International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

MACHINE LEARNING BASED APPROACH FOR LIVER DISEASE PREDICTION

D. NAGESWARA NAIK[1], K. SIVA SANKARA VARA PRASAD[2], S. SASI KUMAR[3], SK SHAHENA[4], S. AARTHI REDDY[5], Dr. M. VEERA KUMARI[6]

[1,2,3,4,5]B-Tech Student, Dept.of CSE,UNIVESAL COLLEGE OF ENGINEERING AND TECHNOLOGY,Andhra pradhesh, India [1]

[6]M.Tech, Ph.D, Associate Professor,HOD Dept.of CSE,UNIVERSAL COLLEGE OF ENGINEERING AND TECHNOLOGY,Andhra pradhesh, India[2]

dgxnageswaranaik@gmail.com , veerakumarimamidi@gmail.com

ABSTRACT

Liver disease is one of the leading causes of disease-related mortality worldwide, with survival rates heavily dependent on early detection and timely treatment. This project presents a Flask-based web application integrated with a machine learning model designed for early liver disease detection. The system employs a Random Forest Classifier trained on diverse patient health parameters, including liver function test results, demographic details, and clinical indicators, to assess an individual's risk of developing liver disease.The machine learning model, stored as Liver2.pkl, leverages robust decision-tree- based ensemble learning techniques to achieve high classification accuracy and predictive performance. Users can input their health parameters through a user- friendly web interface, which then provides a real-time risk assessment. Additionally, the application supports interactive features such as secure

Data set uploads for retraining the model, visualization of training metrics, and comprehensive result dashboards.The integration of advanced machine learning techniques within a simple and accessible web application ensures broader adoption among medical practitioners and researchers. This project contributes to the ongoing efforts to enhance liver disease screening methods and highlights the potential of AI-driven healthcare solutions for early disease detection and prevention.

KEYWORDS: Mortality,Survival,Demographic,Practitioners,Healthcare,Leverages, Medical,Interactive.

1. INTRODUCTION

Liver disease is one of the deadliest forms of disease, ranking among the top causes of disease-related deaths globally. Its aggressive nature and often late diagnosis make it challenging to treat effectively, emphasizing the critical need for early detection and timely intervention. Early identification of liver disease can significantly improve survival rates and enable healthcare providers to administer more effective treatment strategies.The emergence of machine learning and artificial intelligence (AI) in healthcare has revolutionized disease detection and prediction systems. Machine learning models can analyse complex datasets to uncover patterns and correlations that traditional statistical methods might overlook. These models are particularly useful for identifying early warning signs in diseases like liver disease,

where early symptoms can often be non-specific and easily missed.This project introduces a web-based liver disease detection system developed using the Flask framework and integrated with a machine learning model. The system employs a Random Forest Classifier trained on patient health parameters, including liver function test results, demographic factors, and clinical indicators, to assess an individual's likelihood of developing liver disease. This project addresses the critical need for accessible and efficient diagnostic tools for liver disease detection.By integrating machine learning with an intuitive web-based interface, it bridges the gap between advanced AI technologies and practical healthcare applications. Medical professionals, researchers, and even individuals concerned about their health can benefit from this system's ability to provide early assessments and valuable insights. In conclusion, this project demonstrates the potential of AI- driven healthcare solutions in enhancing disease detection and prevention. It aims to contribute to the ongoing efforts in leveraging machine learning for better health outcomes, ultimately saving lives and improving the quality of care for patients.

2. LITERETURE SURVEY

Liver disease detection and diagnosis have been widely studied in the field of medical research and machine learning. Various approaches, ranging from traditional diagnostic methods to AI-driven solutions, have been explored to improve early detection and enhance diagnostic accuracy. This section reviews previous studies and highlights the advancements in machine learning-based liver disease detection.

2.1 Traditional Diagnostic Methods for Liver Disease Detection:

Numerous studies have highlighted the limitations of conventional diagnostic techniques, such as liver function tests (LFTs), imaging techniques (CT, MRI), and biopsies.

Younossi et al. (2017) reported that conventional liver biopsy, while considered the gold standard, is invasive, costly, and subject to sampling errors.

