**PREDICTION ON PROGRESSION AND SURVIVAL**

**RATE OF LIVER CIRRHOSIS BY USING**

**MACHINE LEARNING**

M Pranathi  
*dept. Artificial Intelligence and Machine Learning  
(of Affiliation)*  
*Saveetha Engineering College   
(of Affiliation)*Chennai, India  
srikalamavilla@gmail.com

Varsha G  
*dept. Artificial Intelligence and Data Science  
(of Affiliation)*  
 *Saveetha Engineering College  
(of Affiliation)*Chennai, India

varshagym123@gmail.com

V Sreeja  
*dept. Artificial Intelligence and Machine Learning   
(of Affiliation)*  
*Saveetha Engineering College  
(of Affiliation)*Chennai, India  
yghzinb07@gmail.com

*Abstract*— cirrhosis is a progressive liver disease characterized by replacement of healthy liver by scar tissue, causing significant morbidity and mortality worldwide The aim of this review is to define cirrhosis through clinical data, on population and longitudinal laboratory outcomes. Associated predictive models and survival rates are examined Using a database of patients diagnosed with liver disease, we applied statistical machine learning techniques to develop a predictive model to assess patient outcomes Age , sex, causes of liver disease, Factors such as Child-Pugh score and laboratory markers (e.g., bilirubin, albumin, INR) were analyzed to determine their association with survival rate Our study shows that early recognition and intervention significantly impact survival outcomes.

Keywords—Cirrhosis, Liver Disease, Predictive Models, Survival Rate, Child-Pugh Score, Machine Learning, Longitudinal Analysis, Laboratory Markers, Morbidity, Mortality, Early Recognition, Intervention, Bilirubin, Albumin, INR (International Normalized Ratio).

1. INTRODUCTION

Liver cirrhosis is a chronic and progressive disease of the liver characterized by the replacement of healthy liver tissue with scar tissue, resulting in impaired liver function and structure This is a global health problem a it is important, with an increased incidence of liver diseases, especially hepatitis BC. Due to infection, excessive alcohol consumption, and rising rates of non-fatty liver disease (NAFLD), cirrhosis of the liver is responsible for approximately 2 million deaths per year, according to the World Health Organization (WHO). of it, leading to better understanding, early detection , and effective management. The clinical course of liver disease can be complex and variable, requiring accurate prediction of patient care. Evaluating survival rates and predicting mortality in patients with liver disease is important to guide treatment decisions, including timing of liver transplantation Traditional scoring systems, such as Child Turcotte-Pugh (CTP) score and the Model of End-Stage Liver Disease (MELD) score , They have also been widely used to predict outcome but these models may not capture all relevant clinical variables, prompting a search for statistical and advanced machine learning techniques. The aim of this project is to investigate prognostic models and survival rates in liver disease by analyzing clinical data, demographic data, and laboratory data Using data-driven methods, we seek to identify significant prognostic factors affecting survival outcomes and efficacy -We strive to develop predictive models that can help healthcare professionals make clinical decisions

# **RELATED WORK**

## Cirrhosis is a chronic, progressive disease of the liver marked by fibrous ulcers replacing healthy liver tissue, leading to impaired liver function and increasing the risk of complications Thus this situation represents a major global health challenge, contributing to significant morbidity and mortality worldwide. Although traditional prognostic models, such as the Child Pugh score and end-stage liver disease (MELD) score, have been widely used to predict patient outcomes, they generally fail to capture clinical variables difficult all have been done disease progression, but these models are so often hampered by a small number of cases, specific populations, or a single clinical marker, which prevents their generalizability and real-time processing method.Existing research highlights the importance of a dynamic, data-based approach to better estimate survival rates and disease progression in patients with cirrhosis Machine learning algorithms, when applied in a variety of clinical and demographic contexts, demonstrated the ability to identify significant prognostic factors and improve outcome prediction There is however, still a need for more comprehensive models that affect clinical synthesize long-term data and provide consistent updates based on changing patient conditions.The Cirrhosis Care Companion project aims to address this gap by applying advanced statistical and machine learning techniques to develop robust models to predict liver disease outcomes. By incorporating several variables—age, sex, etiology of liver disease, Child Pugh score, and laboratory markers such as bilirubin, albumin, and INR.

