ISSN: 2581-3404 (Online) IF: 5.68 (SJIF)
IJIRTM, Volume-8, Issue-3, June-2024.

Liver Patients Disease Detection Using Machine Learning Techniques: A Review

Priyanka Thakur[1], Dr. Pinaki Ghosh[2]
M. Tech. Scholar[1], Professor[2]
Department of Artificial Intelligence[1,2]
Sanjeev Agrawal Global Educational University, Bhopal (M.P.) [1,2]

Abstract. *Liver diseases have become a significant health problem worldwide. As a result of its limited symptoms, it is extremely difficult to detect liver disease until the very last stage. An early diagnosis of liver problems will increase patient's survival rate. The introduction of Machine Learning Tools in healthcare has opened up a wide spectrum of applications. Early detection by utilizing supervised learning techniques on a diverse data set can help in reducing mortality. This paper reviews some classification algorithms like Logistic Regression, Support Vector Machines (SVM), K Nearest Neighbor (KNN), Random Forest, Decision Tree and Extra tree classifier for comparing their performance based on the liver patient data.*

Keywords: **-** Disease diagnosis, Machine learning, Classification, Accuracy.

Introduction

Liver acts as one of the most vital internal organs of human body, which functions to work properly even if partly damaged. This causes early detection of issues in the liver difficult. A total of 264,193 deaths as a result of liver disease were reported in India in 2018, according to the latest World Health Organization data. There are about 23.00 deaths per 100,000 people [1]. Thus, developing tools that can aid liver experts such as endocrinologists in making accurate decisions on patients and also with the help of Automatic classification tools for liver diseases (probably mobile enabled or web enabled), one can reduce the patient queue at hospitals as well as provide effective health care in a timely manner.

With the advent of new developments in the field of medicine, lots of data has been gathered. Data mining applications have been effectively USED in different areas. In healthcare huge amount of data is being generated which is processed and analyzed for knowledge extraction. Mining algorithms predict the disease of patients using suitable learning strategy. Diseases like chronic kidney disease (CKD), hepatitis, cancer disease and diabetes have become a worldwide health issue and therefore prediction of such type of diseases is the concerned area for researchers [2,3]. Our work mainly focuses on analyzing classification algorithms like Support Vector Machine, KNN, Decision Tree and Random Forest. There are two main categories of data mining known as supervised learning and unsupervised learning. These techniques mentioned above are used in medical field accordingly to predict diseases and for making decision for treatment of patients. Classification is a supervised technique in which objects are assigned in a collection to target classes.

Decision trees, SVM, NB etc. are approaches of classification. Algorithm like K-Nearest neighbour (KNN) is a simple yet highly efficient data mining algorithm used for classification and pattern recognition. It used a hybrid algorithm that combined Genetic Algorithm (GA) and KNN algorithm to improve the accuracy of classification [5].

In clustering similar types of objects are categorized in the same group. K-means, K- medoids, agglomerative, divisive, DBSCAN etc. are some of the techniques of clustering. Association is the possibility of occurrence of objects in a set.The algorithms used are to predict the liver disorder disease by comparing the accuracy of each algorithm. Finally, clustering divides or clusters the patient symptoms into small partitions based on its similarity. It groups the organized of comparable records into clusters and the clustering analysis will give outcomes directly.

Literature Review

Different researchers have worked on liver disease diagnosis previously and found accuracy of different machine learning algorithms using different tools. In these studies, the decisions made by the prediction systems and input data from patients impacted liver disease diagnoses.

Ketan Gupta et al. [1] Machine Learning is a process which is used to discover patterns in huge data/ large data set to enable decision, thereby allowing machines to go through a learning process (i.e. supervised, unsupervised and semi-supervised or reinforced). 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. The algorithms used here for predicting liver patients are Logistic regression, Decision Tree, Random Forest, KNNeighbor, Gradient Boosting, Extreme Gradient Boosting, LightGB. Based on the analysis and result calculations, it was found that this algorithm has obtained good accuracy after feature selection.