Chang et al. (2018) emphasized that imaging techniques can miss early-stage tumors, making early diagnosis challenging. These limitations underscore the need for non-invasive, accurate, and cost- effective detection methods.

2.2 Machine Learning for Liver Disease Detection:

Kourou et al. (2015) provided an extensive review of machine learning applications in disease detection and concluded that ensemble methods, such as Random Forest, offer robust predictive capabilities.

Chen et al. (2019) developed a decision support system based on Random Forest for predicting liver disease risks using patient clinical parameters. The system achieved an accuracy of 91% on a publicly available dataset.

Zhang et al. (2020) proposed a support vector machine (SVM) model that

outperformed traditional logistic regression in predicting liver disease using clinical data.

2.3 Random Forest Classifier for Medical Predictions:

Random Forest Classifiers have emerged as a popular choice for medical predictions due to their ability to handle high-dimensional and non-linear datasets.

Breiman (2001) introduced the Random Forest algorithm, emphasizing its robustness and accuracy in various applications, including medical diagnostics.

upta et al. (2020) applied Random Forest for liver disease prediction and reported superior performance compared to other algorithms such as k- Nearest Neighbors &Naïve Bayes. The inherent feature selection mechanism and ensemble approach of Random Forest make it suitable for handling complex medical datasets.

2.4 Web-based Health Monitoring Systems:

Several studies have explored the integration of machine learning models with web-based applications for real-time health monitoring.

Razzak et al. (2018) developed a web-based platform for diabetes prediction using machine learning models and highlighted the importance of user-friendly interfaces for non-technical users.

Park et al. (2021) emphasized that Flask-based applications are lightweight and scalable for real-time medical predictions, making them an ideal choice for healthcare solutions.

2.5 Gaps and Challenges Identified in Literature:

Despite the advancements in machine learning-based liver disease detection, several challenges remain:

Data Scarcity: Limited availability of high-quality and diverse datasets hampers model generalization.

Data Privacy: Handling sensitive health data securely remains a critical concern.

Interpretability: Machine learning models often act as black boxes, making it difficult for healthcare professionals to interpret predictions.

Model Update Mechanisms: Few studies address dynamic model retraining and evaluation for evolving datasets**.**

2.6 Contributions of the Proposed Method:

The proposed system addresses several gaps identified in the literature by:

Utilizing a robust Random Forest Classifier for accurate liver disease prediction.

Offering a Flask-based web application for real-time predictions and model retraining.

Incorporating data visualization tools for better interpretability of model outputs.

3. EXISTING SYSTEM

Liver disease is one of the deadliest forms of disease, ranking among the top causes of disease-related deaths globally. Its aggressive nature and often late diagnosis make it challenging to treat effectively, emphasizing the critical need for early detection and timely intervention. Early identification of liver disease can significantly improve survival rates and enable healthcare providers to administer more effective treatment strategies.The emergence of machine learning and artificial intelligence (AI) in healthcare has revolutionized disease detection and prediction systems. Machine learning models can analyse complex datasets to uncover patterns and correlations that traditional statistical methods might overlook. These models are particularly useful for identifying early warning signs in diseases like liver disease, where early symptoms can often be non-specific and easily missed.

3.1 DIS-ADVANTAGES:

1) Its aggressive nature and often late diagnosis make it challenging to treat effectively, emphasizing the critical need for early detection and timely intervention.

2) Highly subjective and dependent on the clinician's experience.

3) Early-stage liver disease symptoms are often non-specific and difficult to identify

4) Elevated AFP levels are not specific to liver disease and may also present in benign liver conditions.

4. PRORPOSED SYSTEM

This project introduces a web-based liver disease detection system developed using the Flask framework and integrated with a machine learning model. The system employs a Random Forest Classifier trained on patient health parameters, including liver function test results, demographic factors, and clinical indicators, to assess an individual's likelihood of developing liver disease. This project addresses the critical need for accessible and efficient diagnostic tools for liver disease detection.By integrating machine learning with an intuitive web-based interface, it bridges the gap between advanced AI technologies and practical healthcare applications. Medical professionals, researchers, and even individuals concerned about their health can benefit from this system's ability to provide early assessments and valuable insights. In conclusion, this project demonstrates the potential of AI- driven healthcare solutions in enhancing disease detection and prevention. It aims to contribute to the ongoing efforts in leveraging machine learning for better health outcomes, ultimately saving lives and improving the quality of care for patients.

