

Comparative analysis of machine learning models on predicting non-alcoholic fatty liver disease stages in Sri Lanka

Udeshika G.A.N.P.^{1*}, Adeeba S.¹

¹Department of Computing and Information Systems,
Faculty of Computing, Sabaragamuwa University of Sri Lanka

*ganpudeshika@std.appsc.sab.ac.lk

Non-Alcoholic Fatty Liver Disease (NAFLD) is emerging to be a significant problem in the worldwide, including in Sri Lanka, where it has become a major contributor to chronic liver disease. The conventional diagnostic methods, including liver biopsy and high-quality imaging, are precise, although they are still invasive, expensive, and unavailable in the rural and resource-constrained clinical environment. The paper is a comparative analysis of machine learning (ML) models of prediction of the non-invasive stages of NAFLD. For training the ML model, regularly measured biochemical and demographic data of 1,280 Sri Lankan NAFLD patients who had confirmed the stage according to the imaging and biopsy-confirmed diagnoses. Parameters that were used include age, sex, liver enzymes, bilirubin fractions, serum proteins, renal markers, electrolytes, and inflammatory indicators. Clinical labels for the patient's stage were given according to the imaging and biopsy-confirmed diagnoses from expert consultants. The patients were divided into four stages of NAFLD, which included Simple Steatosis, Non-Alcoholic Steatohepatitis (NASH), Fibrosis, and Cirrhosis. Six ML models, like Logistic Regression (LR), KNearest Neighbors (KNN), Random Forest (RF), Gradient Boosting, Light-GBM, and XGBoost, were trained and assessed to find out the most successful predictive algorithm. Gradient Boosting showed the best performance with an accuracy of 91.7% followed by XGBoost with 90.6 indicating that the models have a good predictive ability on the basis of routine laboratory markers. The comparative findings suggest that the staging of noninvasive NAFLD using ML is very practical and clinically significant in the Sri Lankan healthcare setting. This research represents a region-specific, datadriven model for NAFLD stage prediction in Sri Lanka. These findings indicate the potential of ML to support early detection, enhance risk stratification, improve clinical decision-making, and prioritize patients who need immediate treatment, especially in hospital settings with limited access to special diagnostic tools.

Keywords: *Biochemical indicators; Comparative model analysis; NAFLD staging; Non-invasive diagnosis; Sri Lankan patient data*