WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Liver Disease Prediction Technique using Machine Learning

Chauhan Bhavna, Prof. Saket Swarndeep

Chauhan Bhavna, Post Graduate Scholar, Post Graduate Department LJ University
Assistent Prof.Saket Swarndeep , Head of Department, Post Graduate Department, LJ University

ABSTRACT :-

Every year, liver disease is the leading cause of death. After cancer, stroke, and respiratory disease, liver disease is the fifth leading cause of death in England. Chronic hepatitis B and C, as well as alcohol and non-alcoholic beverages, are the leading causes of liver disease globally. Machine Learning has a lot of potential for automated disease diagnosis. With the current rise in numerous liver disorders, it's more important than ever to detect liver disease early on. We propose a novel classifier by combining a genetic algorithm with the XGBoost classifier. With feature selection, this compares various classification algorithms and visualization techniques used to forecast liver disease. Outlier detection is used to identify severe deviation values, while isolation forest is used to eliminate them. Accuracy, precision, recall f-measure, and time complexity are all used to evaluate performance.

. **Keywords**:- Non-alcoholic ,fatty liver disease,alcoholic Human life, classification algorithm, Linear Regression, SVM, Machine Learning

Introduction:-

Machine Learning techniques have become increasingly essential in the healthcare industry for disease prediction using medical databases. Machine learning is being used by many researchers and firms to improve medical diagnosis. Classification algorithms are commonly utilized in disease prediction among other machine learning approaches. For the prediction of liver illness, Logistic Regression, KNearest Neighbor, and Support Vector Machines are employed in this research. We all know that the liver is the body's largest internal organ, performing vital functions such as producing blood clotting components and proteins, triglyceride and cholesterol manufacturing, glycogen synthesis, and bile generation. In order for a loss in function to occur, more than 75% of the liver tissue must be impacted. 1 As a result, it's critical to diagnose the disease early on so that it can be treated before it becomes serious.

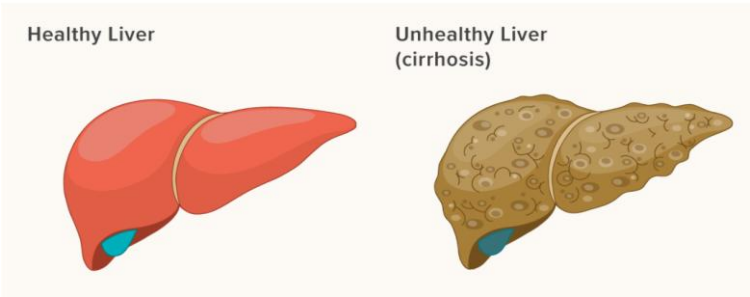

Figure 1.1: Healthy and unhealthy liver [29]

The quantities of proteins, liver enzymes, and bilirubin in your blood will be measured during a liver function test. These measurements can be used to assess your liver function and look for symptoms of inflammation or damage.

Digestion, detoxification, protein synthesis, storage and filtration of some chemicals, and other metabolic tasks are all performed by the liver. The liver's malfunction could result in serious health problems. As a result, it is critical to look after our liver. Poor dietary habits, such as eating fatty foods, drinking alcohol, self-medicating, and leading an unhealthy/sedentary lifestyle, can harm the liver and lead to death.

Alcohol Consumption Control:- Every day, the human liver must conduct roughly 500 activities. A substantial amount of alcohol drinking causes fat to build up around the liver, causing alcoholic fatty liver disease, which can be fatal. The only cure is to stop drinking alcohol or cut back as much as possible.

Drink Plenty of Water: It is always helpful to drink plenty of water. It aids in the removal of toxins from the body and reduces blood thickening, allowing for easier filtration through the liver.