4.1 ADVANTAGES:

1) Medical professionals, researchers, and even individuals concerned about their health can benefit from this system's ability to provide early assessments and valuable insights.

2) This project demonstrates the potential of AI- driven healthcare solutions in enhancing disease detection and prevention.

3) It aims to contribute to the ongoing efforts in leveraging machine learning for better health outcomes, ultimately saving

lives and improving the quality of care for patients.

5. SYSTEM ARCHITECTURE

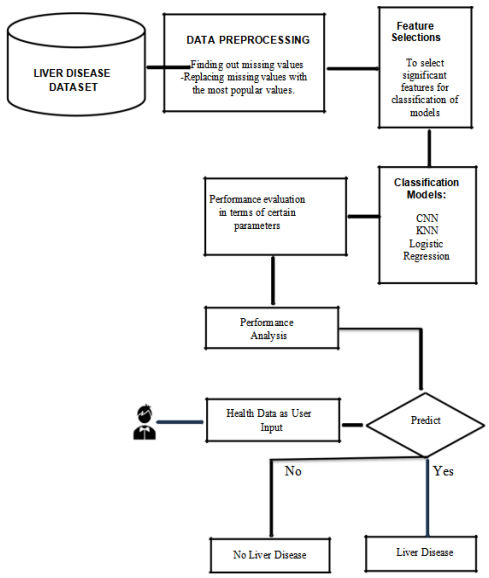

Fig 5.1 system architecture

6. RELATED WORK

A risk factor of a person may be predicted and individual can get treatment for their ailment. Treatment is needed since the incidence rate of cardiovascular illnesses is growing at an unexpectedly rapid pace and a large number of individuals are ignorant that the ML model will be more effective to induce in the task. As a direct result of anticipated that the prevalence of cardiovascular illnesses would continue to increase.

6.1 ALGORITHMS

In this module, implemented different algorithms with the dataset. The first step is training the dataset. In this implementation part 80% of the train dataset and 20%of the test dataset. Implemented various algorithms like CNN, Random Forest, Decision Trees, KNN algorithm.

6.2 Nearest Neighbors (KNN)

KNN is algorithm which is also called as the K-nearest neighboring algorithm. It is also called as the lazy learning algorithm. Its motivation is to utilize a database in which the information focuses are isolated into a few classes to anticipate the order of another example point. A KNN sorts an example to the class that is most decide among K neighboring. K is a limitation for adjusting the classification algorithms.

6.3 Random Forest

Random Forest is also called as Random Decision Trees. Random Forest Algorithm is a machine learning technique where these are a learning tasks, classification and regression tasks. Random forests or random decision forests are an ensemble learning technique for classification, regression and different assignments that works by developing a huge number of decision trees at training time and yielding the class that is the method of the classes (classification) or mean forecast(regression) of the individual trees. Random decision forests right for decision trees' propensity for overfitting to their training set. In the forest of trees has been the immediate connection between the combine trees and the outcome it can get. To get increasingly effective and precise Predictions, random forest inserts an additional layer of irregularity.

6.4 Logistics Regression

Calculated Regression was for the most part utilized in natural research and applications in the mid-20th century. Logistic regression can deal with any number of numerical as well as absolute factors. In addition, it introduces a discrete parallel item somewhere in the range of 0 and 1. Strategic Regression processes the connection between the element factors by surveying

probabilities(p) utilizing an underlying logistic function.

6.5 Decision Tree

Decision Tree calculation has a place with the supervised learning algorithms. In contrast to other supervised learning algorithms, a decision tree algorithm can be utilized for taking care of regression and classification issues as well. The general thought process of utilizing Decision Tree is to make a training model that can use to predict class or estimation of objective factors by taking in choice standards derived from earlier data (training data).

