IJRAR.ORG E-ISSN: 2348-1269, P-ISSN: 2349-5138

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYTICAL REVIEWS (IJRAR) | IJRAR.ORG

An International Open Access, Peer-reviewed, Refereed Journal

Forecasting Liver Disease Risk with Machine Learning Models

[1]Manoj Allenki, [2]Nakka Venkatesh, [3]Kailash Balerao, Goutham Macharla, Varshitha Vemula

[1345]Student, [2]Assistant Professor

[12345] Department of Computer Science Engineering (AI&ML)

[12345]CMR Engineering College, Hyderabad, Telangana

Abstract: In humans, the liver is the most essential organ in the body. Performs various functions, such as the production of bile, which aids in the digestion of fats. The process of eliminating bile and bilirubin from the body, as well as the breakdown of proteins, are essential functions of the liver. Carbohydrates, enzyme activation, and glycogen storage are all important processes in the body. Vitamins and minerals are essential for the synthesis of plasma proteins and the process of clotting. Factors: The liver is highly susceptible to damage as a result of consuming alcohol. Pain relievers, dietary choices, and incorporates a significant amount of technology. By utilizing various data mining techniques, it is feasible to expedite the process of diagnosing liver diseases. As more data is incorporated, the accuracy of the prediction will increase. To avoid the local storage scarcity experienced in many healthcare centers, cloud storage is used. As the documents generated are voluminous in size in health care centers cloud storage would be an appropriate choice. It is a very challenging task for the common person to predict the disease due to lack of awareness. To overcome this issue, in this study, we proposed a machine Learning approach using Logistic Regression, Support vector machine and Random Forest algorithms [1]. The performance of each algorithm is evaluated with respect to parameters like accuracy, sensitivity, precision and specificity then based on these parameters an appropriate algorithm is used for disease prediction and a user interactive website will be developed so the user will know about whether he/she has liver disease or not by entering the blood test report values in the website.

***Index Terms* -** Machine Learning, Liver Function, Logistic Regression, Support vector machine, Random Forest.

I. Introduction

Liver disease is a major global health concern, requiring prompt diagnosis for effective treatment. In the medical field, machine learning has become an essential tool for predictive analysis. By leveraging various patient data and sophisticated algorithms, machine learning can facilitate early detection and risk evaluation of liver diseases. This introduction discusses the importance of predicting liver disease, the process of data collection, associated challenges, and the potential for enhancing patient outcomes through the use of artificial intelligence in clinical practice.

II. LITERATURE REVIEW

Varun Vats, et.al (2018) considered three different ML (Machine Learning) algorithms. A comparison of these algorithms had been carried out for evaluating their forecasting accuracy and computing intricacy [3]. These algorithms included AP (Affinity Propagation), K means and DBSCAN. This work was dedicated to the medical dataset based on lever disorders. This work made use of the Silhouette coefficient to measure the comparative efficiency of the considered algorithmic approaches.

M. B. Priya, et.al (2018) stated that the healthcare sector had a lot of data but this data was of no use. This ample data required a leading analytic tool so that the hidden relationship and the valuable knowledge could be determined. The liver disease referred to the medical condition of the human liver-related to the human liver. The liver diseases led to sudden changes in health conditions that governed the functioning of the liver affecting other internal body organs. This work made use of several classification algorithms based on data mining. These algorithms included DT (Decision Tree), LD (Linear Discriminant), SVM Fine Gaussian, and LR (Logistic Regression) [6]. This work made use of Lab-based metrics of patients in the form of a liver dataset.

N Chandra Sekhar Reddy, et.al (2017) described different classification approaches by implementing them on the dataset of patients suffering from liver diseases. The main objective here was to accurately predict liver disorder by means of several data mining algorithms [7]. This work performed the analysis using the dataset of real-time patients to build classification paradigms for the prediction of liver diseases. This work implemented five classification algorithms on the used dataset. This work analyzed different metrics such as precision, recall, and accuracy for determining the efficiency of the implemented classification models [8].

III. PROPOSED METHODOLOGY

We are proposing a module that focuses on analyzing a dataset containing information about liver disease patients. There are several stages involved in this process.

In the initial stage of data pre-processing, the raw data is transformed into valuable information and external disturbances are eliminated. When there are missing values in the dataset, the process of imputation is used to replace them.