Regular Exercise: One of the most important aspects is to begin exercising. Obesity and overweight are linked to non-alcoholic fatty liver disease (NAFLD). By maintaining a healthy weight, regular exercise can help avoid liver disease. Eat a healthy diet to maintain your liver in good shape. Reduce your carb intake and increase your intake of nuts. Controlling calorie intake with a low carb diet helps to maintain a healthy liver.
Reduce your intake of fizzy drinks: Many cold drinks on the market include added sugars that contribute to obesity. Reduce your consumption of these beverages.

Figure 1.2: progressive of liver disease

Prothrombin Time (PT): A prothrombin time, or PT, is a test used to determine whether or not someone is taking the proper dose of the blood thinner warfarin (Coumadin). It also looks for problems with blood coagulation.

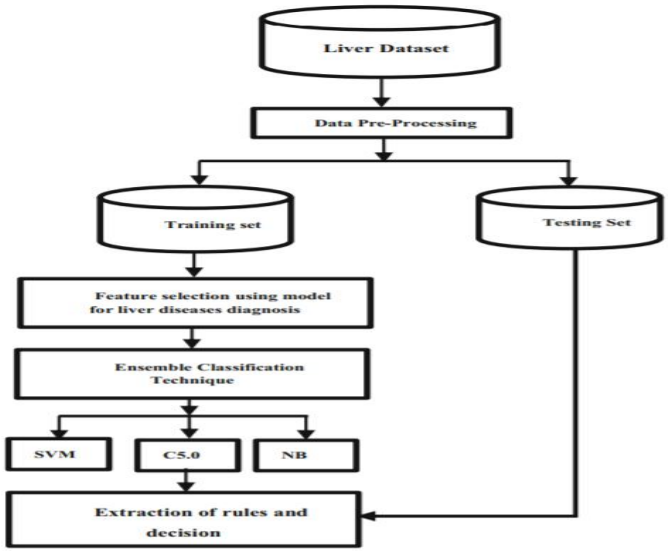

Figure 1.3 : Proposed System

Types of liver Disease

Hepatitis is a liver inflammation caused by viruses such as hepatitis A, B, and C. Hepatitis can also be caused by non-infectious factors such as excessive drinking, medications, allergic reactions, or obesity.
Cirrhosis: Long-term liver injury from any source can result in persistent scarring, which is known as cirrhosis. The liver is then unable to operate properly.
Hepatocellular carcinoma, the most prevalent type of liver cancer, nearly often develops after cirrhosis has developed.Liver failure can be caused by a variety of factors, including infection, hereditary illnesses, and excessive alcohol consumption.Cirrhosis causes the liver to leak fluid into the abdomen, causing it to become bloated and heavy.
Gallstones: Hepatitis and bile duct infection (cholangitis) can occur if a gallstone becomes lodged in the bile duct draining the liver.
Hemochromatosis is a condition in which iron deposits in the liver, causing damage. The iron deposits all over the body, creating a slew of severe health issues.

Primary sclerosing cholangitis is an uncommon condition that produces inflammation and scarring in the bile ducts of the liver. Its etiology is unclear. Primary biliary cirrhosis is an uncommon illness in which the bile ducts in the liver are slowly destroyed by an unknown process. Cirrhosis (permanent scarring of the liver) occurs over time.

Machine learning techniques

Machine learning algorithms have lately been employed as a non-invasive way for staging chronic liver illnesses in order to avoid the limitations of biopsy. The goal of this study is to compare different machine learning algorithms for predicting advanced fibrosis by combining blood biomarkers and clinical data to create classification models.

Naive Bayes classifiers are a type of simple probabilistic classifier based on applying Bayes' theorem to characteristics with strong (naive) independence assumptions. They are one of the most basic Bayesian network models, but when combined with kernel density estimation, they can achieve higher levels of accuracy.

The number of parameters required by Naive Bayes classifiers is linear in the number of variables (features/predictors) in a learning problem. Instead of expensive iterative approximation, which is used for many other types of classifiers, maximum-likelihood training can be done by evaluating a closed-form expression, which requires linear time.