6.6 Support Vector Machine (SVM)

SVM algorithm tries to give out hyper planes and split the data into different categories. The scikit-learn package in python is employed for implementing SVM. The pre-processed information is split into check information and coaching set that is of twenty fifth and seventy fifth of the entire dataset severally. An SVM technique builds hyper planes in an exceedingly dimensional area. A decent separation is achieved by the hyper plane that has the most important distance to the closest coaching information of any category (so-called purposeful margin), since generally the larger the margin the lower the generalization error of the classifier.

7.RESULTS

Fig 7.1 above results explain overview of liver like healthy liver and damaged liver.

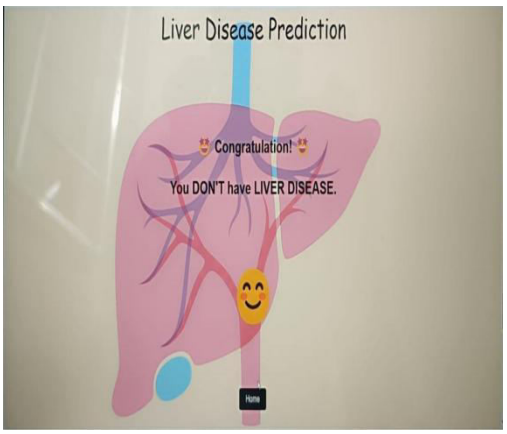

Fig 7.2 above results explain liver disease predict like nagative results like no liver disease.

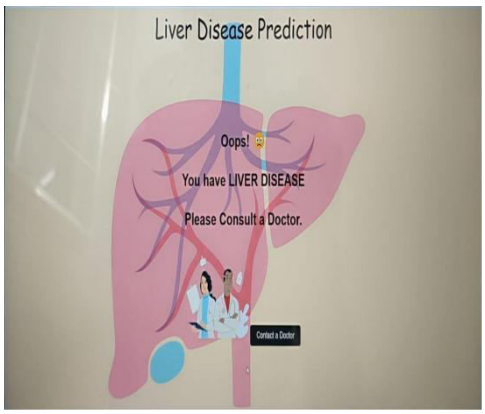

Fig 7.3 above results explain liver disease predict pasitive results like have liver disease.

8. CONCLUSION

This project demonstrates a robust and efficient approach to liver disease detection by integrating machine learning with a Flask-based web application. By utilizing a Random Forest Classifier trained on comprehensive health parameters, the system provides accurate predictions regarding the risk of liver disease, thereby assisting users and healthcare providers in early diagnosis and decision-making. The user-centric design allows seamless data input, model predictions, and visualization of results. Key features such as dataset

upload, dynamic model retraining, and visualization tools ensure adaptability and scalability for real-world healthcare applications. The system's architecture promotes efficient data handling, secure transactions, and a streamlined workflow from input to prediction. This solution underscores the transformative role of AI in healthcare, particularly in disease diagnostics, where early detection significantly improves survival rates. As a future enhancement, the model can be improved by incorporating larger and more diverse datasets, advanced algorithms, and clinical validation. Additionally, extending the system with cloud-based deployment and real-time patient monitoring capabilities can further enhance its utility and impact.

9. REFFERENCES

1. Md. F. Rabbi, S. M. Mahedy Hasan, A. I. Champa, Md. Asif Zaman, and Md. K. Hasan,"Prediction of Liver Disorders using Machine Learning Algorithms: A Comparative Study,"2020 2nd International Conference on Advanced Information and Communication Technology (ICAICT), Nov. 2020, https://doi.org/10.1109/icaict51780.2020.9333528.

2. C.-C. Wu et al., "Prediction of fatty liver disease using machine learning algorithms,"Computer Methods and Programs in Biomedicine, vol. 170, pp. 23–29, Mar. 2019, https://doi.org/10.1016/j.cmpb.2018.12.032.

3. M. Ghosh et al., "A Comparative Analysis of Machine Learning Algorithms to Predict Liver Disease," Intelligent Automation & Soft Computing, vol. 30, no. 3, pp. 917– 928, 2021, https://doi.org/10.32604/iasc.2021.017989.

