ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/386321889

A Study on Cirrhosis Prediction Based on Machine Learning Techniques

Article *in* Journal of Education for Pure Science · December 2024

DOI: 10.32792/jeps.v14i4.461

CITATIONS
0

READS
116

2 authors, including:

Maalim A. Aljabery
University of Basrah
21 PUBLICATIONS **29** CITATIONS

SEE PROFILE

All content following this page was uploaded by Maalim A. Aljabery on 04 December 2024.

The user has requested enhancement of the downloaded file.

Journal of Education for Pure Science- University of Thi-Qar
Vol 14., No. 4 (2024)
Website: jceps.utq.edu.iq Email: jceps@eps.utq.edu.iq

DOI: https://doi.org/10.32792/jeps.v14i4.461

A Study on Cirrhosis Prediction Based on Machine Learning Techniques

Duaa S.Ali, Maalim A. Aljabery

duaasabeeh1@gmail.com maalim.aljabery@uobasrah.edu.iq

Department of Computer Science, College of Computer Science and Information Technology , University of Basarh, Basrah, Iraq

Received 24/6/2024, Accepted 5/9/2024, Published 1/12/2024

This work is licensed under a Creative Commons Attribution 4.0 International License.

Abstract

Cirrhosis is an advanced stage of many chronic liver diseases, primarily caused by viral hepatitis. Early detection is crucial to prevent further liver tissue scarring and to prolong patient survival. AI-based computer-assisted diagnostics, utilizing Machine Learning (ML) and Deep Learning (DL) methods, offer significant advantages over conventional approaches by reducing time, effort, and risks, in addition to improving the accuracy and efficiency of diagnosis. This paper aims to review recent key studies on diagnosing various liver diseases, with a focus on cirrhosis, using ML and DL techniques. Additionally, it will cover publicly accessible liver disorder datasets and metrics for evaluating model performance and discuss existing research restrictions and future works for the automatic detection of cirrhosis.

Keywords: Artificial intelligence, Random Forest, liver disease, Machine Learning, Cirrhosis.

1 Introduction

Liver-related disorders are a leading cause of death worldwide and are increasingly prevalent. These conditions can progress to liver cirrhosis over time [1]. Cirrhosis, an advanced stage of chronic liver diseases, causes about 1.3 million deaths annually, ranking as the 14th leading cause of death globally [2]. It is characterized by the replacement of normal liver tissue with scar tissue, a dynamic process involving inflammation, hepatocyte injury, necrosis, fibrosis, and regeneration [3]. Various factors, such as hepatitis viruses (B and C), alcohol consumption, Non-Alcoholic Fatty Liver Disease (NAFLD), and autoimmune diseases, can lead to cirrhosis [1]. Clinically, cirrhosis is divided into two stages: compensated and

decompensated. Compensated cirrhosis is often asymptomatic, while decompensated cirrhosis is marked by symptoms such as jaundice, variceal bleeding, hepatic encephalopathy, ascites, and portal hypertension [4] [5].

Liver biopsy is the gold standard for diagnosing cirrhosis and conducting histological evaluations [6]. However, it has a recorded fatality rate of up to 1.6%, and it is invasive, expensive, and associated with risks such as bleeding, pain, and even death [6][7]. Additionally, sample mistakes (inadequate sample size, improper puncture placement, etc.) could lead to an underestimation of liver damage [8].

Due to these limitations, there is a need for alternative, non-invasive methods to identify the stages of liver cirrhosis. Recognized non-invasive diagnostic algorithms, such as (DL) and (ML) methods, offer promising alternatives.

1.1 Overview of Machine Learning (ML) and Deep Learning (DL)

ML is a branch of AI designed to enable computers to learn specific skills without human intervention [9]. It is extensively used in healthcare for diagnosing and treating illnesses because the complexity of medical data makes manual detection challenging; medical data, which includes patient records, is increasingly stored electronically as information technology advances. Based on clinical and biochemical data, various ML algorithms can be used to predict conditions such as cirrhosis and liver disorders in patients [10][11].

DL is a subcategory of ML, with several layers in its structure that are utilized to extract high-level features from the input. These layers convert the picture-based input data into an output that identifies the disease [12]. One of the primary advantages of utilizing DL methods is that they automatically extract features without the need for human interaction. In contrast, ML models require manually designed feature extraction procedures [13]. DL approaches are the main engine of artificial intelligence (AI) in hepatology [14]. ML algorithms are, in general, categorized into four primary types, which are listed below:

