OPEN ACCESS JOURNALS
Indiana Journal of Multidisciplinary Research
Abbriviate Tittle- Ind J Multi Res
ISSN *(Online)*- 2583-3820
Journal Homepage Link- https://indianapublications.com/Journals/IJMR
DOI: https://doi.org/10.5281/zenodo.12685527

INDIANA PUBLICATIONS
PRODUCTIVE AND QUALITY RESEARCH

Research Article

Volume-04|Issue-03|2024

Predictive Analytics for Liver Cirrhosis: Leveraging Machine Learning to Enhance Early Detection and Intervention

Prof. Samatha R Swamy[1], S Harini*[2], Sanjana A[3]

[1]Assistant Professor, Information Science & Engineering Department, RV Institute of Technology and Management, Bengaluru, Karnataka, India.
[2,3]Student in Information Science & Engineering Department, RV Institute of Technology and Management, Bengaluru, Karnataka, India.

Article History
Received: 20.05.2024
Accepted: 05.06.2024
Published: 30.06.2024

Citation
Swamy, S. R., Harini, S., & Sanjana, A. (2024). Predictive Analytics for Liver Cirrhosis: Leveraging Machine Learning to Enhance Early Detection and Intervention. *Indiana Journal of Multidisciplinary Research, 4*(3), 248-253.

Abstract: Liver disease poses a vital issue for public health, all over the world, liver cirrhosis being one of its severe manifestations. This study endeavors to make a prognostic model for liver cirrhosis utilization of machine learning methods, specifically employing a Random Forest classifier. The dataset used in this research comprises clinical parameters of patients, including age, gender, and different biochemical indicators linked to liver functionality. Through rigorous data pre-processing, including handling missing values and encoding categorical variables, followed by feature selection, the dataset is prepared for model training. The Random Forest classifier is optimized using hyperparameter tuning with GridSearchCV to enhance predictive performance. The resulting model demonstrates promising accuracy and generalization ability, as evidenced by the evaluation metrics. The results indicate that machine learning approaches can aid healthcare practitioners in the early detection and prognosis of liver cirrhosis, thereby facilitating timely interventions and improving patient outcomes.

Keywords: Liver cirrhosis, Machine learning, Random Forest classifier, Predictive model, Grid Search CV, Hyper parameter tuning.

INTRODUCTION

In a world where liver disease remains a significant health concern, affecting millions of individuals globally, those who have been diagnosed with liver cirrhosis face unique challenges in managing their condition effectively **[1]**. This research addresses a critical need in liver disease management; Accurately predicting cirrhosis using advanced machine-learning techniques. Given the global burden that cirrhosis imposes on healthcare systems, the development of predictive models becomes imperative to enable timely diagnosis and intervention. This study aims to create a model predicting liver cirrhosis by utilizing clinical information taken from patients through random forest distribution. Since cirrhosis is a multifactorial disease affected by many demographic and biochemical factors, including factors such as age, gender, and biochemical markers related to liver function, it is very important for accurate prediction. This study aims to contribute to global health by providing a reliable prediction model for cirrhosis also its local impact. By emphasizing efficiency over technology, the aim is to create solutions that are easily implemented in the clinic to benefit patients around the world.

The proposed predictive model is intended for equipping healthcare professionals with resources that help identify individuals at risk of developing high blood pressure. Through the use of machine learning algorithms, the model is designed to increase the accuracy of diagnosis and facilitate timely intervention, ultimately improving patient outcomes. The approach adopted by this study extends from data collection to the distribution model, focusing on optimizing the performance of the competition valley random forest. The goal is to create a powerful and reliable model to be integrated into existing treatment, leveraging the latest educational technology. The next section of this article explains this approach by describing the steps in data collection, feature selection, model training, and evaluation. The results are presented and discussed to demonstrate the effectiveness of the proposed prediction model in the accurate diagnosis of cirrhosis. This study aims to close the divide between technological development and clinical practice by developing a cirrhosis testing model that will assign doctors to early diagnosis and intervention. It aims to improve patient care and liver disease management by providing the highest level of measurement to the healthcare system. Additionally, this research focuses on advancing personalized medicine in the treatment of liver disease by adjusting predictive models based on the patient's condition. By integrating patient-specific information such as medical history, genetic predisposition, and lifestyle into the prediction process, the plan is designed to enhance the accuracy and specificity of liver disease prediction. From the patient's perspective, physicians can gain a deeper understanding

***Corresponding Author:** S Harini

of each patient's disease risk and disease progression, allowing for personalized treatment strategies and intervention planning. This approach not only increases the predictive power of the model but also supports a more patient-centered healthcare system, ultimately leading to better assessment and quality of life for people with cirrhosis.