Jingwei Wei, et al. [3] The goal of this work is to evaluate the performance of different Machine Learning algorithms in order to reduce the high cost of chronic liver disease diagnosis by prediction. In this work, we used six algorithms Logistic Regression, K Nearest Neighbors, Decision Tree, Support Vector Machine, Naïve Bayes, and Random Forest. The performance of different classification techniques was evaluated on different measurement techniques such as accuracy, precision, recall, f-1 score, and specificity. We found the accuracy 75%, 74%, 69%, 64%, 62% and 53% for LR, RF, DT, SVM, KNN and NB. The analysis result shown the LR achieved the highest accuracy. Moreover, our present study mainly focused on the use of clinical data for liver disease prediction and explore different ways of representing such data through our analysis.

Mandakini Priyadarshani Behera, et al. [4] Liver diseases, a wide spectrum of pathologies from inflammation to neoplasm, have become an increasingly significant health problem worldwide. Noninvasive imaging plays a critical role in the clinical workflow of liver diseases, but conventional imaging assessment

may provide limited information. Here, they review the methodological process in liver disease radiomics studies in a stepwise fashion from data acquisition and curation, region of interest segmentation, liver-specific feature extraction, to task-oriented modeling. Furthermore, the applications of radiomics in liver diseases are outlined in aspects of diagnosis and staging, evaluation of liver tumour biological behaviors, and prognosis according to different disease type. Finally, we discuss the current limitations of radiomics in liver disease studies and explore its future opportunities.

Se-Yeol Rhyou, et al. [5] In this study for heart and liver data classification, a hybrid model is created by combining support vector machine (SVM) approach and modified particle swarm optimization model. The data sets are collected from UCI machine learning repository. The results are calculated based on classification accuracy, error, correctness, recall as well as F1 score. The results obtained is compared with SVM, hybrid particle swarm optimization support vector machine algorithm (PSOSVM), hybrid Crazy particle swarm optimization support vector machine algorithm (CPSOSVM).

D. Devikanniga, et al. [7] This research work uses different ensemble methods to investigate the early detection of liver disease. The selected dataset for this analysis is made up of attributes such as total bilirubin, direct bilirubin, age, sex, total protein, albumin, and globulin ratio. This research mainly aims at measuring and comparing the efficiency of different ensemble methods. AdaBoost, LogitBoost, BeggRep, BeggJ48 and Random Forest are the ensemble method used in this research. The study shows that LogitBoost is the most accurate model than other ensemble approaches.

Jaime Lynn Speiser, et al. [8] In this research, author propose a fully automated liver steatos is prediction model using three deep learning neural networks. As a result, liver steatosis can be automatically detected with high accuracy and precision. First, transfer learning is used for semantically segmenting the liver and kidney (L-K) on parasagittal US images, and then cropping the L-K area from the original US images. The second neural network also involves semantic segmentation by checking the presence of a ring that is typically located around the kidney and cropping of the L-K area from the original US images. These cropped L-K areas are inputted to the final neural network, SteatosisNet, in order to grade the severity of fatty liver disease. The experimental results demonstrate that the proposed model can predict fatty liver disease with the sensitivity of 99.78%, specificity of 100%, PPV of 100%, NPV of 99.83%, and diagnostic accuracy of 99.91%, which is comparable to the common results annotated by medical experts.

Nazim Razali, et al. [9] In the present review, author analyzed the diagnostic performance of ultrasonography (US) in the non-invasive evaluation of NAFLD and NAFLD-related HCC, as well as possibilities of optimizing US diagnosis with the help of artificial intelligence (AI) assistance. To date, US is the first-line examination recommended in the screening of patients with clinical suspicion of NAFLD, as it is readily available and leads to a better disease-specific surveillance. However, the conventional US presents limitations that significantly hamper its applicability in quantifying NAFLD and accurately characterizing a given focal liver lesion (FLL). Ultrasound contrast agents (UCAs) are an essential add-on to the conventional B-mode US and to the Doppler US that further empower this method, allowing the

evaluation of the enhancement properties and the vascular architecture of FLLs, in comparison to the background parenchyma. The current paper also explores the new universe of AI and the various implications of deep learning algorithms in the evaluation of NAFLD and NAFLD-related HCC through US methods, concluding that it could potentially be a game changer for patient care.