- **Supervised Learning (SL):** is a crucial ML technique in the medical field. It involves training an algorithm on labeled data to learn the associations between input and output data. This approach is commonly used in medicine to construct predictive models that can accurately diagnose and classify diseases. (e.g., present/absent illness or result) [15].
- **Unsupervised Learning (UL):** An ML method where algorithms learn patterns exclusively from unlabeled data. One of the three primary ML techniques, clustering, association rules, and dimensionality reduction, is used to get the output of this kind of learning. In the medical field, clustering can be used, for instance, to find coherent groups of patients that are well-separated and have comparable demographics (such as age and gender) and shared clinical histories [16].
- **Semi-Supervised Learning (SSL):** Semi-supervised learning combines the strengths of both SL and UL ML techniques by leveraging both labeled and unlabeled data during training. In a diagnostic context, SSL provides the means to improve classification performance by using the massive volumes of unlabeled medical data obtained during normal clinical procedures, all without requiring large, fully-labeled data sets [17].
- **Reinforcement Learning (RL):** In RL, the model is trained by providing positive environmental feedback for desired actions and negative feedback for undesired actions.RL is less common than

supervised and unsupervised learning, but it has significant applications in various fields, including clinical research [18].

1.1.1 Supervised Learning Methods

AI employs algorithms specifically created to assimilate information from extensive medical datasets and identify their correlation with a certain condition (such as a disease or disease stage) or outcome (such as time to clinical events or death) in order to aid in clinical practice. These algorithms incorporate self-updating instructions to enhance accuracy by incorporating regular feedback input. This helps to minimize clinical errors and provides the opportunity for real-time diagnostic and prognostic judgments [15], including supervised learning approaches are widely used across various domains, including Gradient Boosting (GB), Decision Trees (DT), Logistic Regression (LR), Random Forests (RF), and Support Vector Machines (SVM). These techniques are detailed below:

- **Logistic Regression (LR):** This method categorizes data by predicting the probability that a data point belongs to a certain class, with the output value ranging between 0 and 1. It is particularly popular for binary classification problems, utilizing the sigmoid function as a crucial component of the classifier [19].
- **Decision Tree (DT):** This technique addresses classification problems and can also be applied to regression tasks. It constructs a model based on simple decision rules derived from the training data, where each leaf node represents an outcome, internal nodes signify attributes, and branches indicate decision rules [20].
- **Gradient-boosting (GB):** This ML technique builds strong predictive models by applying the boosting approach, where a series of trees are constructed sequentially, each one improving upon the previous iteration. It comprises three elements: a loss function, a weak learner (such as a DT), and an additive model [21].
- **Random Forest (RF):** Introduced by Breiman in 2001, RF is an ensemble learning algorithm that employs multiple training subsets derived from the original dataset using the Bootstrap sampling method. Each subset is used to train a decision tree, and the final model is an aggregation of these trees [22].
- **Support Vector Machines (SVM):** SVM is a supervised method used for both regression and classification tasks. It works by finding a hyperplane in an N-dimensional space that best separates the different classes, aiming to create the most effective decision boundaries [23].

The most important DL algorithms are:

- **Artificial Neural Network (ANN):** is made to mimic human brain anatomy in order to duplicate studies on humans. Neural networks consist of input and output layers in addition to a hidden layer that houses components that transform data into a format that the output layer can use. ANN is made up of three linked layers. Another name for it is a Feed-Forward Neural Network [24].
- **Convolutional Neural Networks (CNNs):** represent a subset of the widely utilized artificial neural networks (ANNs) in image processing. Over the past thirty years, since CNN was first developed in 1989, a distinct kind of CNN has been introduced and has shown remarkable performance in disease identification. The input layer, hidden layer, and output layer make up the three layers of a CNN architecture [19].

2 Related works

The following is an overview of earlier research that used ML and DL approaches to diagnose liver problems. **Table 1** summarises these studies and includes the source number, publication year, topic, number of patients, techniques employed, and best accuracy attained. The accuracy of each ML and DL technique from past research that was collated in this study is shown in **Figure 2**.

The proportion of each (ML) and (DL) technique used in previous studies, which have been compiled into this research, is shown in **Figure 1**.

