International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 21 No. 4 (2025)



ijerst.editor@gmail.com

editor@ijerst.com

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
Vol. 21, No. 4, 2025
© 2025 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

Research Paper

Liver Disease Prediction using Machine learning and Deep Learning

[1]Tiragamalla Akhila,[2]Mr.Bblv Prasad

[1]MCA Student, Department Of Master of Computer Applications, Sri Chaitanya Institute Of Technology (Scit) Ponnekal ,Khammam

[1]Email : akhilatiragamalla@gmail.com

[2] Professor, Department Of Master of Computer Applications, Sri Chaitanya Institute Of Technology (Scit) Ponnekal ,Khammam

[2]Email : masterbhanuprasad@gmail.com

ABSTRACT

Recently liver diseases are becoming most lethal disorder in a number of countries. The count of patients with liver disorder has been going up because of alcohol intake, breathing of harmful gases, and consumption of food which is spoiled and drugs. Liver patient data sets are being studied for the purpose of developing classification models to predict liver disorder. This data set was used to implement prediction and classification algorithms which in turn reduces the workload on doctors.The data set used in this paper is Liver Patient taken from UCI Repository (i.e. Supervised Learning). There is a plenty of data on patients undergoing medical examination at hospitals and these data has been extracted on liver patients whose information can be further used for future improvement of their conditions. In other words, historical and classified input of patients and output data is fed into various algorithms or classifiers for predicting the future data of patients. In this work, we proposed apply machine learning and deep learning algorithms to check the entire patient's liver disorder.The algorithms used here for predicting liver patients are Decision Tree, KNNeighbor and Artifical Neural Network. Based on the analysis and result calculations, it was found that these algorithm has obtained good accuracy after feature selection. A Decision Tree boasts remarkable accuracy at 99.96%, leading the pack in precise liver disease predictions. Meanwhile, the KNearestNeighbour model follows closely at 97.42%, with Artificial Neural Networks achieving 71.55% accuracy, offering diverse options for predictive analysis.

Keywords: Liver Disease Prediction, Machine Learning, Deep Learning, Decision Tree, K-Nearest Neighbor (KNN), Artificial Neural Network (ANN), Classification, Feature Selection, Medical Diagnosis, Healthcare Analytics, Liver Patient Dataset, UCI Machine Learning Repository, Predictive Analysis, Disease Detection, Clinical Decision Support System.

I. INTRODUCTION

The liver is the most imperative structure in a human build. Insulin is broken down by the liver. The liver breaks bilirubin with glucuronidation, which further helps its defecation into bile [1]. It is also accountable for the breaking down and excretion of many unwanted products. It shows a noteworthy role in altering toxic materials. It shows a noteworthy role in collapsing medicinal products.

It's named Drug metabolism. The weight would be 1.3 kg. The liver consists of 2 immense portions namely the privileged portion, and the left estimate. The gallbladder is located below the liver, near the pancreas. The Liver along with these organs helps to consume and give nutrition. Its job is to help the flow of the wounding materials in the stream of blood from the stomach, before passing it to whatsoever is left of the body. Liver sicknesses are triggered when the working of the liver is affected or any injury has happened to it [2]. The development of liver disorders [3] is complicated and varied in character, influenced by a number of variables that determine disease susceptibility. Sex, ethnicity, genetics, environmental exposures (viruses, alcohol, nutrition, and chemicals), body mass index (BMI), and coexisting diseases like diabetes are among them. A high mortality rate is associated with liver problems, which are life-threatening diseases. The usual urine and blood tests are the first step in the prognosis of liver disorders. A LFT (liver functions test) is recommended for the patient based on the symptoms seen [4]. Liver disease is a significant health issue affecting millions of people globally. Early detection and accurate classification of liver diseases can lead to better patient outcomes and reduce the burden on the healthcare system. One-third of adults and an increasing proportion of youngsters in affluent nations suffer from non-alcoholic fatty liver disease (NAFLD) [5], a growing health issue. The abnormal buildup of triglycerides in the liver, which in some people causes an inflammatory reaction that can lead to cirrhosis and liver cancer, is the first sign of the condition. While there is a significant correlation between obesity, insulin resistance, and non-alcoholic fatty liver disease (NAFLD), the pathophysiology of NAFLD remains poorly understood, and treatment options are limited. However, machine learning techniques have demonstrated encouraging results in predicting and categorizing liver diseases based on patient data. By utilizing sophisticated algorithms to analyze and learn from large datasets, these techniques can identify patterns and anticipate outcomes. The employment of machine learning techniques in liver disease prediction and classification is a dynamic area of research, with continual advancements being made to enhance accuracy and decrease healthcare costs.

