Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

PREDICTION OF HEPATITIS DISEASE USING THE MACHINELEARNING TECHNIQUE

Dr. M. Anantha Guptha Associate professor, Department of ECE, GATES Institute of Technology, Gooty.

N.M.Eswar Adithya[2], A.Deekshitha[3], J.Girish Kumar[4], G.Kiran Kumar [5] M.Akhil Kumar[6] UG student, Department of ECE, GATES Institute of Technology, Gooty.

Abstract:

This work is to choose the best tool forthe diagnosis and detection of hepatitisas well as for the prediction of the life expectancy of hepatitis patients, In this work, A comparative study betweenvarious Machine learning tools and Neural networks were carried out. Theperformance metrics based on the accuracy rate and the mean square error The machine learning (ML)algorithms such as Random Forest Classifiers (RFC) , K Nearest Neighbor (KNN) , and Artificial Neural Networks (ANN) were Considered as the classification and prediction tools for diagnosing Hepatitis disease. A brief study on the above algorithms were performed based on the pre diction accuracy of disease diagnosis. All the ML Algorithms were implemented and validated usingpython as platform. Hepatitis Bincluding chronic liver disease is quite common in this world, and may cause damage to hepatocytes. The range may form healthy carrier to decompensate cirrhosis. I have medicine, diagnosis is the "Recognition of a disease or stipulation by its apparent signs and symptoms". An important issue in medical diagnosis is a stratification, which refers to the sorting of patients based on the severity of the disease.

In case the clinical problems lies beyond the physician's competence, thesolution is to consult a specialist, can ncommon, expert opinion is either unavailable or not available in a timely fashion. The physician is left without the adequate time to devote to each case and struggling to keep up with thenewest on developments in his field owing too or to increasing expectationsof the highest in quality health care andthe rapid grow of ever more detailed medical knowledge In this, a the per wehave described and the Hepatitis can be generalized.

1. Introduction

Hepatitis is a medical condition characterized by inflammation of the liver. The liver is an essential organ in the body that performance many vital functions, including detoxifying harmful substances, producing bile, storing glucose, and synthesizing proteins. Hepatitis can be caused by various factors, including viral infections, alcohol, drugs, autoimmune diseases, and metabolic disorders. Symptoms of hepatitis can vary depending on the cause and severity of the disease. Some of the com Mon symptoms include fatigue, jaundice (yellowing of the skin and eyes), abdominal pain, nausea, vomiting, and loss of appetite. In some cases,hepatitis can cause chronic liver disease liverfailure, or liver cancer.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

2. PROPOSED METHOD

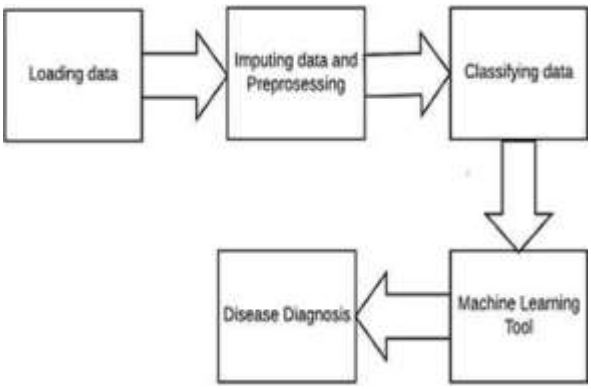
We proposed a machine learning-based diagnosis method for the identification of Hepatitis disease in this work. Machine learning predictive models include ANN, RFC, KNN and SVM are used for the identification of Hepatitis disease. The data set is extracted from UCI repository which consists of 155 instances with 20 attributes. Inthis hepatitis database, out of 155 instances 75of them have missing values. To get sufficient data for training, validation and testing, the data augmentation technique was used.