- Elias et al. [25] constructed a prediction model for liver disease using ML approaches, with the voting classifier showing superior performance with an Area Under Curve (AUC) of 88.4% and other metrics such as recall, accuracy, F-measure, and precision around 80%.
- Zheyu et al. [26] proposed a unique Decision Tree model to improve the accuracy of liver cirrhosis diagnosis in Hepatocellular Carcinoma patients. The DT model demonstrated exceptional diagnostic accuracy in both the training and testing populations, achieving an AUC of 0.853 and 0.817 in the ROC curve (Receiver Operating Characteristic), respectively.
- Fei Chen et al. [27] constructed a prediction model based on individual Bile Acid (BA) profiles to detect Compensated Advanced Chronic Liver Disease (CACLD) using SVM and RF, with equivalent scores for the improved SVM model around 0.86, 0.84, and 0.85.
- Ruhul et al. [28] proposed an integrated feature extraction method for classifying liver injury patients. They used various ML methods including LR, K-nearest Neighbor (KNN), Voting Classifier, SVM, Multilayer Perceptron (MLP), and RF achieving an accuracy of 88.10%.
- Jing et al. [29] introduced ML models for liver disease prediction using SVM, Gaussian process (GP), eXtreme Gradient Boosting(XGBoost), bagging, and RF algorithms, achieving the highest accuracy of 80.35%.
- Ke Chena et al.[30] developed a prediction model for liver cirrhosis in Wilson Disease patients, achieving an AUC of 0.78 and an accuracy of 0.76 in the testing set using XGBoost.
- Anıl Utku [31] suggested a deep learning model using Multilayer Perceptron to predict cirrhosis likelihood, outperforming other ML models with 80.48% accuracy and 85.71% recall, F1-score, and precision.
- Mikolaj et al. [32] developed a Deep Neural Network using CNN to distinguish cirrhosis patients with an accuracy of 86%.
- Ji-Yuan et al. [33] proposed a non-invasive model to predict HBV-related liver inflammation using RF, achieving 69.17% sensitivity, 81.44% specificity, and 73.8% accuracy.
- Xiangyu et al. [34] developed a non-invasive diagnostic algorithm for identifying Chronic Hepatitis B (CHB) related liver cirrhosis using binary logistic regression and Lasso regression for feature selection. Their model achieved an AUC of 0.852 in validation cohorts.

Table 1 - An overview of the previously listed studies.

SN	Reference	Year	Subject	No. of patients	Methods of Classification	Outcomes
1.	Elias et al. [25]	2023	Liver Disease	579	MLP, Voting, LR, RF, Random Tree (RT), J48, Naive Bayes (NB), SVM, ANN	Accuracy 80 %
2.	Zheyu et al. [26]	2023	Liver Cirrhosis	240	DT	AUC 81%
3.	Fei Chen et al. [27]	2023	CACLD	159	RF, SVM	Accuracy 82%
4.	Ruhul Amin et al. [28]	2022	Liver Disease	583	MLP, SVM, RF, LR, KNN	Accuracy 88.10 %
5.	Jing et al. [29]	2022	Liver Disorder	345	SVM, GP, XGBoost, Bagging, RF	Accuracy 80.35%
6.	Ke Chena et al.[30]	2022	Liver Cirrhosis	346	XGBoost	Accuracy 76 %
7.	Anıl Utku [31]	2022	Liver Cirrhosis	418	MLP	Accuracy 80.48%
8.	Mikolaj et al. [32]	2022	Liver Cirrhosis	46	CNN	Accuracy 86%
9.	Xiangyu et al. [34]	2021	HBV-related liver cirrhosis	754	logistic regression	AUC 85 %
10.	Ji-Yuan et al.[33]	2021	hepatic inflammation	650	RF	Accuracy 73.8%

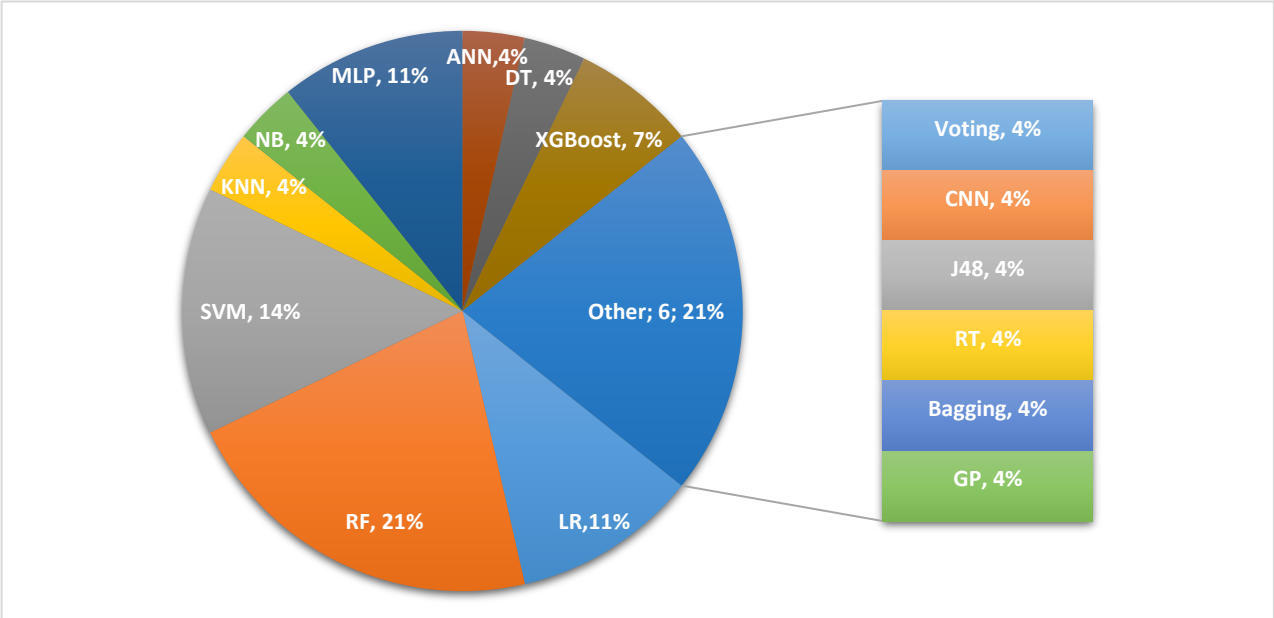

Figure .1: A pie chart for the utilization of ML and DL methodologies in the studies referred to above.