## **PROBLEM STATEMENT**

Hepatitis is a malignant, life-threatening, progressive form of liver failure, which can lead to complications such as portal hypertension, liver failure, hepatocellular carcinoma, and more results The unpredictable progression of the disease makes it difficult for clinicians to accurately predict patient outcome, especially in the progressive stage -Despite scoring systems such as Turcotte-Pugh (CTP) and liver disease (MELD) score models exist, these models often fail to account for all relevant clinical factors, resulting in potential prognostic errors in survival rates and treatment outcomes.

The current healthcare system faces major challenges in optimizing the care of cirrhotic patients, as heterogeneity in the pathogenesis and progression of the disease complicates prognosis Many prognostic models rely on static clinical variables on a large scale, failing to adjust dynamically to changes in patient condition.

In addition, factors such as age, comorbidities, and other complications are often underrepresented in traditional models, reducing their effectiveness in personalized treatment. Therefore, there is an urgent need to improve the accuracy of predictive models to better estimate survival rates in liver disease Use modern data analysis techniques, such as machine learning, and clinical demographic variables types together to create more powerful prognostic tools that provide clinicians with deeper insights into disease progression for predictions of survival and enable healthcare professionals to make informed treatment and independent resource decisions will be involved

**IV. SYSTEM ARCHITECTURE AND DESIGN**

The systematic design of the classification task starts with importing data from various sources such as databases, files, and APIs, using Python for data transformation and then preprocessing the data to deal with missing values, . encodes categorical variables, and normalizes features, resulting in cleanup of an embedded dataset. This data set is divided into training and validation sets using separator class, including stratified K-Fold cross validation to ensure robust analysis. Class weights are calculated to address any imbalances in the data set. The next section involves specifying classifiers (such as XGBoost, LightGBM, and CatBoost) with appropriate hyperparameters. These models are trained on a training dataset, and their performance is evaluated using metrics such as accuracy and F1-score. Model evaluation follows, resulting in reporting and confusion matrices to evaluate performance. They choose models that work best for applications, including developing APIs or web interfaces to serve the models and using monitoring tools to monitor performance If you want to create user interfaces for user interaction access, such as sending data and viewing results. Overall, this framework emphasizes flexibility, scalability, and security by ensuring a systematic approach to the creation, training, and implementation of the classification model

**V.PROPOSED METHODOLOGY**

The proposed approach to predicting liver disease progression and survival rates using machine learning consists of several steps, aimed at developing robust models for early detection and prediction The process begins with information a collected on, treat relevant patient information including medical history, clinical tests and liver function parameters -will be obtained from publicly available records or data-sets These data will be pre-processed to assess standards missingness control, normalize features, and encode categorical variables such as demographics and disease stages Exploratory data analysis (EDA) is conducted to identify patterns and significant associations between various factors and outcomes of liver disease.

Once the data is generated, it will be divided into training and testing phases. In model training, along with neural networks, ensemble methods such as Random Forest, Gradient Boosting (XGBoost, LightGBM), and support vector machines (SVM) will be explored Feature engineering techniques will be used to generate new meaningful features which can improve the prediction power of the model. Layered K-fold cross-validation will be used to ensure model robustness and prevent overfitting.

Class balance methods such as SMOTE (Synthetic Minority Over-sampling Technique) or square-weight adjustment will be used to overcome the imbalances common in medical data sets for model analysis. Once the optimal model is identified, it will be optimized using hyperparameter optimization methods such as Grid Search or Random Search.

Once model evaluation is complete, the high-performance model will be used in a clinical scenario to be integrated into an application for physicians to use to predict patient outcomes. A key component of the approach is model interpretation capability, using tools such as SHAP (SHapley Additive description) to provide insight into the factors that contribute most to the predictions, ensuring model clarity and reliability so for physicians

**VI. IMPLEMENTATION AND RESULTS**

The project, Predicting Liver Cirrhosis Progression Survival Rate Using Machine Learning, was developed through a series of well-defined steps, which were used to ensure model accuracy and calibration. Key factors such as demographic data, liver function tests, disease stage, and other relevant biomarkers were extracted to serve as input for model training The data were preprocessed to address missing values, for continuous variables were normalized, and used to record categorical variables.

