadily increasing as a result of excessive alcohol consumption, inhalation of harmful gases, and consumption of contaminated food, pickles, and drugs. As a result, in this study, liver patient data is considered and evaluated using univariate analysis and a feature selection method for determining predicator attributes. A comparative study [5] of artificial neural network-based predictive models like BP, RBF, SOM, and SVM is also provided.

To determine the condition of the Liver Cancer, new metrologies for early Liver Cancer are required, according to the paper [6]. To address this issue, various data classification techniques or algorithms are employed. Decision tree, C4.5, Association rule, Bayesian networks, Support vector Machine, K-NN, Neural networks, and other classification techniques or algorithms are examples. Classification is a data mining technique or algorithm that categorises data based on attributes contained in datasets. To classify the data and build a model, classification techniques used the training and test data sets. This model is then used to categorise the new objects. So, datasets are taken from the Pt. B D Sharma Postgraduate Institute of Medical Sciences Rohtak to classify or build a model for medical diagnosis or health-care research. The Naive Bayes algorithm is probability-based, whereas the J48 algorithm is decision-tree-based. The purpose of this paper is to compare the classifiers NAIVE BAYES and J48 in the context of a bank dataset in order to maximise true positive rate and minimise false positive rate of defaulters rather than achieving only higher classification accuracy using the WEKA tool. The results of the experiments presented in this paper concern classification accuracy, sensitivity, and specificity. The results in the paper [7] on this dataset also show that j48 is more efficient and accurate than Nave bayes.

III METHODOLOGY

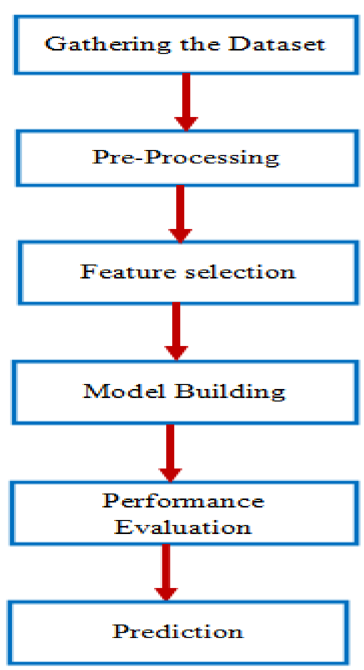

Fig. 1. Proposed Methodolgy

Data Gathering

The first step is to collect data, which can be obtained from hospitals or from the website. The dataset in this case is from the UCI [8] standard repository.

Data pre-processing:

Many records in the collected data may have missing data or values that take age into account. In general, missing values are replaced with the feature's nearest or closest value. And the chronic liver disease target data are divided into two categories: group 1 represents the presence of chronic liver disease patient records, and group 2 represents the absence of chronic liver disease patient records. In a classification model, the target label values are converted to non-numeric values. Following that, the dataset is divided into two groups: training and testing. The data are ready for classification of the models [9]. When the data pre-processing is finished, the stage of model selection for classification begins.

Feature selection:

The selection of important features of chronic liver disease is one of the main segments[10] in chronic liver disease prediction. Several features, such as age and gender, that represent each patient's personal information, are chosen in this step. Other clinical features are gathered from various medical tests.

Data classification:

In data mining, classification is an important process and function. The collected items' function is assigned to the target class or category. The classification aims[11] to get the target class to predict accurately for all case data. Following data pre-processing, features are added to a classification model. Random Forest, SVM, and Nave Bayes are some popular classification models.

Performance Evaluation:

For classification performance evaluation, various classification criteria such as accuracy, precision, sensitivity, specificity, f-measure, and FRP are computed.

Prediction:

In this step, the mapping of selected features onto the training model is performed in order to classify the given features and predict chronic liver disease. A specialist doctor labels the gathered chronic liver disease dataset to generate predictions. The classification is developed as a multi-class issue, and medical data is classified into various classes. As a result, each class refers to a specific type of chronic liver disorder [12]. Based on the selected important features, this process can determine the likelihood of a patient having liver disease.

The prediction is used to obtain accurate results for chronic liver diseases. The discovery of associations among independent and dependent variables is done with the help of prediction as suggested by its name. A historical chronic liver disease database is used to recognize and extract hidden knowledge of chronic liver disease. The complex questions are answered in order to diagnose chronic liver disease. As a result, practitioners can make informed clinical decisions. To summarize, the predicting analysis approach aids in forecasting future possibilities based on current data. The proposed model was enhanced by combining three classifiers: logistic regression, random forest, and the KNN algorithm. Python is used for the implementation of the proposed model, and the accuracy achieved is 77.58 percent. In the future, the clustering algorithm will be executed using the hybrid classifier technique for data division.

IV RESULT AND DISCUSSION

In this study, various classification algorithms, including Random Forest, Nave Bayes, Support Vector Machine, and KNearest Neighbour, were used to predict liver disease. All of these algorithms were compared based on classification accuracy, which was determined using a confusion matrix. As a result, it is possible to conclude that Logistic Regression is appropriate for predicting liver disease.

The system can be improved further by expanding the liver disease dataset to obtain more accurate liver prediction results.

REFERENCE

[1] Hsieh Jun-Wei, C.-Hung Lee, Y.-Chih Chen, W.-Shan Lee and H.-Fen Chiang, "Stage Classification in Chronic Kidney Disease by Ultrasound Image", *International Conference on Image and Vision Computing New Zealand ACM*, pp. 271-276, 2014.

[2] A. Charleonnan,T. Fufaung, T. Niyomwong, W Chokchueypattanakit,S. Suwannawach, N. Ninchawee "Predictive Analytics for Chronic Kidney Disease Using Machine Learning Techniques " MITiCON- 2016.

[3] Anatomy and function of the liver. [Online]. Available: https://www.medicinenet.com/liver_anatomy_and_function/article.ht m

[4] Abhishek Chowdhury, Thirunavukkarasu K, Sidhyant Tejas(2017), Predicting whether song will be hit using Logistic Regression. Volume 6 Issue 9 September 2017.

[5]K-Nearest Neighbours. [Online]Available: https://www.saedsayad.com/k_nearest_neighbors.htm

[6] P.Mazaheri, A. Narouziand A. Karimi (2015), Using Algorithms to Predict Liver Disease Classification,Electronics Information and Planning. 3:255-259.

[7] H. Jin , S. Kim and J.Kim (2014), Decision Factors on Effective Liver Patient Data Prediction,International Journal of Bio-Science and Bio-Technology

[8] Available: http://www.ics.uci.edu/~mlearn/databases/

[9] Comparative Study of Artificial Neural Network based Classification of Liver Patient, Journal of Information Engineeringand Application. 3(4).

[10] David Diez, Christopher Barr and Mine Cetinkaya-Rundel. Open Intro Statistics. 3rd Edition. OpenIntro.org

[11] Reetu and N.Kumar (2015), Medical Diagnosis for Liver Cancer Using Classification Techniques, International Journal of Recent Scientific. Volume 6. Issue, 6, pp 4809-4813.

[12] Tina R. Patil, Mrs. S. S. Sherekar. Performance Analysis of Naïve Bayes and J48 Classification Algorithm for Data Classification. International Journal Of Computer Science And Applications Vol. 6, No.2, Apr 2013.