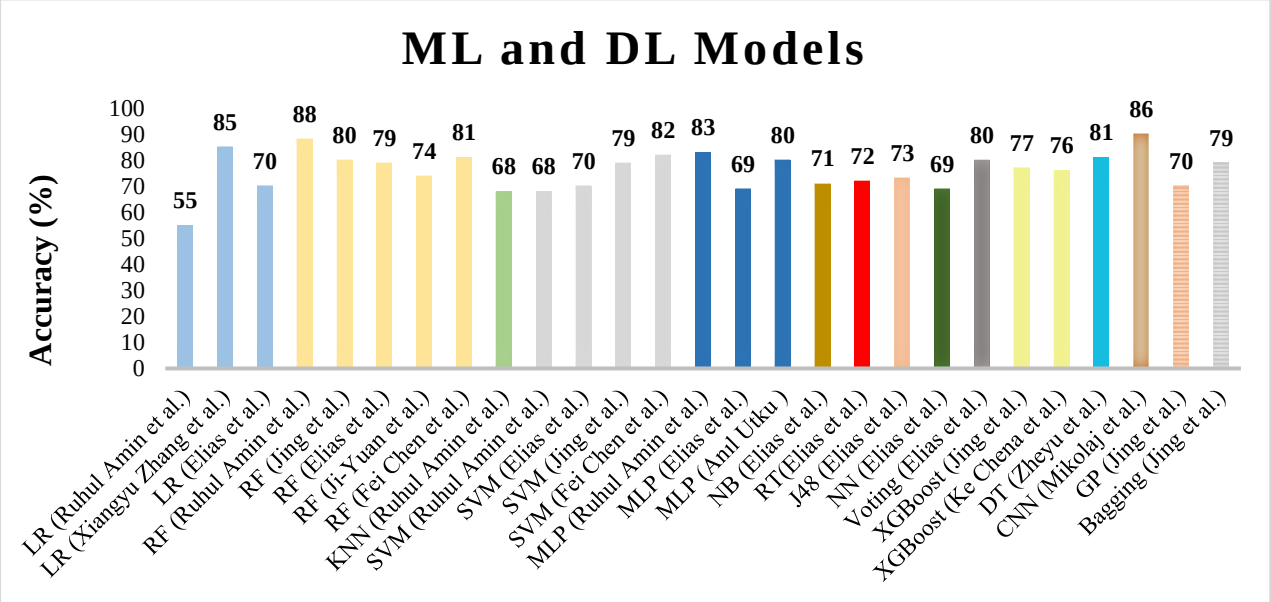

Figure .2: Evaluation of the identification of liver disease using ML and DL models for accuracy**.**

3 Publicly Accessible Liver Disorder Datasets

The general datasets pertaining to liver disease detection are enumerated below and are summarized in **Table 2**, it encompasses classes, datasets, features, and samples. The most common characteristics found in the liver disease datasets are outlined in **Table 3.**

3.1 Related Datasets

a) Indian Liver Patient Dataset (ILPD): Utilized for categorizing liver illnesses, the ILPD consists of 583 patients and 11 attributes, including class (both with and without liver disease), Albumin and Globulin Ratio, Total Proteins, Alkaline Phosphatase, Total Bilirubin, Alamina Aminotransferase, gender, age, Albumin, Aspartate Aminotransferase, and Direct Bilirubin, accessible via the University of California Irvine ML Repository repository (UCI) [21].

b) Hepatitis Dataset (HD): This dataset encompasses 20 distinctive characteristics accessible in the UCI Repository, comprising 155 patient records. Attributes include histopathology, steroids, age, weariness, malaise, liver_big, anorexia, and antivirals, among others such as protime, gender (male or female),alk_Phosphate, ascites, liver_firm, sgot, varices, bilirubin, albumin, and class (lived or died) [23].

c) HCV Dataset: Downloadable from Kaggle, the HCV dataset comprises 14 characteristics and 615 records encompassing Hepatitis C patients (categorized as fibrosis, hepatitis C, cirrhosis). Attributes include age, gender, class, ID, and healthy patients (blood donors) ALT, CREA, GGT, AST, CHE, PROT, BIL, and CHO [35].