Binish Khan, et al. [12] The objective of this study is to propose a rule-based classification model with basic decision-making techniques to predict various types of heart disease. To get better results the experiment was done using a different data mining algorithm compared with previous liver disease predictions. All experiments have been implemented at Azure Machine Learning tool. This paper is about to study the prediction of liver disease to produce better performance accuracy by comparing various mining data classification algorithms.

Md. Shafiul Azam, et al. [13] The aim of this research is to develop a mechanism to help medical practitioners predict and diagnose liver disease. Several systems have been proposed to help medical experts by diminishing error and increasing accuracy in diagnosing and predicting diseases. Among many existing methods, a few have considered the class imbalance issues of liver disorder datasets. As all the samples of liver disorder datasets are not useful, they do not contribute to learning about classifiers. A few samples might be redundant, which can increase the computational cost and affect the performance of the classifier. In this paper, a model has been proposed that combines noise filter, fuzzy sets, and boosting techniques (NFFBTs) for liver disease prediction. Firstly, the noise filter (NF) eliminates the outliers from the minority class and removes the outlier and redundant pair from the majority class. Secondly, the fuzzy set concept is applied to handle uncertainty in datasets. Thirdly, the AdaBoost boosting algorithm is trained with several learners viz, random forest (RF),support vector machine (SVM), logistic regression (LR), and naive Bayes(NB). The proposed NFFBT prediction system was applied to two datasets(i.e., ILPD and MPRLPD) and found that AdaBoost with RF yielded 90.65%and 98.95% accuracy and F1 scores of 92.09% and 99.24% over ILPD and MPRLPD datasets, respectively.

Problem Statement

It is very difficult to identify in early stages of liver disease even liver tissue has damaged moderately, in these case many medical expert system difficult to identify the disease. This leads to fail in treatment and medication. In order to avoid this early prediction is crucial to give proper treatment and save life of patient. As per the existing system of medical expert system for diagnosis of liver disease has been useful to the society, moreover easy detection and prediction of the disease can be easy done with the use of the expert system. With the repeated improving in Artificial intelligence different types of machine learning algorithm has been developed this will help in improving the quality and accuracy of the detection or prediction of the liver disease. Data preprocessing is one of the most critical steps in data mining process. The sequences of steps identified in extracting knowledge from data are data cleaning, data integration, data transformation, and data reduction.

This study will use classification and regression as data mining tasks. Classification is used for predicting responses that can have just a few known values based on column in the dataset. Regression can predict one or more continuous variables. In classification, models are evaluated using Bayes point machine and neural

network as algorithms, while regression using linear regression and Poisson regression. The liver is a very vital organ of the human body. Its failure can be fatal, and the only solution is a transplant within a given time. Various features such as total bilirubin, gender, age, SGPT, ALP, Albumin, etc., can be used for the early detection of liver illnesses in a person. Several research works discussed above make use of these features to detect liver disease. Though many machine learning classification-based algorithms are used in the literature, they have some drawbacks. In most existing works, we compare and studied different machine learning models are used, and have some limitation over the performance parameters.

Discussion

Machine learning is a subfield of Artificial Intelligence that allows a computer system to learn from the environment, through re-iterative processes and improve itself from experience. Machine learning algorithms organize the data, learn from it, gather insights and make predictions based on the information it analyzed without the need for additional explicit programming. Training a model with data and after that using the model to predict any new data is the concern of machine learning. Machine learning algorithms are widely composed of supervised, unsupervised, semi-supervised and reinforcement learning. The development of machine learning has proven to better describe data as a result of providing both engineering solutions and an important benchmark.