II. LITERATURE SURVEY

Deep learning in liver biopsies using convolutional neural networks

AUTHORS: R. Forlano, P. Manousou, and N. Giannakeas,

ABSTRACT: Nonalcoholic fatty liver disease (NAFLD) presents a wide range of pathological conditions, varying from nonalcoholic steatohepatitis (NASH) to cirrhosis and hepatocellular carcinoma (HCC). Their prevalence is characterized by increased fat accumulation and hepatocellular ballooning. They have become a cause of concern among physicians and engineers, as significant implications tend to occur regarding their accurate diagnosis and treatment. Although magnetic resonance, ultrasonography and other noninvasive methods can reveal the presence of NAFLD, image quantitative interpretation through histology has become the gold standard in clinical examinations. The proposed work introduces a fully automated diagnostic tool, taking into account the high discrimination capability of histological findings in liver biopsy images. The developed methodology is based on deep supervised learning and image analysis techniques, with the determination of an efficient convolutional neural network (CNN) architecture, performing eventually a classification accuracy of 95%.

2.2 Accuracy prediction using machine learning techniques for indian patient liver disease

AUTHORS: L. A. Auxilia,

ABSTRACT: The utilization of medicinal datasets has pulled in the consideration of specialists around the world. Machine Learning methods have been broadly utilized as a part of creating choice emotionally supportive networks for ailments forecast through an arrangement of therapeutic datasets. Grouping systems have been broadly utilized as a part of the restorative field for exact order than an individual classifier. Liver malady (additionally called hepatic infection) is a sort of harm to or illness of the liver. There are in excess of a hundred various types of liver ailment. In this task, I have taken the datasets of general Indian liver

ailment patient's records to help basic leadership. Indian Liver Patient's datasets demonstrate that proposed technique amazingly enhances the illnesses expectation precision.

2.3 Applying machine learning in liver disease and transplantation: a comprehensive review

AUTHORS: B. Wang, A. Goldenberg, and M. Bhat

ABSTRACT: Machine learning (ML) utilizes artificial intelligence to generate predictive models efficiently and more effectively than conventional methods through detection of hidden patterns within large data sets. With this in mind, there are several areas within hepatology where these methods can be applied. In this review, we examine the literature pertaining to machine learning in hepatology and liver transplant medicine. We provide an overview of the strengths and limitations of ML tools and their potential applications to both clinical and molecular data in hepatology. ML has been applied to various types of data in liver disease research, including clinical, demographic, molecular, radiological, and pathological data. We anticipate that use of ML tools to generate predictive algorithms will change the face of clinical practice in hepatology and transplantation. This review will provide readers with the opportunity to learn about the ML tools available and potential applications to questions of interest in hepatology.

2.4 Diagnosis of liver diseases using machine learning

AUTHORS: S. Sontakke, J. Lohokare, and R. Dani,

ABSTRACT: Liver Diseases account for over 2.4% of Indian deaths per annum. [14] Liver disease is also difficult to diagnose in the early stages owing to subtle symptoms. Often the symptoms become apparent when it is too late. [1] This paper aims to improve diagnosis of liver diseases by exploring 2 methods of identification-patient parameters and genome expression. The paper also discusses the computational algorithms that can be used in the aforementioned methodology and lists demerits. It proposes methods to improve the efficiency of these algorithms.

2.5 Human fatty liver disease: old questions and new insights

AUTHORS: J. C. Cohen, J. D. Horton, and H. H. Hobbs,

ABSTRACT: Nonalcoholic fatty liver disease (NAFLD) is a burgeoning health problem that affects one-third of adults and an increasing number of children in developed countries. The disease begins with the aberrant accumulation of triglyceride in the liver, which in some individuals elicits an inflammatory response that can progress to cirrhosis and liver cancer. Although NAFLD is strongly associated with obesity and insulin resistance, its pathogenesis remains poorly understood, and therapeutic options are limited. Here, we discuss recent mechanistic insights into NAFLD, focusing primarily on those that have emerged from human genetic and metabolic studies.

III. EXISTING SYSTEM

In the existing system, different classifiers were implemented on liver patient diseases dataset to predict liver diseases based on developed software. Dataset was processed and implemented on WEKA tool using feature selection techniques with 10-fold cross validation testing option. The results of the proposed work were compared using feature selection and without using feature selection techniques after the implementation of different classifiers in terms of execution time and accuracy. During the research work the result of other parameters like kappa statistic, correctly classified instances, and mean absolute error were also compared on liver patient diseases dataset.