d) Liver Disorders Dataset (BUPA): Available in the UCI repository, comprising 345 entries, the BUPA dataset encompasses 7 characteristics such as selector, drinks, MCV (Mean Corpuscular Volume), Gammagt (gamma-glutamyl transpeptidase), SGOT (aspartate aminotransferase), SGPT (alanine aminotransferase) and alkphos (alkaline phosphatase) [29].

e) **Hepatocellular Carcinoma Dataset (HCC):** This dataset consists of demographic, risk factors, laboratory, and overall survival characteristics of 165 records diagnosed with HCC, featuring 49 variables used for predicting an HCC patient's survival, essential for clinical decision-making [36] .

f) Cirrhosis Prediction Dataset (CP): Encompassing 418 Primary biliary cholangitis (PBC) patients and 20 features are symptoms of the disease such as jaundice, ascites, edema, etc., as well as blood tests for liver function and patient information, such as age and gender, the CP dataset, is available to all users on Kaggle [31].

Table 2 - An overview of the liver disease public datasets.

Datasets	Number of classes	Number of records	Number of features	Resource
ILPD dataset	2	583	11	[37]
HD dataset	2	155	20	[38]
HCV dataset	3	615	14	[39]
BUPA dataset	2	345	7	[40]
HCC dataset	2	165	49	[41]
CP dataset	4	418	20	[42]

Table 3 - recurring features in open datasets.

Important features	Datasets
Gamma-glutamyl transpeptidase (GGT)	HCV, BUPA, HCC
Alanine transaminase (ALT)	ILPD, HCV, BUPA, HCC
Aspartate aminotransferase (AST)	ILPD, HD, HCV, BUPA, HCC, CP
Albumin (ALB)	ILPD, HD, HCV, CP
Alkaline phosphatase (ALP)	ILPD, HD, HCV, HCC, CP
Bilirubin	ILPD, HD, HCV, HCC, CP

3.2 Suggested Dataset

➢ Liver Disorders Dataset (LD)

LD dataset was acquired from the Zenodo website [43], accessible through a link specified on the GitHub website. It encompasses 70 features and includes data from 10,000 patients, featuring patient information, liver function tests, and symptoms related to liver diseases such as age, gender, Alt, Ast, diabetes, obesity, Hyperbilirubinemia, triglycerides, bleeding, jaundice, among others. The target class in this dataset is Cirrhosis, with three categories: absent, compensated, and decompensated cases.

Given that data extracted from patient records may not always be entirely accurate, the dataset underwent pre-processing operations to ensure its readiness for use in ML models and to yield more precise diagnostic outcomes. Pre-processing involves converting raw data into a structured format [35].

Table 4 provides detailed information about the suggested dataset.

Table 4 - Details of the dataset suggested.

Dataset	Number of instances	Attributes	Class	Resource
Liver disorders	10,000	30	3	Zenodo

4 Measures of Performance Evaluation

Researchers evaluating the efficacy of prediction algorithms for liver disease have employed various metrics, including accuracy, F1-score, specificity, and sensitivity. Among these measures accuracy and area under the curve (AUC) are the most commonly used. Here are the specifics of each measure:

- **Accuracy:** This metric is widely utilized and is defined as the ratio of correctly identified instances to all instances within each test dataset [25].It can be mathematically represented as follows:

$$\textbf{Accuracy} = \frac{\textbf{TP+TN}}{\textbf{TP+TN+FP+FN}} \qquad \textbf{(1)}$$

  - **True Positive (TP):** indicates a positive output, resulting in the accurate classification of the expected outcome.
  - **False Positive (FP):** indicates a positive output, leading to an incorrect classification of the expected outcome.
  - **True Negative (TN):** indicates a negative output so that the anticipated outcome is correctly categorized.
  - **False Positive (FN):** indicates a negative output, which leads to an incorrect classification of the expected outcome [44].
- **Precision:** When retrieving information, precision only yields favorable outcomes. In the event that the real positive plus false positive equals zero, precision will return zero [45]. The computation is as follows:

$$\textbf{Precision} = \frac{\textbf{TP}}{\textbf{TP+FP}} \qquad \textbf{(2)}$$

- **Specificity:** To calculate true negatives (TN) using false positives (FP) and TN, specificity assesses the accuracy of negative predictions [45]. The computation is as follows:

$$\textbf{Specificity} = \frac{\textbf{TN}}{\textbf{TN+FP}} \qquad \textbf{(3)}$$

- **Sensitivity:** It measures the classifier's ability to detect positive samples accurately, indicating the proportion of actual positive samples identified by the model [20]. The mathematical calculation of sensitivity is represented by the following equation:

$$\textbf{Sensitivity} = \frac{\textbf{TP}}{\textbf{TP+FN}} \qquad \textbf{(4)}$$

- **F1-score:** The F-Measure (F1) is the weighted average of precision and recall, incorporating false positive and false negative readings. It is considered more valuable than accuracy as it accounts for both types of errors [45]. Following is the computation of the F-Measure:

$$\textbf{F1-score} = \frac{\textbf{2(Recall*Precision)}}{\textbf{(Recall+Precision)}} \qquad \textbf{(5)}$$

- **ROC AND AUC:** The Receiver Operating Characteristic Curve (ROC) assesses the performance of a classification model by comparing the True Positive (TP) rate against the False Positive (FP) rate at different classification thresholds. The area under the ROC curve (AUC) represents the total area under the curve, offering an overall performance metric across all possible classification thresholds [44].