LITERATURE SURVEY

Many academic publications on liver diagnosis and prognosis, including scientific articles, clinical studies, and medical journals, were systematically reviewed. This evidence formed the main basis for the development of the methods and methods of this research, with special reference to studies that have led to positive results towards the understanding of various conditions of cirrhosis. Significant publications, including scientific literature and comprehensive reviews, have played an important role in guiding the range of diagnostic tests, identifying biomarker-related indicators, and demonstrating differences in demographic and biochemical factors that deal with the development of cirrhosis. By gaining insights from the wealth of existing knowledge in these studies, the objective of this research is to link advances in liver disease research to predictive models that show the potential for early diagnosis and intervention, synthesizing information from diverse resources exemplifies the process of combining research outcomes, our commitment to rigor and excellence in advancing our understanding of cirrhosis and improving clinical patient care.

Huang *et al.* [1], a thorough examination of existing literature underscores the significant impact of cirrhosis on global morbidity and mortality, underscoring cirrhosis as a worldwide health concern. In a comprehensive study featured in The Lancet Gastroenterology & Hepatology 2019, researchers noted a shift in the landscape of cirrhosis, with the prevalence of cirrhosis on the rise alongside factors such as obesity, alcohol consumption, and liver disease. Furthermore, the authors emphasized the advancement of non-alcoholic fatty liver disease (NAFLD) and alcohol-induced cirrhosis, as well as chronic liver disease, the primary cause of cirrhosis globally, in their 2019 publication in the Journal of Hepatology. An analysis published in the Journal of the American Medical Association in 2018 shed light on the pattern of infectious disease transmission by investigating rates of hospital-acquired cirrhosis and mortality among U.S. Medicare beneficiaries. There is an immediate necessity for early intervention and enhanced access to care to alleviate the burden of cirrhosis in the years ahead. Through collaborative efforts to address these challenges, researchers strive to diminish the morbidity and mortality linked to cirrhosis and enhance the overall well-being of affected individuals worldwide. Dritsas *et al.* [2], when unraveling the complexity of the liver through machine learning, it is important to understand the important role of the liver as the largest tumor in the body. The liver performs numerous functions., including processing nutrients in food and water, filtering toxins from the blood, and improving the immune system. However, exposure to bacteria or toxins may result in liver damage, ultimately leading to liver disease. Liver diseases encompass conditions that impair liver function, pose a threat to human health, and necessitate prompt medical intervention. In light of this context, the current study explores the field of early liver disease prognosis using machine learning techniques. By utilizing a range of machine learning and integration models, the research aims to develop reliable prediction algorithms that can accurately identify individuals who are susceptible to liver disease. The primary objective of this investigation is to assess the efficacy of different predictors in forecasting the onset of liver disease by meticulously scrutinizing and contrasting these models based on metrics such as accuracy, precision, regression, F measure, and area under the curve (AUC). Experimental findings demonstrate the significance of voting, which emerges as the most influential model in the prediction framework. The voter exhibits a dislike towards the identification of liver disease, showcasing an intrinsic accuracy, recall, and F-measure of 80.1%. Additionally, it showcases a sensitivity of 80.4% and an AUC of 88.4% after undergoing 10-fold cross-validation following SMOTE. Ability to predict onset. To confirm these findings, we conducted a literature review, extracting information from seminal studies and academic publications on the prediction of liver disease. Key results include research studies on the use of machine learning in liver diagnosis and prognosis, showing interactions between treatments and predicting outcomes. Research has demonstrated the efficacy of different machine-learning techniques and algorithms in identifying patterns of liver disease, paving the way for advances in predictive models and personalized medicine. By combining findings from studies conducted in this course, this study leads to a body of knowledge for improving early detection and intervention strategies. Through collaboration and collaborative research, the ultimate goal is to reduce the burden of liver disease and improve patient outcomes worldwide.