IV. PROPOSED SYSTEM

In the proposed system, we have to import the liver patient dataset (.csv). Then the dataset is pre-processed and the anomalies and full-up empty cells in the dataset are removed, so that we can further improve the effective liver disease prediction. Then we construct a Confusion matrix for accomplishing an enhanced lucidity of the no of correct/incorrect predictions. Formerly, several classification and prediction procedures and if possible, combinations of different algorithms are implemented and check the accuracy.

V. SYSTEM ARCHITECTURE

The proposed Liver Disease Prediction System follows a structured machine learning architecture

that transforms raw patient medical records into accurate disease predictions through a sequence of data processing and classification stages. The architecture consists of seven major components: Dataset, Data Preprocessing, Feature Selection, Machine Learning Model, Classifier, Data Classification, and Result. Each component performs a specific task to improve prediction accuracy and support early diagnosis of liver disease.

The process begins with the Dataset, which contains historical medical records of liver patients collected from the UCI Machine Learning Repository. The dataset includes various clinical attributes such as age, gender, total bilirubin, direct bilirubin, alkaline phosphatase, alanine aminotransferase, aspartate aminotransferase, total proteins, albumin, albumin and globulin ratio, and the target class indicating whether the patient has liver disease. This dataset serves as the primary input for training and testing the prediction models.

The collected data is then passed to the Data Preprocessing stage, where missing values, duplicate records, and inconsistent data are identified and removed. Categorical values are converted into numerical form, and numerical features are normalized or standardized to ensure uniformity across all attributes. Data preprocessing improves the quality of the dataset, reduces noise, and prepares the information for efficient machine learning analysis.

After preprocessing, the system performs Feature Selection, which identifies the most relevant medical attributes that significantly influence liver disease prediction. Irrelevant and redundant features are eliminated to reduce computational complexity and improve the efficiency of the learning algorithms. Selecting the most informative features helps increase classification accuracy while minimizing overfitting and training time.

The selected features are then provided to the Machine Learning Model, where different algorithms such as Decision Tree, K-Nearest Neighbor (KNN), and Artificial Neural Network (ANN) are trained using historical patient data. These algorithms learn the relationships between clinical features and liver disease conditions, enabling the system to recognize hidden patterns associated with liver disorders. The trained models are optimized to provide reliable predictions for new patient records.

The trained models are utilized by the Classifier, which receives unseen or test patient data samples and determines whether the patient is likely to have liver disease. The classifier compares the characteristics of new patient records with the knowledge learned during training and assigns the appropriate disease class. This classification process forms the core decision-making component of the system.

Next, the Data Classification module organizes the classifier's outputs into meaningful prediction categories such as Liver Disease or No Liver Disease. It evaluates the prediction results using performance metrics such as accuracy, precision, recall, and F1-score to measure the effectiveness of each machine learning algorithm. This stage helps identify the best-performing model for liver disease prediction.

Finally, the Result module presents the prediction outcome to healthcare professionals or users. The system displays whether the patient is affected by liver disease based on the selected machine learning model. According to the experimental results, the Decision Tree algorithm achieved the highest prediction accuracy of 99.96%, followed by K-Nearest Neighbor (97.42%) and Artificial Neural Network (71.55%). These results demonstrate the effectiveness of machine learning techniques in supporting early diagnosis and clinical decision-making.

Overall, the proposed system architecture provides an intelligent, efficient, and reliable framework for liver disease prediction. By integrating data preprocessing, feature selection, machine learning algorithms, and automated classification, the system assists healthcare professionals in detecting liver disorders at an early stage, reducing diagnostic workload, improving prediction accuracy, and enabling timely medical intervention.