Artificial neural networks, often known as neural networks, are computational algorithms.

They want to imitate the behavior of neuron-based biological systems. ANNs are computational models based on the central nervous systems of animals.

The KNN algorithm is a simple, easy-to-implement supervised machine learning technique that can address both classification and regression issues.

The KNN algorithm believes that objects that are similar are close together. To put it another way, related items are close together. The usefulness of the KNN algorithm is contingent on this assumption being correct.

KNN combines the concept of similarity (also known as distance, proximity, or closeness) with basic arithmetic we may have learnt as children while calculating the distance between points on a graph.

The Support Vector Machine (SVM) algorithm is a simple but powerful Supervised Machine Learning approach that may be used to create regression and classification models. Both linearly separable and non-linearly separable datasets can benefit from the SVM method. Even with a small amount of data, the support vector machine method performs admirably.

The goal of the SVM algorithm is to find a hyperplane in an N-dimensional space that categorizes data points clearly. The hyperplane's size is determined by the number of features. If there are only two input characteristics, the hyperplane is just a line. When there are three input features, the hyperplane becomes a two-dimensional plane. In other words, given n features, the hyperplane is (n-1)- D plane.

Conclusion

The approaches for identifying liver illness in patients have been proposed and assessed in this paper utilizing machine learning techniques. Logistic deterioration and SVM are two of the most common machine learning algorithms. The prediction analysis was carried out using all of the models, and their performance was evaluated. The probability of liver disease prediction was achieved with a 96 percent accuracy. The current scenario can be compared to other strategies in the future, such as nave bayes classification, Random forest, and so on. This work might also be focused on using bio-inspired optimization techniques to implement parametric classifications.

References:-

[1] Ailem, M., Role, F. and Nadif, M., 2015, October. Co-clustering document-term matrices by direct maximization of graph modularity. In Proceedings of the 24th ACM International on Conference on Information and Knowledge Management (pp. 1807-1810). ACM.

[2] Ailem, M., Role, F. and Nadif, M., 2017. Sparse poisson latent block model for document clustering. IEEE Transactions on Knowledge and Data Engineering, 29(7), pp.1563-1576.

[3] Liao, K., Liu, G., Xiao, L. and Liu, C., 2013. A sample-based hierarchical adaptive K-means

[4] clustering method for large-scale video retrieval. Knowledge-Based Systems, 49, pp.123- 133.

[5] Onan, A., Bulut, H. and Korukoglu, S., 2017. An improved ant algorithm with LDA-based representation for text document clustering. Journal of Information Science, 43(2), pp.275- 292.

[6] Rossi, R.G., Marcacini, R.M. and Rezende, S.O., 2013. Benchmarking text collections for classification and clustering tasks. Institute of Mathematics and Computer Sciences, University of Sao Paulo.

[7] Jain, A.K., 2010. Data clustering: 50 years beyond K-means. Pattern recognition letters, 31(8), pp.651-666.

[8] Park, H.S. and Jun, C.H., 2009. A simple and fast algorithm for K-medoids clustering. Expertsystems with applications, 36(2), pp.3336-3341.

[9] Neha, D. and Vidyavathi, B.M., 2015. A survey on applications of data mining using clustering techniques. International Journal of Computer Applications, 126(2).

[10] Djenouri, Y., Belhadi, A. and Belkebir, R., 2018. Bees swarm optimization guided by data mining techniques for document information retrieval. Expert Systems with Applications, 94,pp.126-136.

[11] Reddy, G.S., Rajinikanth, T.V. and Rao, A.A., 2014, February. A frequent term based text clustering approach using novel similarity measure. In 2014 IEEE International Advance Computing Conference (IACC) (pp. 495-499). IEEE.

[12] Alhawarat, M. and Hegazi, M., 2018. Revisiting K-Means and Topic Modeling, a Comparison Study to Cluster Arabic Documents. IEEE Access, 6, pp.42740-42749.