5 Restrictions and Future Work

This section will focus on the key limitations identified in previous studies related to predicting liver disease, followed by the proposed prospects outlined below:

5.1 Restrictions

- The existing liver disease datasets, including those for cirrhosis, are often small in size (usually under one thousand samples) and publicly available. This small sample size can lead to lower predictive accuracy in artificial intelligence models, potentially resulting in inaccurate disease diagnosis. Obtaining larger clinical datasets from hospitals and clinics can be challenging due to concerns regarding patient privacy.
- Research on the diagnosis of cirrhosis is comparatively rare, and what is known about it frequently focuses on forecasting the disease's presence or absence rather than assessing its severity or stage. Effective management and therapy of cirrhosis depend on early detection and precise staging of the disease.
- Limited availability of publicly available datasets devoted to liver disease, especially those including individuals with cirrhosis.

5.2 Future Work

For future work aimed at predicting the stage of cirrhosis using (AI) models, we plan to leverage a large dataset that has not been previously utilized. Our approach involves comprehensive data processing techniques to preprocess and refine the dataset. Additionally, we intend to integrate several methods for (ML) methodologies to improve the accuracy of diagnosing cirrhosis.

6 Conclusion

This study explores recent research on ML and DL methods in predicting liver disorders related to cirrhosis. Among the algorithms summarized in previous studies, RF and SVM are the most widely used ML approaches for this topic. Accuracy, specificity, and sensitivity are the performance metrics used.

Common characteristics were also identified among publicly available liver disease datasets. The study examines important challenges to AI-assisted early detection of liver cirrhosis stages and our future endeavors aim to increase the size of the dataset and explore additional machine learning models and feature selection techniques to refine the predictive accuracy further.

References

[1] L. A. P. Faccioli, M. L. Dias, B. A. Paranhos, and R. C. dos Santos Goldenberg, “Liver cirrhosis: An overview of experimental models in rodents,” Jul. 15, 2022, Elsevier Inc. doi: 10.1016/j.lfs.2022.120615.

[2] H. Zhangdi et al., “BMSCs alleviate liver cirrhosis by regulating Fstl1/Wnt/β-Catenin signaling

pathway," Heliyon, vol. 9, no. 11, Nov. 2023, doi: 10.1016/j.heliyon.2023.e21010.

[3] X. Liu, R. L. Ma, J. Zhao, J. L. Song, J. Q. Zhang, and S. H. Wang, "A clinical decision support system for predicting cirrhosis stages via high frequency ultrasound images," Expert Syst. Appl., vol. 175, Aug. 2021, doi: 10.1016/j.eswa.2021.114680.

[4] E. Hasa, P. Hartmann, and B. Schnabl, "Liver cirrhosis and immune dysfunction," Int. Immunol., vol. 34, no. 9, pp. 455–466, Sep. 2022, doi: 10.1093/intimm/dxac030.

[5] C. C. Cardoso et al., "B-cell compartment abnormalities are associated with ACLF and mortality in patients with liver cirrhosis," Clin. Res. Hepatol. Gastroenterol., vol. 45, no. 4, Jul. 2021, doi: 10.1016/j.clinre.2021.101698.

[6] X. lan Xu et al., "The role of fibrosis index FIB-4 in predicting liver fibrosis stage and clinical prognosis: A diagnostic or screening tool?," Feb. 01, 2022, Elsevier B.V. doi: 10.1016/j.jfma.2021.07.013.

[7] Y. Yin, D. Yakar, R. A. J. O. Dierckx, K. B. Mouridsen, T. C. Kwee, and R. J. De Haas, "Liver fibrosis staging by deep learning: a visual-based explanation of diagnostic decisions of the model", doi: 10.1007/s00330-021-08046-x/Published.

[8] U. Teufel-Schäfer, C. Flechtenmacher, A. Fichtner, G. F. Hoffmann, J. P. Schenk, and G. Engelmann, "Transient elastography correlated to four different histological fibrosis scores in children with liver disease", doi: 10.1007/s00431-021-04001-6/Published.