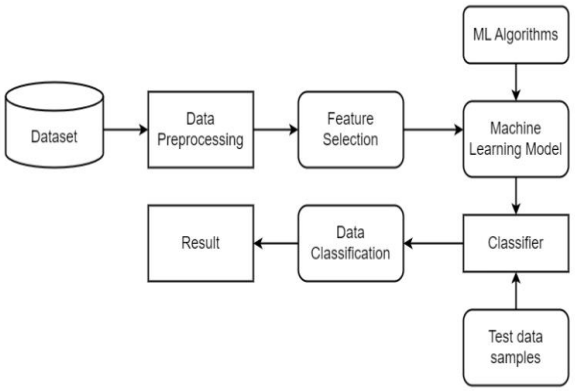


Fig 5.1: System Architecture

VI. IMPLEMENTATION

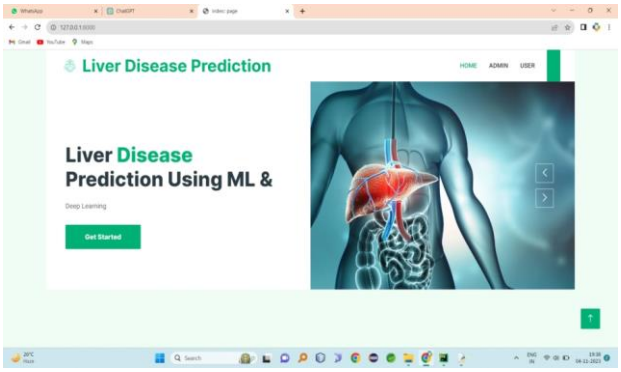


Fig 6.1 Index page

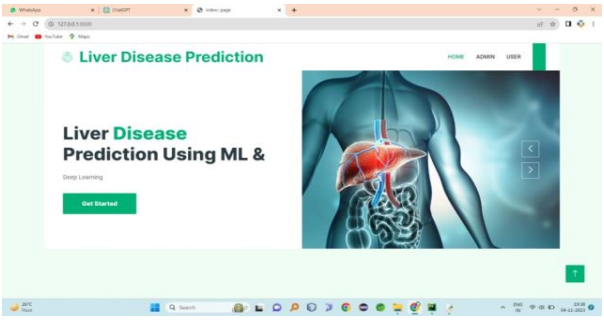


Fig 6.2 : Admin home

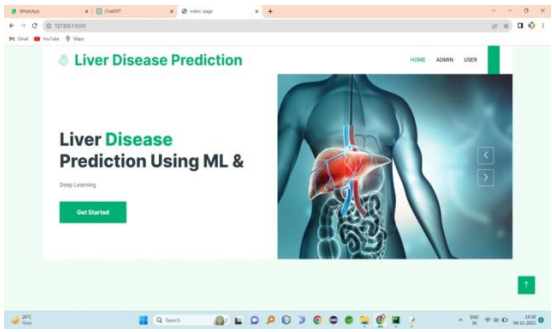


Fig 6.3: Upload dataset

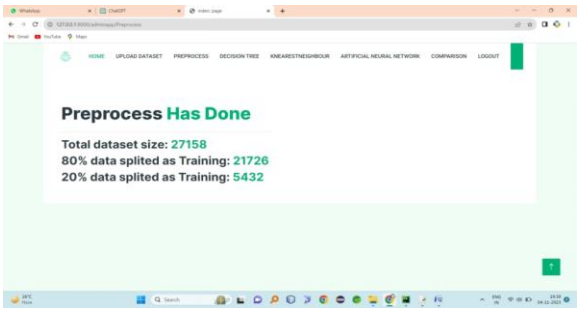


Fig 6.4: Preprocessing

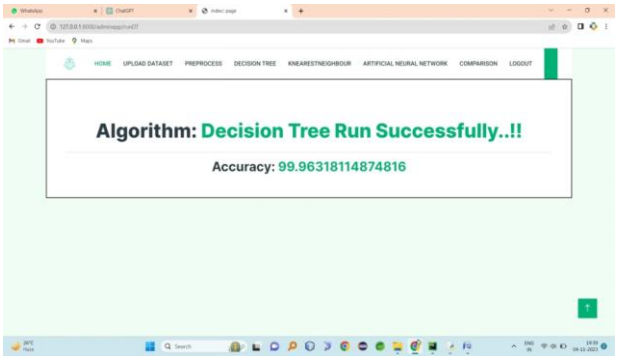


Fig 6.5: Model training

VII. CONCLUSION

The Prediction of liver illness in patients has been examined and analysed in this paper. By using various techniques the data has been cleaned by imputation of missing values with median, then dummy encoding is applied followed by outlier eliminated to improve the performance. In this research paper, various classification algorithm have been applied such as Decision Tree, KNNeighbor and Artifical Neural Network Based on algorithm applied, it is observed that models Decision Tree, KNNeighbor algorithm gives better accuracy than the other classification algorithm. Therefore

concluding that Decision Tree is appropriate for predicting liver disease. When a training data set is available, our proposed classification schemes can significantly enhance classification performance. Then, using a machine learning classifier, good and bad values are classified. Thus, the outputs of the proposed classification model show accuracy in predicting the result.