3. Working

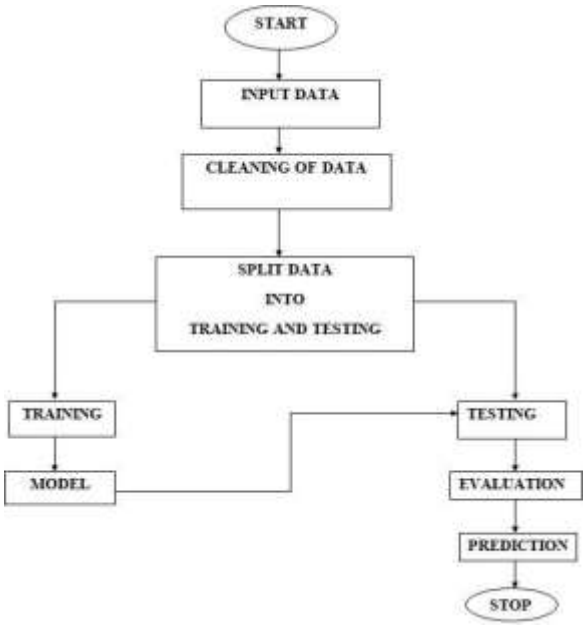
The data set is extracted from UCI repository which consists of 155 instances with 20 attributes. Since ML learns from examples, sufficient and smoothened data have to be given to the network model. To get sufficientdata, data imputation was performed on the available dataset.

The presence of missing data in any field, if not handled wisely may result in incorrect prediction and affects the quality of the result. In this hepatitis database, out of 155 instances
75 of them have missing values. To get sufficient data for training, validation and testing, the data augmentation technique was used. Data augmentation was performed thrice to increase the precision of the results. Out of all the instances of the dataset, the missing values corresponding to the instancesare removed and the imputation process was performed on the basis of remaining data. Theperformance of the imputation method was evaluated using the metric error rate, in which the imputed data is compared with the data with non-missing values in the attribute. After loading the data and pre-processing phase, classification of the data is carried out. Data

Classification is carried out in two phases, namely training phase and classification phase. In the training phase, the data is classified into training set and validation set. After that, the classifier algorithm builds the classifier with the training dataset. In the second classification phase, the trained modelis used for disease classification and the life expectancy of the Hepatitis person. The data is divided into training (60%), testing (20%) and validation (20%) in a stratified manner. There are mainly three stages in implementing the machine learning code as well as neural network namely; training, validating and testing. In our research, with the given hepatitis dataset, initially the data was split in to these three categories using stratified splitting. After this, we train the datausing suitable machine learning tool and thenthe data for validation is given to the network. Using the trained network, the testing data is validated, and this is the phase which gives the accuracy of prediction of life expectancy of patients with Hepatitis. SVM, KNN and neural network are the ML tools that we haveused in this work.



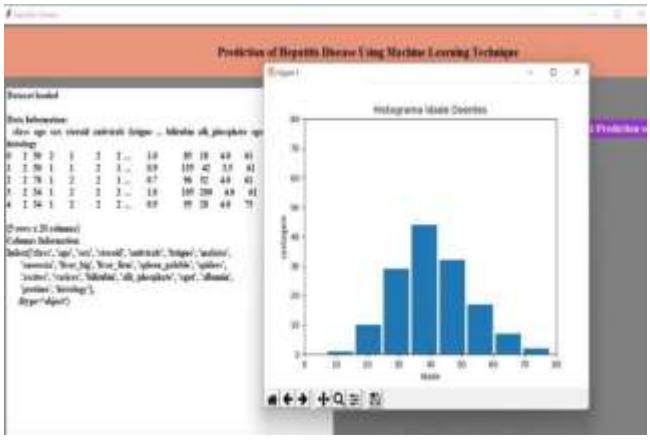
Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023