Feng et al [3], the development of non-invasive techniques for forecasting liver cirrhosis in autoimmune hepatitis (AIH) represents a major advance in the field of hepatology. It offers a hopeful route for early detection and intervention in a pioneering study by researchers aimed to develop a prediction model based on two-dimensional shear wave elasticity (2D-SWE) technology, ultrasonic features, and serological indicators to identify Cirrhosis in AIH patients. This study focused on AIH patients diagnosed by liver biopsy, ultrasound, and serology, and carefully collected summary, ultrasound, and serological tests to define cirrhosis and non-cirrhosis. Using multivariable logistic regression, researchers identified independent risk factors predicting cirrhosis, including liver stiffness (LS), splenomegaly, and complement C4 levels. Following the development of the autoimmune hepatitis cirrhosis (AIHC) prediction model, it was found to showcase superior performance compared to other parameters. The cut-off agreement was established based on sensitivity and specificity values from the Youden Index, emphasizing the model's usefulness in clinical practice. Validation of the AIHC model showcased its strength and clinical applicability with a high area under the curve (AUC), good net reclassification index (NRI), and enhanced discrimination compared to alternative measures. The visualization of the AIHC model simplifies its implementation, making it more accessible for widespread use in clinical settings. This research contributes to the current literature on the unpredictable mechanisms of hypertension in AIH, shedding light on potential new infection-related mechanisms. The AIHC model is a significant advancement toward personalized medicine and enhancing patient outcomes through innovative technology and patient-centered data. Further research and collaboration are necessary to validate these findings and promote the extensive adoption of non-invasive predictive models in clinical practice. Trinh *et al.* [4], researchers investigating the connection between cirrhosis and glucose metabolism investigated the mechanisms underlying insulin resistance and impaired glucose tolerance in these patients. Trinh *et al.* [4] elucidated the dynamic interaction between cirrhosis and glucose dynamics in the body and skeletal muscle during parenteral nutrition. This study uses isotope dilution techniques and arteriovenous measurement methods that give new ideas to glucose metabolism the patients with cirrhosis and avoid adverse effects such as portosystemic shunt and impaired hepatic uptake. This study builds on previous studies that used euglycemic hyperinsulinemia clamps, or blood sugar tests, and examined abnormal glucose levels under circumstances that replicate typical dining and provide information about the body. Through careful analysis, researchers confirmed that liver disease and insulin resistance were present in patients with cirrhosis; This is due to a decrease in endogenous glucose production despite an increase in glucose in skeletal muscles. These discoveries imply a compensatory mechanism of cirrhosis in muscle that increases insulin secretion to promote regulating the absorption of glucose to ensure the balance of glucose levels within the body, in the event of insulin resistance. This study elucidates this complex process, leading to a deeper understanding of glucose metabolism in cirrhosis and demonstrating the interplay between liver function, insulin sensitivity, and peripheral glucose. This study contributes to the growing body of research on the pathophysiology of glucose dysregulation in cirrhosis and provides insight into therapeutic targets and interventions to maintain glucose homeostasis in these patients. Further studies are needed to confirm these findings and investigate the implications in the treatment of cirrhotic patients with poor glucose metabolism. Trinh *et al.* [4], the intersection of obesity and cirrhosis is receiving increasing attention due to its significant impact on health and management strategies. An important study by Kalaitzakis *et al.* [5] investigated the effectiveness of a four-year weight loss intervention following a Mediterranean diet in obese patients with compensated cirrhosis. The study assessed not just the influence of weight loss on various health aspects but also examined the impact of dietary changes on overall health. Based on previous research on the negative effects of obesity on liver health, the study included 62 patients with cirrhosis and 44 healthy patients, all with a BMI over 30 kg/mÂ². Throughout the duration of the research, both sets of participants underwent a weight loss intervention centered around the Mediterranean diet. Over four years, the participants' anthropometric, biochemical, and hematological parameters were evaluated at different intervals to monitor any health-related changes. The findings of the study revealed noteworthy enhancements in anthropometric parameters, including body weight, body mass index, waist and hip circumference, fat percentage, and grip strength, particularly within the initial year of the intervention. Furthermore, the patients experienced a significant improvement in liver function tests and lipid profiles throughout the intervention period. A noticeable improvement in hematological and biochemical parameters such as hematocrit and ferritin levels was also seen in the patient group. These findings demonstrate the potential for a weight loss intervention based on the Mediterranean diet to have beneficial effects on health in obese patients with compensated cirrhosis. By showing significant improvements in several health aspects, especially at the beginning of the intervention, this study highlights the importance of dietary changes in the management of poor health in this patient group. More research is needed in this area to confirm these findings and examine the lasting impacts of diet on heart health and overall health.