VIII. FUTURE SCOPE

The extent of our work is that we will apply deep learning techniques to predict liver disease. Some of the future directions are improve the accuracy of liver disease prediction and classification models is to include more diverse data sources, improving liver disease prediction and classification is to combine multiple machine learning techniques, machine learning models can be trained to predict the likelihood of liver disease in individuals based on their unique characteristics. Another important direction in liver disease prediction and classification using machine learning is to develop models that are explainable. This means that the models should provide clear and interpretable insights into the factors that contribute to liver disease. Explainable models can help healthcare professionals to make better decisions and provide better care for patients.

IX. REFERENCES

[1] M. Sameer and B. Gupta, "Beta Band as a Biomarker for Classification between Interictal and Ictal States of Epileptical Patients," in 2020 7th International Conference on Signal Processing and Integrated Networks (SPIN), 2020, pp. 567–570, doi: 10.1109/SPIN48934.2020.9071343.

[2] S. K. B. Sangeetha, N. Afreen, and G. Ahmad, "A Combined Image Segmentation and Classification Approach for COVID-19 Infected Lungs," J. homepage http//iieta. org/journals/rces, vol. 8, no. 3, pp. 71–76, 2021.

[3] M. Sameer, A. K. Gupta, C. Chakraborty, and B. Gupta, "Epileptical Seizure Detection: Performance analysis of gamma band in EEG signal Using Short-Time Fourier Transform," in 2019 22nd International Symposium on Wireless Personal Multimedia Communications (WPMC), 2019, pp. 1–6, doi: 10.1109/WPMC48795.2019.9096119.

[4] A. Mahajan, K. Somaraj, and M. Sameer, "Adopting Artificial Intelligence Powered ConvNet To Detect Epileptic Seizures," in 2020 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES), 2021, pp. 427–432, doi: 10.1109/IECBES48179.2021.9398832.

[5] N. Nasir, N. Afreen, R. Patel, S. Kaur, and M. Sameer, "A Transfer Learning Approach for Diabetic Retinopathy and Diabetic Macular Edema Severity Grading," Rev. d'Intelligence Artif., vol. 35, pp. 497–502, Dec. 2021, doi: 10.18280/ria.350608.

[6] M. Sameer and B. Gupta, "ROC Analysis of EEG Subbands for Epileptic Seizure Detection using Naive Bayes Classifier," J. Mob. Multimed., pp. 299–310, 2021.

[7] M. Sameer and B. Gupta, "Time–Frequency Statistical Features of Delta Band for Detection of Epileptic Seizures," Wirel. Pers. Commun., 2021, doi: 10.1007/s11277-021-08909-y.

[8] S. M. Beeraka, A. Kumar, M. Sameer, S. Ghosh, and B. Gupta, "Accuracy Enhancement of Epileptic Seizure Detection: A Deep Learning Approach with Hardware Realization of STFT," Circuits, Syst. Signal Process., 2021, doi: 10.1007/s00034-021 01789-4.

[9] S. Gupta, M. Sameer, and N. Mohan, "Detection of Epileptic Seizures using Convolutional Neural Network," in 2021 International Conference on Emerging Smart Computing and Informatics (ESCI), 2021, pp. 786–790, doi: 10.1109/ESCI50559.2021.9396983.

[10] P. Porwal et al., "Indian Diabetic Retinopathy Image Dataset (IDRiD): A Database for Diabetic Retinopathy Screening Research," Data , vol. 3, no. 3. 2018, doi: 10.3390/data3030025.

[11] M. Sameer and P. Agarwal, "Coplanar waveguide microwave sensor for label-free real-time glucose detection," Radioengineering, vol. 28, no. 2, p. 491, 2019.

[12] M. Sameer and B. Gupta, "Detection of epileptical seizures based on alpha band statistical features," Wirel. Pers. Commun., vol. 115, no. 2, pp. 909–925, 2020, doi: 10.1007/s11277-020-07542-5.

[13] M. Sameer, A. K. Gupta, C. Chakraborty, and B. Gupta, "ROC Analysis for detection of Epileptical Seizures using Haralick features of Gamma band," in 2020 National Conference on Communications (NCC), 2020, pp. 1–5, doi: Computer Networks 10.1109/NCC48643.2020.9056027.

[14] N. Afreen, R. Patel, M. Ahmed, and M. Sameer, "A Novel Machine Learning Approach Using Boosting Algorithm for Liver Disease Classification," in 2021 5th International Conference on Information Systems and Computer Networks (ISCON), 2021, pp. 1–5.

[15] N. Jiwani, K. Gupta, and P. Whig, "Novel HealthCare Framework for Cardiac Arrest With the Application of AI Using ANN," in 2021 5th International Conference on Information Systems and (ISCON), 2021, pp. 1–5, doi: 10.1109/ISCON52037.2021.9702493.