In the second stage of feature selection, the process of choosing the most relevant features is carried out. After completing the visualization of attributes to generate a statistical report on the health status of individuals, chemical compositions, and the gender ratio, the necessary data is obtained.

In the third stage, we analyze the training data and testing data to forecast the values and evaluate their accuracy in comparison to the actual values. Three distinct classification algorithms are employed, including logistic regression, support vector machine, and random forest classifier. After evaluating the accuracy score, we have chosen one classifier.

In the fourth stage, we create a web application that allows users to interact with websites, input their details, and receive a prediction on whether they are at risk of developing liver disease or not.

System Architecture:

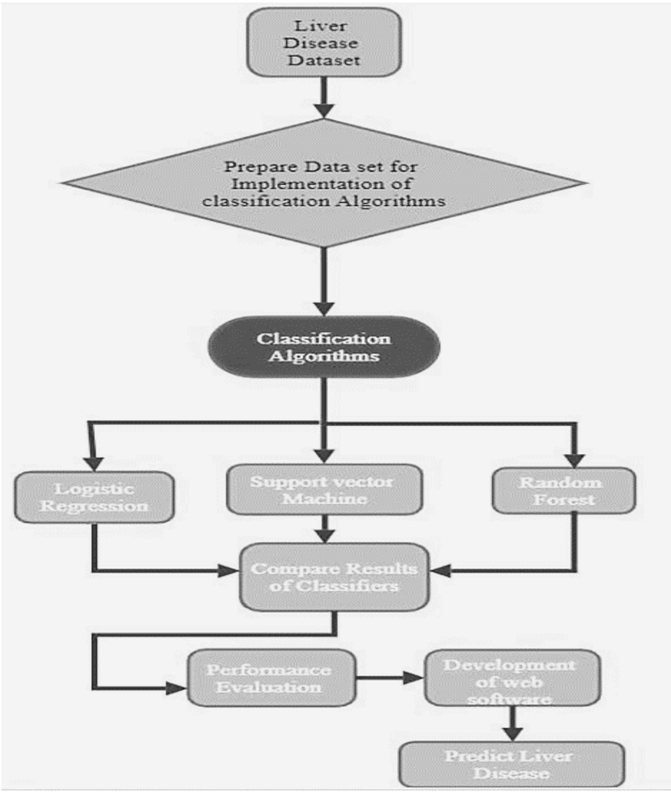


Figure 1: Liver Disease Prediction Model Architecture

Advantages of proposed methods:

No Medical Expertise Required: The system is designed to be user-friendly and accessible, enabling non-medical personnel and researchers without extensive medical knowledge to utilize it effectively. This broadens its usability to a wider audience.

High Accuracy: Machine learning models enhance the system's accuracy in predicting risk factors. These algorithms can analyze and learn from large datasets, resulting in more precise predictions than traditional statistical methods.

Immediate Results: The system delivers instant results, facilitating rapid decision-making. This is crucial in medical contexts where timely risk factor identification can lead to early interventions and improved patient outcomes.

Early Risk Factor Prediction: Machine learning models can be trained to detect risk factors at an early stage, often before symptoms or conditions manifest. Early detection allows for preventive measures and interventions, potentially mitigating the severity of health issues.

Cost & Time Effective: The machine learning-based system offers a cost-effective alternative to traditional diagnostic methods and expensive medical tests. It also saves time by automating the risk factor prediction process, reducing the need for manual analysis and diagnosis.

Scalability: The system's ability to efficiently manage large datasets makes it scalable to accommodate increasing volumes of medical data.

Continuous Learning: Machine learning models can continuously improve and adapt as they process more data, ensuring the system remains current and effective.

Personalized Medicine: By tailoring machine learning models to individual patient data, the system can provide personalized risk factor predictions and treatment plans.

Reduced Human Error: Automating the risk factor prediction process helps minimize the potential for human error that can occur during manual analysis.

Research and Insights: The system can extract valuable insights from data, fostering new discoveries and research opportunities in the medical field.

It is crucial to remember that while machine learning-based systems provide significant benefits, they are not a substitute for the expertise of medical professionals. These systems should serve as supportive tools to assist in decision-making and risk assessment, with medical experts responsible for validating and interpreting the results. Additionally, ensuring data privacy and security is essential when implementing such systems in the medical domain.