After preprocessing, several machine learning algorithms such as Random Forest, Gradient Boosting (XGBoost and LightGBM), and CatBoost were applied. These models were trained with processed data, and class weighting, SMOTE and other class balance methods were used to address imbalances in the dataset, ensuring that the model did not favor most classes in the prophecy of salvation. The dataset was divided into training and testing sessions using stratified K-fold cross-validation to ensure consistency and robustness in the performance evaluation of the model

In model evaluation, various metrics including accuracy, precision, recall, F1-score and AUC-ROC were used to evaluate the ability of the model to predict progression and survival rates of liver disease Hyperparameter tuning was performed using grid search and random search, further optimizing the models and improving their predictive capabilities.

The final results indicated that the models could predict disease progression and patient survival probability well with high accuracy. For interpretation, SHAP (SHapley Additive exPlanations) values ​​were developed to reveal high-impact prognostic factors, such as bilirubin levels, age, and albumin levels

Overall, the results successfully demonstrated the ability of machine learning to predict liver disease progression and survival, providing a valuable tool for clinicians to help deliver early intervention and personalized treatment strategies.

**VII .CONCLUSION AND FUTURE WORK**

In conclusion, a device-based approach to predicting the progression and survival of patients with liver disease has shown potential to aid in clinical decision-making When important findings are available clinical resources, this system can help healthcare providers better understand risk factors in individual patients, enabling more personalized treatment plans and early interventions that can be available to patients life expectancy has improved.

However, there are still areas for improvement and expansion in the future. Including larger and more diverse data sets can increase the generalizability of the model, allowing it to be applied more broadly to different population groups. In addition, the integration of other medical data such as genetic markers, lifestyle factors, and imaging data will provide a comprehensive view of patient health and further improve the accuracy of prognosis. Additionally, an easy-to-use, real-time clinical tool has been developed that uses these predictions to continuously monitor patients, providing immediate benefits to healthcare systems. Interpretive approaches such as SHAP or LIME should be explored to improve prediction interpretation, ensure that physicians can rely on and understand the resulting models Future work could also include collaboration with physicians to provide such predictions this tool has been added to the existing healthcare workflow.

**ACKNOWLEDGMENT**

The authors would like to express their sincere gratitude to Saveetha Engineering College, whose support and resources made the development of this project possible. Special thanks to Mr. VASANTH KUMAR C.H sir, Dr. JAYANTHI G mam for their invaluable guidance and constructive feedback throughout the research process. The authors also extend their appreciation to the AI&ML department for providing technical assistance and encouragement. Lastly, we are grateful to our family and friends for their continuous support and motivation during the completion of this project.

**REFERENCES**

[1] Asrani, S. K., Devarbhavi, H., Eaton, J., & Kamath, P. S. (2019). Burden of liver diseases in the world. Journal of Hepatology, 70(1), 151-171.

[2]Yip, T. C. F., Wong, V. W. S., Tse, Y. K., Chan, H. L. Y., & Wong, G. L. H. (2020). Accuracy of non-invasive liver fibrosis scores for diagnosis of advanced fibrosis and cirrhosis in chronic hepatitis B: A prospective study with internal validation. Alimentary Pharmacology & Therapeutics, 51(9), 985-995.

[3]Malekzadeh, R., Mohamadnejad, M., & Alizadeh, A. H. M. (2020). Liver Cirrhosis: A Review. International Journal of Liver Disease, 10(3), 1-9.

[4]Friedman, S. L. (2010). Mechanisms of hepatic fibrogenesis. Gastroenterology, 134(6), 1655-1669.

[5] Hoang, D. N., & Oh, H. S. (2021). Liver fibrosis progression prediction using machine learning techniques: A comparative study. Journal of Biomedical Informatics, 113, 103627.

+