MATERIALS AND METHODS

Dataset Description:

The dataset utilized in this research comprised patient data gathered from the northeastern area of Andhra Pradesh, India. It encompassed demographic details like age and gender, along with a range of biochemical indicators linked to liver well-being. The variables considered in this research encompass the total bilirubin level, direct bilirubin level, alkaline phosphatase level, alanine aminotransferase level, aspartate aminotransferase level, total protein level, albumin level, and the albumin to globulin ratio. This dataset provides a comprehensive understanding of liver health, making it suitable for developing predictive models to diagnose cirrhosis.

Load the dataset:

***Corresponding Author:** S Harini

The information utilized in this research was obtained from the Indian Liver Patient Dataset, comprising de-identified patient data from the northeastern area of Andhra Pradesh, India. These data include patient demographics and various biochemical markers related to liver health. Upload datasets to the analysis center for further pre-processing and modelling. Accurate control of missing values and coding of categorical variables are important steps in preparing the data set for analysis. Upon completion of loading, the dataset is partitioned into a training set and a test set in order to streamline the training and testing procedures.

Data pre-processing:

Before design, pre-processing steps are involved appropriately and the data is completed. Handling missing data is a main part of preliminary data as data that is missed can impact the performance model. The dropna() function is used for removing non-significant rows to ensure data used for analysis is complete and unbiased.

Encode categorical variable 'Gender':

Code categorical variables (such as “gender”) using one-time coding and convert them to a numerical format appropriate to the study model. This transformation allows the model to process data across categories by representing each category as a binary feature.

Feature selection:

Data are partitioned into characteristics (X) and objective variables (y), covering a large range of demographic, clinical, and biochemical factors that influence liver health. This step is important for preparing data for training and evaluation models because it identifies variables obtained independently from various methodologies.

Split the dataset into training and testing sets:

To test the performance of the model, use the train_test_split() function to split the dataset into a training set and a test set. This method allows the model to be trained on a small set of objects and analyzed on a separate, undetected subset, thus simulating world performance and preventing over performance.

Model selection- Random Forest:

The random forest algorithm was chosen as the prediction model due to its versatility, robustness, and ability to handle complex data. Random forests reduce competition and increase the accuracy of predictions by creating multiple decision trees and combining their predictions.

Hyper parameter tuning using Grid Search CV:

Hyperparameter tuning is necessary to improve model performance. GridSearchCV is used to search a set of predefined hyperparameters and determine the best combination that results in best model performance. This extensive research allows the model to be precisely tuned for optimum performance.

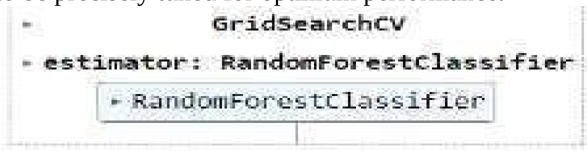


Figure 1: Hyper parameter tuning using Grid Search CV

Best parameters and best score:

After the hyper parameters were adjusted, the visual parameters and similarity scores were determined as cross-validation results. These parameters represent the best configuration of the random forest model for a given data set and serve as a blueprint for subsequent model training and evaluation.

Predictions using the best estimator from GridSearchCV:

The best guess obtained by GridSearchCV encapsulates the random forest algorithm model with the optimal hyperparameters to perform parameter estimation. This step allows the model to make predictions on unobserved data and evaluate its performance.