4. N. Afreen, R. Patel, M. Ahmed, and Mustafa Sameer, "A Novel Machine Learning Approach Using Boosting Algorithm for Liver Disease Classification," Oct.2021, https://doi.org/10.1109/iscon52037.2021.9702488.

5. X. Pei, Q. Deng, Z. Liu, X. Yan, and W. Sun, "Machine Learning Algorithms for Predicting Fatty Liver Disease," Annals of Nutrition and Metabolism, vol. 77, no. 1, pp. 38–45, 2021, https://doi.org/10.1159/000513654.

6. Suryaprakash Reddy, Lamu Samuel Kiran, and Xavier, "Comparative Analysis of Liver Disease Detection using Diverse Machine Learning Techniques," May 2022, https://doi.org/10.1109/iciccs53718.2022.9788208.

7. Shaker, O. Hashim Yahya, A. Hadi N. Ali Hussein, H. TH. and T. A.-Q. , "A Predictive model for liver disease progression based on logistic regression algorithm," Periodicals of Engineering and Natural Sciences (PEN), vol. 7, no. 3, p. 1255, Sep. 2019, https://doi.org/10.21533/pen.v7i3.667.

8. k. Thirunavukkarasu; Ajay. S. Singh, Md. Irfan and A. Chowdhury, "Pre-diction of Liver Disease using Classification Algorithms", 2018 4th International Conference on Computing Communication and Automation (ICCCA). https://ieeexplore.ieee.org/document/8777655

9. G. Veena, D. Sneha, Deepti and Tripti Tanvi, "Effective Analysis and Diagnosis of Liver Disorder", International Conference on Communication and Signal Processing, pp. 86-90, April 3–5, 2018. https://ieeexplore.ieee.org/document/8524347

10. L. A. Auxilia, "Accuracy Prediction Using Machine Learning Techniques for Indian Patient Liver Disease", 2018 2nd International Conference on Trends in

Electronics and Informatics (ICOEI), pp. 45-50, 2018. https://ieeexplore.ieee.org/document/8553682

11. Menard, G., & Torielle, N. (2014). Training and assessing classification rules with imbalanced data. Data Mining and Knowledge Discovery, 28(1), 92–122. https://ieeexplore.org/10.1007/s10618-012-0295-5.

12. Nahar, N., & Ara, F. (2018). Liver disease prediction by using different decision tree techniques. International Journal of Data Mining & Knowledge Management Process, 8(2), 01–09. https://ieeexplore.org/10.5121/ijdkp.2018.8201.

13. Ortega, J. H. J. C., Layman, A. C., Natividad, L. R. Q., Ban tug, E. T., Resurrection, M. R., & Manalo, L. O. (2020). Analysis of performance of classification algorithms in mushroom poisonous detection using confusion matrix analysis. International Journal, 9(1.3). https://doi.org/10.30534/ijatcse/2020/7191.32020.

14. Ossawa, F. Y., Akinola, J. E. T., Ayodele, O., Himyarite, J. O., Alazani, O., & Ajimobi, J. (2017). Supervised machine learning algorithms: Classification and comparison. International Journal of Computer Trends and Technology (IJCTT), 48(3), 128–138. https://doi.org/10.14445/22312803/IJCTT-V48P126.

FIRST AUTHORS:

D. NAGESWARA NAIK pursuing his B.Tech in Computer Science And Engineering in Universal College Of Engineering And Technology.

K. SIVA SANKARA VARA PRASAD pursuing his B.Tech in Computer Science And Engineering in Universal College Of Engineering And Technology.

S. SASI KUMAR pursuing his B.Tech in Computer Science And Engineering in Universal College Of Engineering And Technology.

SK SHAHENA pursuing her B.Tech in Computer Science And Engineering in Universal College Of Engineering And Technology.

S. AARTHI REDDY pursuing her B.Tech in Computer Science And Engineering in Universal College Of Engineering And Technology.

SECOND AUTHOR:

Dr. M. VEERA KUMARI,M.Tech, Ph.D and B.Tech received her degree's in computer science and engineering. She is currently working as Associate Professor and HOD of CSE in Universal College Of Engineering And Technology.