[13] Chen, Y. and Sun, P., 2018, August. An Optimized K-Means Algorithm Based on FSTVM. In 2018 International Conference on Virtual Reality and Intelligent Systems (ICVRIS) (pp. 363-366). IEEE.

[14] Janani, R. and Vijayarani, S., 2019. Text document clustering using Spectral Clustering algorithm with Particle Swarm Optimization. Expert Systems with Applications, 134, pp.192-200.

[15] Jin, C.X. and Bai, Q.C., 2016, June. Text clustering algorithm based on the graph structures of semantic word co-occurrence. In 2016 International Conference on Information System and Artificial Intelligence (ISAI) (pp. 497-502). IEEE.

[16] Afzali, M. and Kumar, S., 2019, February. Text Document Clustering: Issues and Challenges. In 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon) (pp. 263-268). IEEE.

[17] 16] H. Bunke, M. Roth and E.G.Schukat-Talamazzini, "Offline Cursive Handwriting Recognition Using Hidden Markov Models", Pattern Recognition, Vol. 28, No. 9, 1995 Elsevier Science Ltd.

[18] [17] Nafiz Arica, "An Off-line Character Recognition System for free style Handwriting", a thesis submitted to the graduate school of natural and applied sciences of the middle east technical university,1998.

[19] [18] Yong Haw Tay, Pierre-Michel Lallican, Marzuki Khalid, Christian Viard-Gaudin, Stefan Knerr," An Offline Cursive Handwritten Word Recognition System", IEEE Catalogue No. 01 CH37239 2001.

[20] [19] Anshul Gupta, Manisha Srivastava and Chitralekha Mahanta," Offline Handwritten Character Recognition International Conference on Computer Applications and Industrial Electronics (ICCAIE), 2011.

[21] [20] J. Pradeep, E. Srinivasan and S. Himavathi,"Neural Network Based Recognition System Integrating Feature Extraction and Classification for English Handwritten", International Journal of Engineering (IJE) Transactions B: Applications Vol. 25, No. 2, (May 2012) 99-106.

[22] [21] D. K. Patel, T. Som and M. K Singh," Improving the Recognition of Handwritten Characters using Neural Network through Multiresolution Technique And Euclidean Distance Metric", International Journal of Computer Applications (0975 – 8887) Volume 45– No.6 May 2012.

[23] [22] M. Blumenstein, B. Verma and H. Basli," A Novel Feature Extraction Technique for the Recognition of Segmented Handwritten Characters", Proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDAR'03) 0-7695-1960-1/03 $17.00 © 2003 IEEE

[24] [23] Sumedha B. Hallale, Geeta D. Salunke, " Twelve Directional Feature Extraction for Handwritten English Character Recognition", International Journal of Recent Technology and Engineering (IJRTE)ISSN:2277-3878, Volume-2, Issue-2, May 2013.

[25] [24] Amit Choudhary, Rahul Rishi and Savita Ahlawat, "Off-Line Handwritten Character Recognition using Features Extracted from Binarization Technique", 2212-6716 © 2013 American Applied Science Research Institute doi:10.1016/j.aasri.2013.10.045

[26] [25] Rafael M. O. Cruz, George D. C. Cavalcanti and Tsang Ing Ren," An Ensemble classifier for offline Cursive character recognition using multiple features Extraction technique", 978-1-4244-8126- 2/10/$26.00 ©2010 IEEE

[27] [26] Singh, Sameer, Mark Hewitt,"Cursive Digit And Character Recognition on Cedar Database", Pattern Recognition, 2000.Proceedings. 15th international conference on. Vol. 2. IEEE 2000.

[28] [27] K. Gaurav and Bhatia P. K., "Analytical Review of Preprocessing Techniques for Offline Liver disease predictoin", 2nd International Conference on Emerging Trends in Engineering & Management, ICETEM, 2013.