[9] M. Méndez, M. G. Merayo, and M. Núñez, "Machine learning algorithms to forecast air quality: a survey," Artif. Intell. Rev., vol. 56, no. 9, pp. 10031–10066, Sep. 2023, doi: 10.1007/s10462-023-10424-4.

[10] M. A. Aljabery and S. Kurnaz, "Applying datamining techniques to predict hearing aid type for audiology patients," J. Inf. Sci. Eng., vol. 36, no. 2, pp. 205–215, 2020, doi: 10.6688/JISE.202003_36(2).0002.

[11] P. I. Adegbola, A. B. Aborisade, and A. Adetutu, "Liver disease prediction in rats exposed to environmental toxicants using Machine-learning techniques," Informatics Med. Unlocked, vol. 42, Jan. 2023, doi: 10.1016/j.imu.2023.101369.

[12] D. E. M. Nisar, R. Amin, N. U. H. Shah, M. A. A. Ghamdi, S. H. Almotiri, and M. Alruily, "Healthcare Techniques through Deep Learning: Issues, Challenges and Opportunities," IEEE Access, vol. 9, pp. 98523–98541, 2021, doi: 10.1109/ACCESS.2021.3095312.

[13] H. Shaheen, K. Ravikumar, N. Lakshmipathi Anantha, A. Uma Shankar Kumar, N. Jayapandian, and S. Kirubakaran, "An efficient classification of cirrhosis liver disease using hybrid convolutional neural network-capsule network," Biomed. Signal Process. Control, vol. 80, Feb. 2023, doi: 10.1016/j.bspc.2022.104152.

[14] D. Nam, J. Chapiro, V. Paradis, T. P. Seraphin, and J. N. Kather, "Artificial intelligence in liver diseases: Improving diagnostics, prognostics and response prediction," 2022.

[15] G. D'Amico, A. Colli, G. Malizia, and G. Casazza, "The potential role of machine learning in modelling advanced chronic liver disease," Jun. 01, 2023, Elsevier B.V. doi: 10.1016/j.dld.2022.12.002.

[16] M. A. Makroum, M. Adda, A. Bouzouane, and H. Ibrahim, "Machine Learning and Smart Devices for Diabetes Management: Systematic Review," Mar. 01, 2022, MDPI. doi: 10.3390/s22051843.

[17] J. N. Eckardt, M. Bornhäuser, K. Wendt, and J. M. Middeke, "Semi-supervised learning in cancer diagnostics," Front. Oncol., vol. 12, no. July, pp. 1–10, 2022, doi: 10.3389/fonc.2022.960984.

[18] J. Peng, E. C. Jury, P. Dönnes, and C. Ciurtin, "Machine Learning Techniques for Personalised Medicine Approaches in Immune-Mediated Chronic Inflammatory Diseases: Applications and Challenges," Sep. 30, 2021, Frontiers Media S.A. doi: 10.3389/fphar.2021.720694.

[19] M. M. Ahsan, S. A. Luna, and Z. Siddique, "Machine-Learning-Based Disease Diagnosis: A Comprehensive Review," Mar. 01, 2022, MDPI. doi: 10.3390/healthcare10030541.

[20] M. Hanafy and R. Ming, "Machine learning approaches for auto insurance big data," Risks, vol. 9, no. 2, pp. 1–23, 2021, doi: 10.3390/risks9020042.

[21] A. Q. Md, S. Kulkarni, C. J. Joshua, T. Vaichole, S. Mohan, and C. Iwendi, "Enhanced Preprocessing Approach Using Ensemble Machine Learning Algorithms for Detecting Liver Disease," Biomedicines, vol. 11, no. 2, Feb. 2023, doi: 10.3390/biomedicines11020581.

[22] G. S. Ohannesian and E. J. Harfash, "Epileptic Seizures Detection from EEG Recordings Based on a Hybrid System of Gaussian Mixture Model and Random Forest Classifier," Inform., vol. 46, no. 6, 2022, doi: 10.31449/inf.v46i6.4203.

[23] G. Obaido et al., "An Interpretable Machine Learning Approach for Hepatitis B Diagnosis," Appl. Sci., vol. 12, no. 21, Nov. 2022, doi: 10.3390/app122111127.

[24] S. S. Islam, M. S. Haque, M. S. U. Miah, T. Bin Sarwar, and R. Nugraha, "Application of machine learning algorithms to predict the thyroid disease risk: an experimental comparative study," PeerJ Comput. Sci., vol. 8, 2022, doi: 10.7717/PEERJ-CS.898.

[25] E. Dritsas and M. Trigka, "Supervised Machine Learning Models for Liver Disease Risk Prediction," Computers, vol. 12, no. 1, p. 19, Jan. 2023, doi: 10.3390/computers12010019.

[26] Z. Zhou et al., "A decision tree model to predict liver cirrhosis in hepatocellular carcinoma patients: a retrospective study," PeerJ, vol. 11, 2023, doi: 10.7717/peerj.15950.