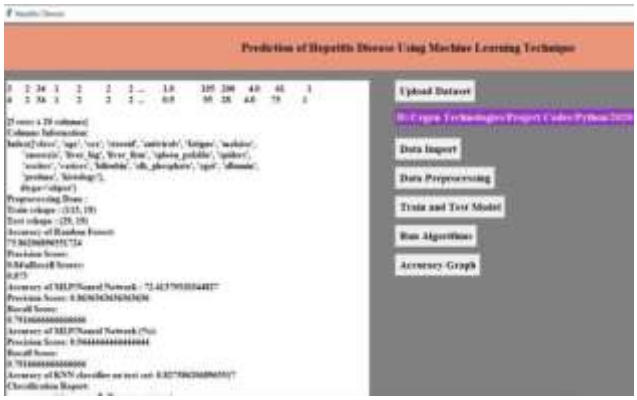
4. Results and Discussion

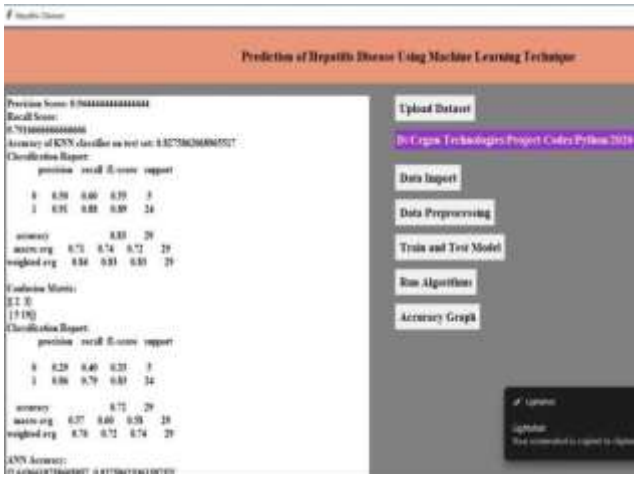
A histogram is a graphical representation of the distribution of a set of continuous data. Itis a bar graph-like representation where the data is divided into a set of intervals, called bins, and the height of each bar represents the number of data points that fall into that bin. The width of each bin represents the range of values that fall into that bin. Histograms are commonly used in data analysis to visualize the shape of the distribution of the data.

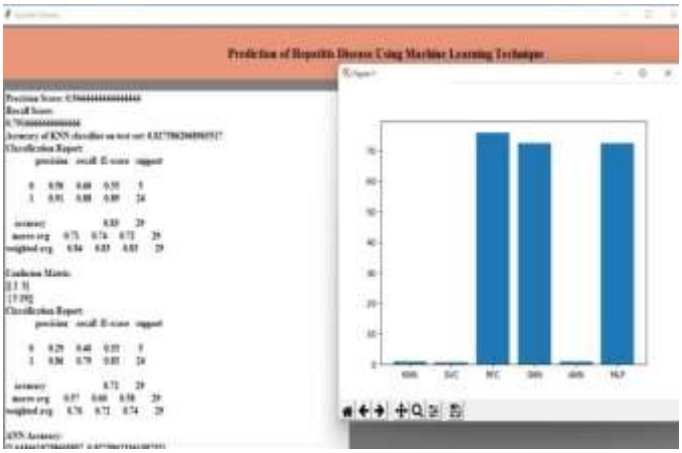
By examining the histogram, we can identify features such as the central tendency, spread, skewness, and the presence of outliers in the data. Histograms are particularly useful whendealing with large data sets, as they can quickly provide an overview of the data distribution. They are also a useful tool for identifying patterns in data and for identifyingtrends over time. Histograms can be created using many statistical software packages, and many spreadsheet programs also include built-in tools for creating histograms. The shape of the histogram is affected by the number of bins, so it's important to choose anappropriate number of bins to ensure that the distribution is accurately represented. Herethe histogram is taken between count of patients and their respective age and divided into set of similar continuous data.



Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023







hepatitis diagnosis and treatment, leading to better patient outcomes. However, it's crucial to note that machine learning models should complement expert medical diagnosis rather than replacing it. Overall, the application of machine learning in hepatitis prediction is a significant development in healthcare that canmake a significant impact on patient care andoutcomes.