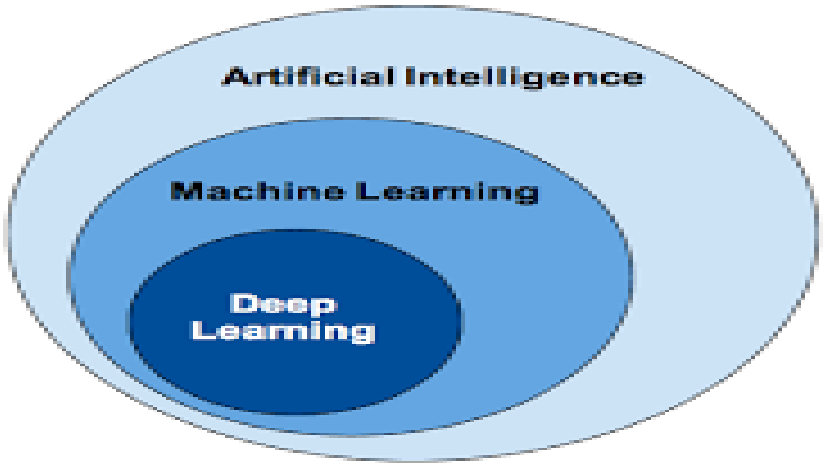

Fig. 1: Representation of artificial intelligence and its subfields.

There are various tools and technology for the prediction and classification re available machine learning is most popular tools among them. Machine learning is popular programming tools for the prediction and classification of large or huge amount of data, the classification is also a method of data mining which primary object is to grouping of the same data or the same pattern of the data in a single group. Machine learning is basically differing from the traditional programming language. As we know that the machine learning tolls provide the concept of pattern analysis or behavioral analysis for the large amount of data, basically it is categorized the data according to their properties or behavior such as their shape, size, used

places, their attributes etc. classification is also challenging task for the dynamic nature or regularly updated data there is classification are supervised classification, up-supervised classification, semi-supervised classification. Machine learning play big role in pattern recognition, the recognition of pattern faced the series of training process. The training process of classification technique generates the accuracy performance of classifier and method of pattern recognition.

Machine learning is popular programming tools for the prediction and classification of large or huge amount of data, the classification is also a method of data mining which primary object is to grouping of the same data or the same pattern of the data in a single group. Machine learning is basically differing from the traditional programming language. As we know that the machine learning tolls provide the concept of pattern analysis or behavioral analysis for the large amount of data, basically it is categorized the data according to their properties or behavior such as their shape, size, used places, their attributes etc. classification is also challenging task for the dynamic nature or regularly updated data there is some classification are supervised classification, up-supervised classification, semi-supervised classification.