Model evaluation:

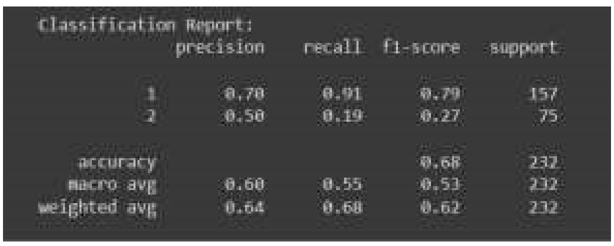
The evaluation of the random forest model's performance was conducted by utilizing different metrics including such as accuracy, precision, recall, and F1 score. These metrics provide significant insights into the model's ability to identify liver disease and its overall predictive capability.

Calculate accuracy in percentage:

Accuracy is a simple measure that evaluates the distribution model and is evaluated as the percentage of events predicted correctly for all events in the experiment. This metric is a main indicator of the model's performance and its effectiveness in predicting liver disease accuracy.

Display classification report:

A detailed report of the classification was generated to give a thorough overview of the model's performance in different groups (e.g., with or without liver disease). The report includes accuracy, recall, and F1 scores for each category, providing insight into model performance and possible domains for improvement.



Classification Report:

	precision	recall	f1-score	support
1	0.70	0.91	0.79	157
2	0.50	0.19	0.27	75
accuracy			0.68	232
macro avg	0.60	0.55	0.53	232
weighted avg	0.64	0.68	0.62	232

Figure 2: Display classification Report

***Corresponding Author:** S Harini

RESULT AND DISCUSSION

The statistical methods evaluated through the utilization of the predictive performance of the random forest model for liver hypertension included such as accuracy, precision, recall, and F1 score. Through these methods, it was determined that the model's accuracy was approximately 67.67%. This indicates that the model successfully classified around 68% of samples into their respective groups. Although this level of accuracy indicates good predictive power, further analysis is needed to better understand the model's strengths and limitations. After examining the distribution map, it is clear that the model achieves the same accuracy and return scores across the class, performing reasonably well in classification controlling for the pros and cons of cirrhosis. The correlation between precision and correct predictions across classes was observed to be strong, indicating that the noted model effectively reduces false positives. Additionally, the model's ability to recognize a significant number of positive outcomes in each category suggests its proficiency in handling the majority of cirrhosis cases. Although the performance of the model appears to be good, potential limitations and areas for improvement should be considered. Even if average accuracy is achieved, the model's performance will vary depending on distinct attributes of data and distribution of events in each category. Moreover, delving deeper into the importance of values could offer important insights into the variables influencing model predictions and guide forthcoming model adjustments.. Overall, the evaluation of this study demonstrates the potential of machine learning, particularly random forests, in predicting cirrhosis based on clinical and demographic factors. However, continued research and improvement are needed to develop a more accurate and comprehensive predictive model and ultimately help improve patient outcomes in liver diagnosis and practice.

CONCLUSION

In summary, this study demonstrates the feasibility and effectiveness of machine learning algorithms, specifically random forests, in predicting cirrhosis on a clinical and population basis. The model achieved an average accuracy of approximately 67.67%, demonstrating that it has the potential to serve as a predictive tool in clinical settings. Improvements and enhanced usability are necessary to boost the robustness and adaptability of the model, irrespective of its current performance. Moving forward, upcoming research should concentrate on various crucial aspects. Initially, integrating essential features and data can enhance the predictive model's accuracy by capturing intricate processes that impact liver disease. Furthermore, conducting external validation studies involving independent data will showcase the model's efficacy across diverse cultures and settings, thereby bolstering its credibility and applicability in practical scenarios. Additionally, efforts must be undertaken to enhance the interpretation and clarity of predictive models, as it is crucial to enable clinical decision-making and provide confidence in search engine tools.

Explaining the logic behind predictive models and identifying the key features that enable predictions that help clinicians to better understand and interpret the resulting models. Overall, the results of this study demonstrate the potential of machine learning techniques to improve early detection and management of cirrhosis. Using predictive analytics, doctors can identify high-risk individuals and initiate timely interventions to slow disease progression and improve patient outcomes. The progression of machine learning will enhance healthcare by integrating predictive models into healthcare environments., ultimately enhancing personalized medicine for managing liver diseases.