5. Conclusion

In summary, machine learning techniques have demonstrated great potential in predicting hepatitis based on patient data. Several algorithms such as Support Vector Machines (SVMs), Random Forest Classifier (RFC), and Artificial Neural Networks (ANNs) have shown high accuracy and sensitivity in this task. These models can be trained on various features such as clinical and laboratory data and can provide useful insights into the underlying factors that contribute to hepatitis.

The use of machine learning in healthcare hasthe potential to improve the accuracy of

References

[1] M. A. Chinnaratha, G. P. Jeffrey, G. Macquillan, E. Rossi, B. W. D. Boer, D. J. Speers, and L. A. Adams, “Prediction of morbidity and mortality in patients with chronic hepatitis c by non- invasive liver fibrosis models,” Liver International, vol. 34, no. 5, pp. 720–727, 2014.

[2] Roslina, A. H., and A. Noraziah. "Prediction of hepatitis prognosis using Support Vector Machines and Wrapper Method." In 2010 Seventh International Conference on Fuzzy Systems and

Knowledge Discovery, vol. 5, pp. 2209-2211. IEEE, 2010.

[3] G. H. Haydon, R. Jalan, M. Alakorpela, Y. Hiltunen, J. Hanley, L. M. Jarvis, C. A. Ludlum, and P. C. Hayes, "Prediction of cirrhosis in patients with chronic hepatitis c infection by artificial neural network analysis of virus and clinical factors," Journal of ViralHepatitis, vol.5, no. 4, pp. 255–264, 2010.

[4] Atif Khan, John A. Doucette, Robin Cohen, "Integrating Machine Learning into a Medical Decision Support System to Addressthe Problem of Missing Patient Data", 2012 IEEE DOI 10.1109/ICMLA.2012.82.

[5] Uhmn, Saangyong, Dong-Hoi Kim,Young-Woong Ko, Sungwon Cho, Jaeyoun Cheong, and Jin Kim. "A study on application of single nucleotide polymorphism and machine learning techniques to diagnosis of chronic hepatitis." Expert Systems 26, no. 1 (2009): 60-69.

[6] Kayvan Joo, Amir Hossein, Mansour Ebrahimi, and Gholamreza Haqshenas. "Prediction of hepatitis C virus interferon/ribavirin therapy outcome based on viral nucleotide attributes using machine learning algorithms." BMC research notes 7, no. 1 (2014): 565.

[7] Vijaya Rani, S., and S. Dayananda. "Liver disease prediction using SVM and Naïve Bayes algorithms." International Journal of Science, Engineering and Technology Research (IJSETR) 4, no. 4 (2015): 816-820.

[8] Sartakhti, Javad Salimi, Mohammad Hossein Zangooei, and Kourosh Mozafari. "Hepatitis disease diagnosis using a novel hybrid method based on support vector machine and simulated annealing (SVMSA)." Computer methods and programs in biomedicine 108, no. 2 (2012): 570-579.

[9] Ba-Alwi, Fadl Mutahir, and Houzifa M. Hintaya. "Comparative study for analysis the prognostic in hepatitis data: data mining approach." spinal cord 11 (2013): 12.

[10] Yildirim, Pinar. "Filter based feature selection methods for prediction of risks in hepatitis disease." International Journal of Machine Learning and Computing 5, no. 4 (2015): 258.

[11] Chen, Hui-Ling, Da-You Liu, Bo Yang, Jie Liu, and Gang Wang. "A new hybrid method based on local fisher discriminant analysis and support vector machines for hepatitis disease diagnosis." Expert Systems with Applications 38, no. 9 (2011): 11796-11803.

[12] Lalithamani N., Amrutha C., "Efficient storage and retrieval of medical records using fusion- based multimodal biometrics", *International Journal of Computer Aided Engineering and Technology*, Volume 10, Issue 5, 2018.