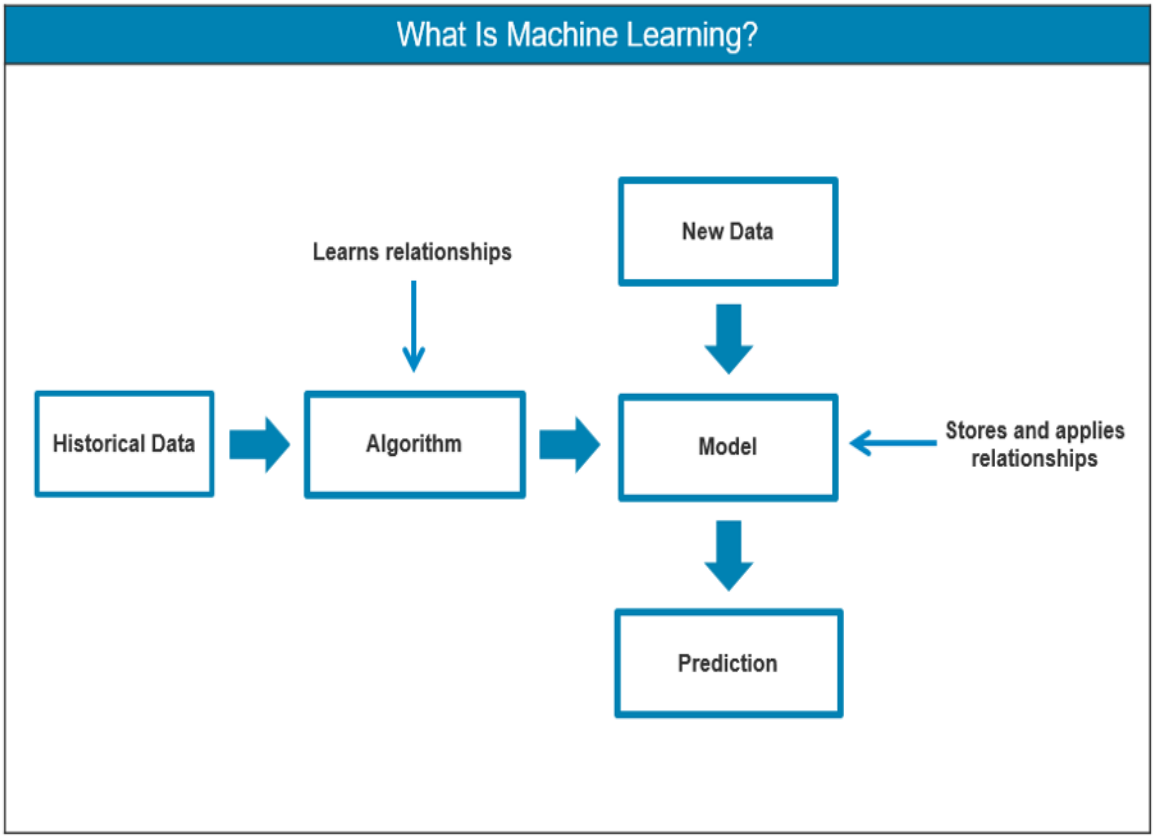

Fig. 2: Machine learning algorithms implementation model.

In this work, some classification algorithms like Logistic Regression, Support Vector Machines (SVM), K Nearest Neighbor (KNN), Random forest, decision tree, and Extra tree classifier have been considered for comparing their performance based on the liver patient data.

Conclusion

Liver disease has been increasing annually in people across the globe. This is mainly due to lifestyle changes. The earlier machine learning models showed limitations in terms of dataset, feature extraction tools and lack of additional parameters. This work presents an approach that compared various techniques which can help in combining various methods and developing a new model or hybrid model for health services. More than one dataset may be used for better approach and comparison. Classification rules and disease identifying techniques may also be generated by using different efficient algorithms. More than one database for comparative analysis, better approach and improved accuracy may also be used. Here we review different machine leaning techniques for liver patients to diagnosis at early stage, in future will propose the artificial intelligence based model to predict the more accurate value for using different performance parameters value.

References

[1] Ketan Gupta, Nasmin Jawani, Neda Afreen, “Liver Disease Prediction using Machine learning Classification Techniques”, IEEE, 2022.

[2] A.K.M Sazzadur Rahman, F. M. Javed Mehedi Shamrat, Zarrin Tasnim, Joy Roy, Syed Akhter Hossain, “A Comparative Study On Liver Disease Prediction Using Supervised Machine Learning Algorithms”, International Journal of Scientific & Technology Research Volume 8, Issue 11, November 2019.

[3] Jingwei Wei,Hanyu Jiang,Dongsheng Gu,Meng Niu,Fangfang Fu, Yuqi Han, Jie Tian, “Radiomics in liver diseases: Current progress and futureopportunities”, Liver International, 2020.

[4] Mandakini Priyadarshani Behera, Archana Sarangi, Debahuti Mishra, Shubhendu Kumar Sarangi, “A Hybrid Machine Learning Algorithm for Heart and Liver Disease Prediction using Modified Particle Swarm Optimization”, International Conference on Machine Learning and Data Engineering, 2023.

[5] Se-Yeol Rhyou, Jae-Chern Yoo, “Cascaded Deep Learning Neural Network for Automated LiverSteatosis Diagnosis Using Ultrasound Images”, Journal of Sensors, 2021.

[6] Monica Lupsor-Platon , Teodora Serban, Alexandra Iulia Silion, George Razvan Tirpe, Alexandru Tirpe and Mira Florea, “Performance of Ultrasound Techniques and the Potential of Artificial Intelligence in the Evaluation of HepatocellularCarcinoma and Non-Alcoholic Fatty Liver Disease”, Journal of Cancer Treatment, MDPI.

