Prediction of Hepatitis Disease Using Machine Learning Techniques

Dr. P.Chandra Kanth[1]
Department of Computer Science & Engineering,
Audisankara College of Engineering and Technology, Gudur, AP

Shaik Riyaz[2]
PG scholar, Department of MCA,
Audisankara College of Engineering and Technology, Gudur, AP

Abstract - The objective of this work is to choose the best tool for diagnosis and detection of Hepatitis as well as for the prediction of life expectancy of Hepatitis patients. In this work, a comparative study between various machine learning tools and neural networks were carried out.The performance metric is based on the accuracy rate and the mean square error. The Machine Learning (ML) algorithms such as Support Vector Machines (SVM), K Nearest Neighbor (KNN) and Artificial Neural Network (ANN) were considered as the classification and prediction tools for diagnosing Hepatitis disease. Brief studies on the algorithms were performed based on the prediction accuracy of disease diagnosis. All the ML algorithms were implemented and validated using MATLAB software.

Keywords – Artificial Neural Network (ANN), Classification, Prediction, Machine learning, Classifiers
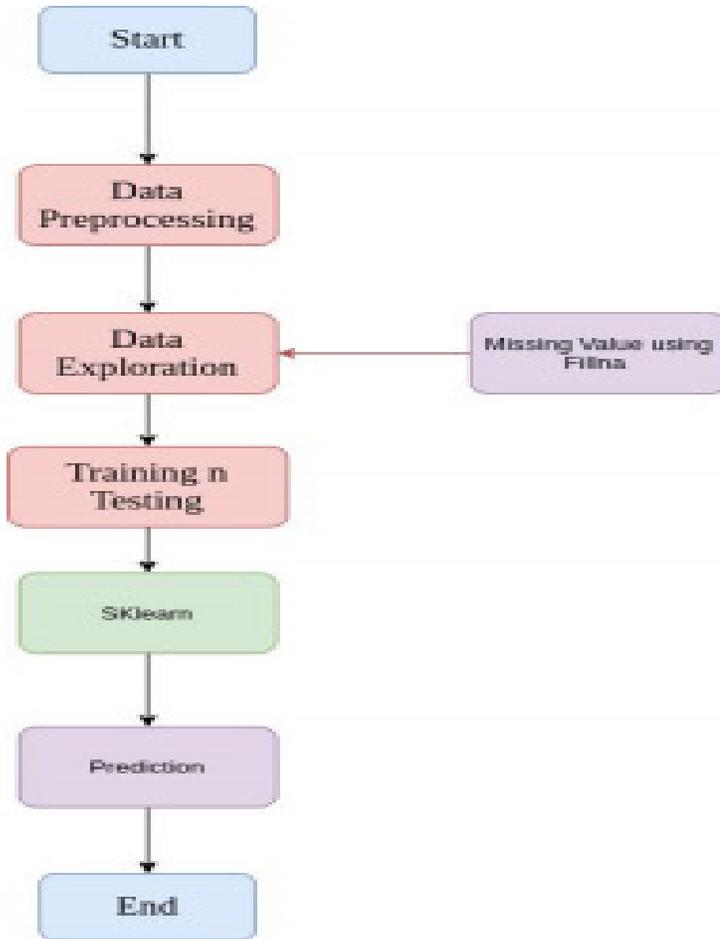
INTRODUCTION

Medical diagnosis is an important and a quite complex task which requires accurate identification. It is important to diagnose the disease at proper time and to be cured at the earliest. Liver is the vital part of a human body. One of the severe diseases that affect the functionality of liver is hepatitis, which causes inflammation of the liver. The main factor for Hepatitis disease is the presence of virus in liver [1]. Hepatitis is a worldwide disease with high mortality rate. If accurate measures are not taken in proper time, it may affect the vital functions of the body and may cause to cirrhosis, severe scarring and increase the risk of liver cancer [2]. Early detection through proper diagnosis and proper medication can cure the disease. For diagnosis of any disease, the two important things are: (i) The selection of right parameters of diagnosis and (ii) proper analysis of the data with an experienced expertise. Machine Learning