[27] F. Chen, Y. Yao, Z. Li, L. Deng, and R. He, "Assessment of compensated advanced chronic liver disease based on serum bile acids in chronic hepatitis B patients," Sci. Rep., vol. 13, no. 1, Dec. 2023, doi: 10.1038/s41598-023-39977-8.

[28] R. Amin, R. Yasmin, S. Ruhi, M. H. Rahman, and M. S. Reza, "Prediction of chronic liver disease patients using integrated projection based statistical feature extraction with machine learning algorithms," Informatics Med. Unlocked, vol. 36, Jan. 2023, doi: 10.1016/j.imu.2022.101155.

[29] J. Zhao, P. Wang, and Y. Pan, "Predicting liver disorder based on machine learning models," J. Eng., vol. 2022, no. 10, pp. 978–984, Oct. 2022, doi: 10.1049/tje2.12184.

[30] K. Chen, Y. Wan, J. Mao, Y. Lai, G. Zhuo-Ma, and P. Hong, "Liver cirrhosis prediction for patients with Wilson disease based on machine learning: A case-control study from southwest China," Eur. J. Gastroenterol. Hepatol., vol. 34, no. 10, pp. 1067–1073, Oct. 2022, doi: 10.1097/MEG.0000000000002424.

[31] A. Utku, "DEEP LEARNING BASED CIRRHOSIS DETECTION," Oper. Res. Eng. Sci. Theory Appl., vol. 6, pp. 95–114, 2023, doi: 10.31181/oresta/060105.

[32] M. Wieczorek, A. Weston, M. Ledenko, J. N. Thomas, R. Carter, and T. Patel, "A deep learning approach for detecting liver cirrhosis from volatolomic analysis of exhaled breath," Front. Med., vol. 9, Sep. 2022, doi: 10.3389/fmed.2022.992703.

[33] J. Y. Zhou, L. W. Song, R. Yuan, X. P. Lu, and G. Q. Wang, "Prediction of hepatic inflammation in chronic hepatitis B patients with a random forest-backward feature elimination algorithm," World J. Gastroenterol., vol. 27, no. 21, pp. 2910–2920, Jun. 2021, doi: 10.3748/wjg.v27.i21.2910.

[34] X. Zhang et al., "Development and validation of a non-invasive model for diagnosing HBV-related liver cirrhosis," Clin. Chim. Acta, vol. 523, pp. 525–531, Dec. 2021, doi: 10.1016/j.cca.2021.11.004.

[35] R. Safdari, A. Deghatipour, M. Gholamzadeh, and K. Maghooli, "Applying data mining techniques to classify patients with suspected hepatitis C virus infection," Intell. Med., vol. 2, no. 4, pp. 193–198, Nov. 2022, doi: 10.1016/j.imed.2021.12.003.

[36] M. Sharma and N. Kumar, "Improved hepatocellular carcinoma fatality prognosis using ensemble learning approach," J. Ambient Intell. Humaniz. Comput., vol. 13, no. 12, pp. 5763–5777, Dec. 2022, doi: 10.1007/s12652-021-03256-z.

[37] "ILPD (Indian Liver Patient Dataset)." Accessed: Jun. 01, 2024. [Online]. Available: https://archive.ics.uci.edu/dataset/225/ilpd+indian+liver+patient+dataset

[38] "Hepatitis Dataset ." Accessed: Jun. 01, 2024. [Online]. Available: https://archive.ics.uci.edu/dataset/46/hepatitis

[39] "HCV data." Accessed: Jun. 01, 2024. [Online]. Available: https://archive.ics.uci.edu/dataset/571/hcv+data

[40] " Liver Disorders." Accessed: Jun. 01, 2024. [Online]. Available: https://archive.ics.uci.edu/dataset/60/liver+disorders

[41] "HCC Survival." Accessed: Jun. 01, 2024. [Online]. Available: https://archive.ics.uci.edu/dataset/423/hcc+survival

[42] "Cirrhosis Patient Survival Prediction", Accessed: Jun. 01, 2024. [Online]. Available: https://www.archive.ics.uci.edu/dataset/878/cirrhosis+patient+survival+prediction+dataset-1

[43] "Liver disorders ." Accessed: Feb. 26, 2024. [Online]. Available: https://zenodo.org/records/6726768#.Y1vxci8Rpz8

[44] P. Chittora et al., "Prediction of Chronic Kidney Disease - A Machine Learning Perspective," 2021, Institute of Electrical and Electronics Engineers Inc. doi: 10.1109/ACCESS.2021.3053763.

[45] V. Anthonysamy and S. K. K. Babu, "Multi Perceptron Neural Network and Voting Classifier for Liver Disease Dataset," IEEE Access, vol. 11, pp. 102149–102156, 2023, doi: 10.1109/ACCESS.2023.3316515.

View publication stats