Use Case Diagram:

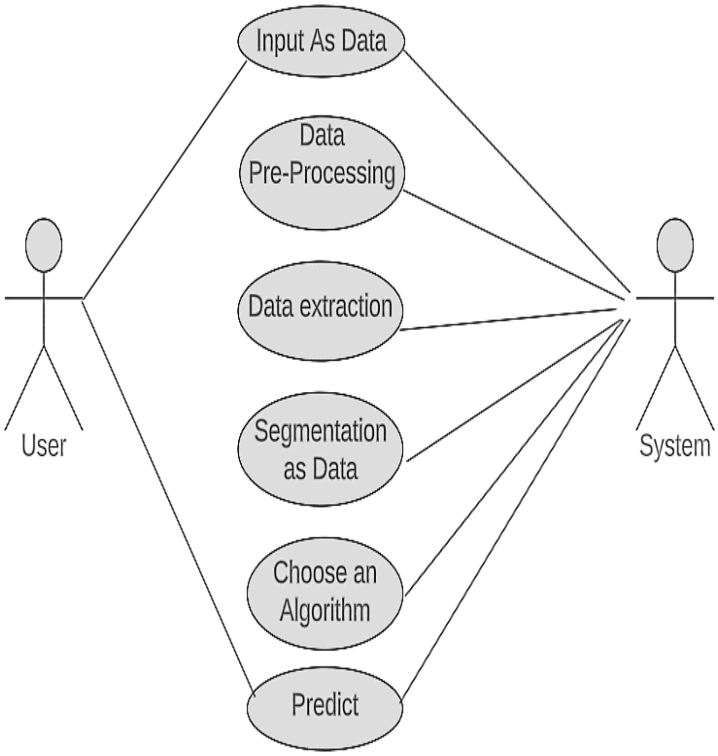
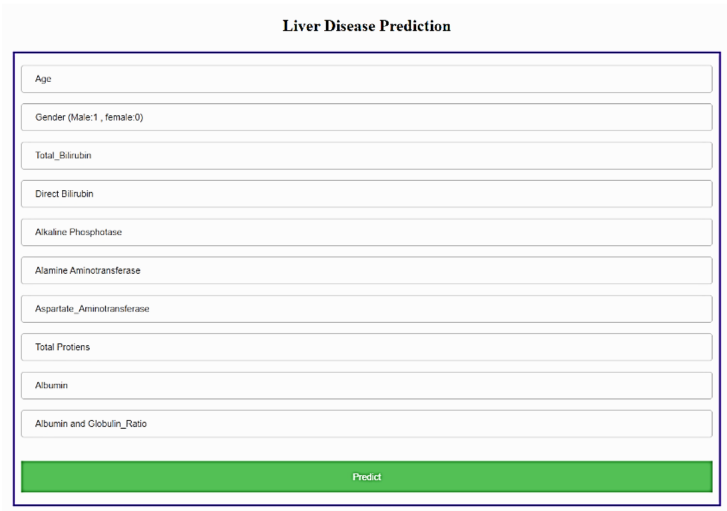


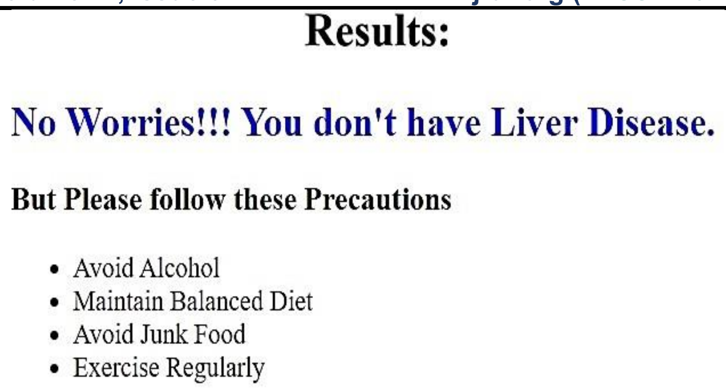
Figure 2: Use Case Diagram

IV. IMPLEMENTATION

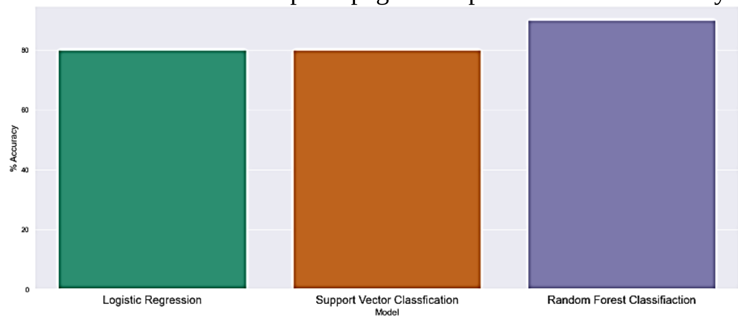


Above is the demonstration of our Prediction Page.