(ML) is the tool which could make a system to learn by itself by detecting different patterns and different relationships for the given data using different algorithms [5]. This would enable automatic diagnosis of any diseases, where the two important things considered with utmost care are: selection of parameters and the tool used for analyzing these parameters. In this work, a study of three different tools that are used for Hepatitis prediction namely: KNN, SVM and ANN are carried out. Different researches have undergone for the diagnosis and the prediction of diseases using machine learning techniques [14]. Somaya et al. evaluated different machine learning techniques in the prediction of advanced fibrosis that incorporates serum biomarkers [7]. Haydon et al. used artificial neural networks for the prediction of cirrhosis in patients using routine clinical host and viral parameters [4]. An automatic diagnosis system was proposed by Jiaxin et al. using extreme learning machine on serum indices data of patients to predict the fibrosis stage and inflammatory activity grade of chronic hepatitis C [8]. Sushrutha et al. proposed a hybrid model for the prediction of hepatitis [9]. They have developed a combination of genetic search algorithm and multilayer perceptron technique. The paper [10] investigates the impact of applying varied different fold for cross validation on missing values using PCA-MLP imputation method. Janhel et al. investigated different ANN models for the prediction of Hepatiti[11]. Cai et al. proposed a classification method using serum biochemistry data of patients in which a collaborative representation model is used with smoothly clipped absolute deviation to diagnosis chronic hepatitis C [3]. Rouhani et al designed various neural network such as RBF, GRNN, PNN, LVQ and SVM to diagnosis hepatitis disease and compared the performance [12].

Proposed System

In this work, the required data set is chosen from UCI repository, considering different clinical cases. This dataset consists of 155 instances with 20 attributes, one among the same attributes is the class to decide the life expectancy of a hepatitis patient. The 19 attributes for classification are shown in table I. Machine learning algorithm such as SVM and KNN were applied to the dataset for training and testing. Followed by this, neural network approach was performed on the same dataset for performance analysis. Comparison was evaluated based on the prediction accuracy of the tool used as well as the mean square error. Lower the mean square error, better the performance.

```
                    ┌──────────────┐
                    │    Start     │
                    └──────┬───────┘
                           │
                    ┌──────▼───────┐
                    │    Data      │
                    │ Preprocessing│
                    └──────┬───────┘
                           │
                    ┌──────▼───────┐         ┌──────────────────┐
                    │    Data      │◄────────│ Missing Value using│
                    │ Exploration  │         │      Filna        │
                    └──────┬───────┘         └──────────────────┘
                           │
                    ┌──────▼───────┐
                    │  Training n  │
                    │   Testing    │
                    └──────┬───────┘
                           │
                    ┌──────▼───────┐
                    │   SKlearn    │
                    └──────┬───────┘
                           │
                    ┌──────▼───────┐
                    │  Prediction  │
                    └──────┬───────┘
                           │
                    ┌──────▼───────┐
                    │     End      │
                    └──────────────┘
```

CONCLUSION

The Implementation is Phase where we endeavor to give the practical output of the work done in designing stage and most of Coding in Business logic lay coms into action in this stage its main and crucial part of the project.It is done by the developer itself in every stage of the project and fine-tuning the bug and module predicated additionally done by the developer only here we are going to solve all the runtime errors

In this work, different machine learning techniques and neural networks were used for the diagnosis of hepatitis. A comparison on the accuracy for a particular data set was performed by using various ML and ANN techniques, for identifying the best tool for Hepatitis disease diagnosis. We have used Support Vector Machine (SVM), Artificial Neural Network (ANN) and K Nearest Neighbor (KNN) to get the accurate prediction of the disease. With this study, it is inferred that out of all models considered and its performance, ANN is most accurate that gives a

good prediction accuracy of 96 percentage and a minimum mean square error. As a future work the same will be implemented using RNN for the prediction of occurrence of other diseases.

REFERENCES

[1] Ghumbre S. U.; Ghalot A.A, "Hepatitis B Diagnosis using Logical Inference And Self Organizing Map", 2008 ; Journal of Computer Science ISSN 1549-3636.

[2] 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.

[3] 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.

[4] 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 Viral Hepatitis, vol. 5, no. 4, pp. 255–264, 2010.

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

[6] Uhmn, Saangyong, Dong-Hoi Kim, Young-WoongKo, 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.

[7] KayvanJoo, Amir Hossein, MansourEbrahimi, and GholamrezaHaqshenas. "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.

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

[9] Sartakhti, JavadSalimi, Mohammad HosseinZangooei, and KouroshMozafari. "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.

[10] Uttreshwar, GhumbreShashikant, and A. A. Ghatol. "Hepatitis B diagnosis using logical inference and generalized regression neural98, no. 2 (2014): 370-379