REFERENCE

1. Huang, D. Q., Terrault, N. A., Tacke, F., Gluud, L. L., Arrese, M., Bugianesi, E., & Loomba, R. (2023). Global epidemiology of cirrhosis—aetiology, trends and predictions. *Nature Reviews Gastroenterology & Hepatology*, *20*(6), 388-398.
2. Dritsas, E., & Trigka, M. (2023). Supervised machine learning models for liver disease risk prediction. *Computers*, *12*(1), 19.
3. Feng, S., Tu, H., & Chen, L. (2023). A novel model predicted liver cirrhosis constructed by ultrasound and serological in autoimmune liver hepatitis. *Medicine*, *102*(38), e35295.
4. Trinh, B., Rinnov, A. R., Iepsen, U. W., Munch, G. W., Winding, K. M., Lauridsen, C., ... & Ellingsgaard, H. (2024). Glucose turnover at whole-body and skeletal muscle level in response to parenteral nutrition in male patients with alcoholic liver cirrhosis. *Clinical Nutrition ESPEN*.
5. Kalaitzakis, Z. E., Giahnakis, E., Koutroubakis, I. E., Mouzas, I. A., & Kalaitzakis, E. (2024). Bariatric nutritional intervention in obese patients with compensated liver cirrhosis: A four-year prospective study. *Digestive Diseases and Sciences*, 1-12.
6. Langer, M. M., Sichelschmidt, S., Bauschen, A., Bornemann, L., Guckenbiehl, S., Gunzer, M., & Lange, C. M. (2023). Pathological neutrophil migration predicts adverse outcomes in hospitalized patients with liver cirrhosis. *Liver International*, *43*(4), 896-905.
7. Lu, G., Zhang, Y., Ren, Y., Shi, J. S., Xu, Z. H., & Geng, Y. (2023). Diversity and comparison of intestinal *Desulfovibrio* in patients with liver cirrhosis and healthy people. *Microorganisms*, *11*(2), 276.
8. Lan, Y., Wang, H., Weng, H., Xu, X., Yu, X., Tu, H., ... & Sheng, J. (2023). The burden of liver cirrhosis and underlying etiologies: Results from

***Corresponding Author:** S Harini

the Global Burden of Disease Study 2019. *Hepatology Communications*, *7*(2), e0026.

9. Ullah, M. I., Alameen, A. A. M., Al-Oanzi, Z. H., Eltayeb, L. B., Atif, M., Munir, M. U., & Ejaz, H. (2023). Biological role of zinc in liver cirrhosis: An updated review. *Biomedicines, 11*(4), 1094.
10. Emenena, I., Emenena, B., Kweki, A. G., Aiwuyo, H. O., Osarenkhoe, J. O., Iloeje, U. N., ... & Casimir, O. E. (2023). Model for end stage liver disease (MELD) score: A tool for prognosis and prediction of mortality in patients with decompensated liver cirrhosis. *Cureus, 15*(5).
11. Swamy, S. R., & KS, N. P. (2022, December). Hybrid machine learning model for early discovery and prediction of polycystic ovary syndrome. In 2022 Second International Conference on Advanced Technologies in Intelligent Control, Environment, Computing & Communication Engineering (ICATIECE) (pp. 1-8). IEEE.
12. Archana, R., Vaishnavi, C., Priyanka, D. S., Gunaki, S., Swamy, S. R., & Honnavalli, P. B. (2022, May). Remote health monitoring using IoT and edge computing. In 2022 International Conference on IoT and Blockchain Technology (ICIBT) (pp. 1-6). IEEE.
13. Swamy, S. R., Prasad, K. N., & Tripathi, P. (2020, October). Smart home lighting system. In 2020 International Conference on Smart Innovations in Design, Environment, Management, Planning and Computing (ICSIDEMPC) (pp. 75-81). IEEE.
14. Hukkeri, S., Malage, R. V., Swamy, S. R., & Honnavali, P. B. (2021). Estimation of engagement of learners in MOOCs using smart visual processing.
15. Singh, N., Shyam, A., Swamy, S. R., & Honnavalli, P. B. (2021). Differential privacy in NoSQL systems. In Data Science and Security: Proceedings of IDSCS 2021 (pp. 374-384). Springer Singapore.

***Corresponding Author:** S Harini