Above is the demonstration of our Explore page if the person is affected by the disease.



Above is the demonstration of our Explore page if the person is not affected by the disease.



Above is the accuracy graph of various classifiers.

V. CONCLUSION

The primary goal of this work is to develop an effective diagnostic system for chronic liver disease patients using various supervised machine learning classifiers. This application aims to predict liver disease early and provide health condition advisories. Such an application can be particularly beneficial in low-income countries where medical infrastructure and specialized experts are limited.

Our study suggests several directions for future research in this field. While we explored some popular supervised machine learning algorithms, incorporating more algorithms could build a more accurate liver disease prediction model and enhance performance further. Additionally, this work has the potential to play a significant role in healthcare research and medical centers by aiding in the prediction of liver disease.

Different classification algorithms, including Logistic Regression, Support Vector Machine, and Random Forest Classifier, were applied to the dataset. The comparison of these algorithms based on accuracy showed Logistic Regression and Support Vector Machine both at 80%, while Random Forest Classifier achieved 90%. Thus, our findings suggest that the Random Forest algorithm is a robust choice for early-stage liver disease prediction.

VI. SUGGESTIONS FOR FUTURE WORK

Incorporating more diverse and comprehensive data sets: Current liver disease predictors often rely on relatively small data sets and may not include all relevant variables. Expanding the data sets to include a wider range of patient characteristics, medical histories, and lifestyle factors could improve the accuracy of the models. Developing more advanced machine learning algorithms: While current machine learning algorithms such as logistic regression, support vector machine, and random forest have shown promise in liver disease prediction, more advanced algorithms such as deep learning or neural networks could provide even greater accuracy and predictive power.

Integrating real-time monitoring: In addition to predicting the risk of liver disease, real-time monitoring of liver function and disease progression could enable earlier intervention and treatment. Developing predictive models that can incorporate real-time monitoring data could improve patient outcomes and reduce health care costs.

VII. REFERENCES

[1] Anju Gulia, Dr. Ranjan Vohra, Praveen Rani, “Liver Patient Classification Using Intelligent Techniques”, International Journal of Computer Science and Information Technologies, Vol. 5 (4), 2014, 5110- 5115.

[2] A.S. Aneesh Kumar, Dr. C. Jothi Venkateswaran, A novel approach for Liver disorder Classification using Data Mining Techniques, Engineering and Scientific International Journal, ISSN 2394- 7179, ISSN 2394- 7187,2015.

[3] Varun Vats, Lining Zhang, “A Comparative Analysis of Unsupervised Machine Techniques for Liver Disease Prediction”, Conference on IEEE International Symposium on Signal Processing and Information Technology, 2018.

[4] Lavesson. N, and Davidson. P., “Generic Methods for Multi- Criteria Evaluation”, in Proc. of the Siam Int. Conference on Data Mining, Atlanta, Georgia, USA: SIAM Press, 2008, pp. 541-546.

[5] S. Bagga, A. Girdhar, M. C. Trivedi. (2017), “SPMD based time sharing intelligent approach for image denoising”, Journal of Intelligent & Fuzzy Systems, vol.32, no. 5, pp. 3561-3573.

[6] M. B. Priya, P. L. Juliet, P.R. Tamilselvi. (2018), “Performance Analysis of Liver Disease Prediction Using Machine Learning Algorithms”, International Research Journal of Engineering and Technology, vol. 5 no. 1, pp. 206-211.

[7] N Chandra Sekhar Reddy, Purna Chandra Rao Vemuri, A Govardhan, " Performance Analysis of Indicative Feature Subset Selection Algorithm for Effective Intrusion Detection", International Journal of Advanced Science and Technology, Vol. 29, Number 6 (2020), pp. 7-20, ISSN 2005- 4238. [SCOPUS].38

[8] N Chandra Sekhar Reddy, Purna Chandra Rao Vemuri, A Govardhan, "Evaluation of PCA and K means Algorithm for Efficient Intrusion Detection”, International Journal of Applied Engineering Research, Vol. 12, Number 12 (2017), pp. 3370-3376, ISSN 0973- 4562. [SCOPUS].