[7] D.Devikanniga, Arulmurugan Ramu and Ananda kumar Haldorai, " Efficient Diagnosis of Liver Disease using Support Vector Machine Optimized with Crows Search Algorithm", EAI Endorsed Transactions on Energy Web, 2020.

[8] Jaime Lynn Speiser, Constantine J. Karvellas, Bethany J. Wolf, Dongjun Chung, "Predicting daily outcomes in acetaminophen-induced acute liver failure patients with machine learning techniques", Computer Methods and Programs in Biomedicine, 2019.

[9] Nazim Razali, Aida Mustapha, Mohd Helmy Abd Wahab, SalamaA Mostafa1, Siti Khadijah Rostam, "A Data Mining Approach to Prediction of LiverDiseases", Journal of Physics, 2020.

[10] Pushpendra Kumar, and Ramjeevan Singh Thakur, "An Approach Using Fuzzy Sets and Boosting Techniques to Predict Liver Disease", Computers, Materials & Continua Tech Science PressDOI:10.32604/cmc.2021.016957.

[11] Mazhar Javed Awan, "Review on Effective Disease Prediction through Data Mining Techniques", International Journal on Electrical Engineering and Informatics · September 2021.

[12] Binish Khan, Piyush Kumar Shukla, "Strategic Analysis in Prediction of Liver Disease Using Different Classification Algorithms", International Journal of Computer Sciences and Engineering · July 2019.

[13]Md. Shafiul Azam, Aishe Rahman, S. M. Hasan Sazzad Iqbal, and Md. Toukir Ahmed, "Prediction of Liver Diseases by Using Few Machine Learning BasedApproaches", Australian Journal of Engineering and Innovative Technology, 2(5), 85-90, 2020.

[14] Abdul Quadir Md, Sanika Kulkarni, Christy Jackson Joshua, Tejas Vaichole, SenthilkumarMohanand Celestine Iwendi, "Enhanced Preprocessing Approach Using Ensemble Machine Learning Algorithms for Detecting Liver Disease", Journal of Biomedicines, 2023, 11, 581.

[15] A. Singh and B. Pandey, "Diagnosis of Liver Disease by Using least Squares Support Vector Machine Approach", ‖ Int. J. Healthc. Inf. Syst. Informatics, vol. 11, no. 2, pp. 62–75, Apr. 2016.

[16] D Rathore, P Mannepalli, "Recent Trends in Machine Learning for Health Care Sector ", International Journal of Innovative Research in Technology and Management, Vol-5, Issue-2, 2021.

[17] X. X. Geng, R. G. Huang, J. M. Lin, N. Jiang, and X. X. Yang, "Transient Elastography in Clinical Detection of Liver Cirrhosis : A Systematic Review and Meta - analysis", vol. 22, no. 4, pp. 294–303, 2016.

[18] T. Orczyk and P. Porwik, "Liver Fibrosis Diagnosis Support System Using Machine Learning Methods BT - Advanced Computing and Systems for Security", Volume 1, R. Chaki, A. Cortesi, K. Saeed, and N. Chaki, Eds. New Delhi: Springer India, 2016, pp. 111–121.

[19] L. Ozyilmaz and T. Yildirim, "Artificial Neural Networks for Diagnosis of Hepatitis Disease", pp. 586–589.

[20] D Rathore, P Mannepalli, "A Review of Machine Learning Techniques and Applications for Health Care", International Conference on Advances in Technology, Management & Education, 2021, IEEE proceeding, 978-1-7281-8586-6/21.

[21] I. Campus, "Diagnosis Of Liver Disease Induced By Hepatitis Virus Using Artificial Neural Networks", pp. 8–12, 2011.

[22] G. S. Uttreshwar, "Hepatitis B Diagnosis Using Logical Inference And Generalized Regression Neural Networks", Advance Computing Conference, no. March, pp. 6–7, 2009.