

Classification of Patients with Suspected Hepatitis C Using Random Forest: Results from a Machine Learning Model

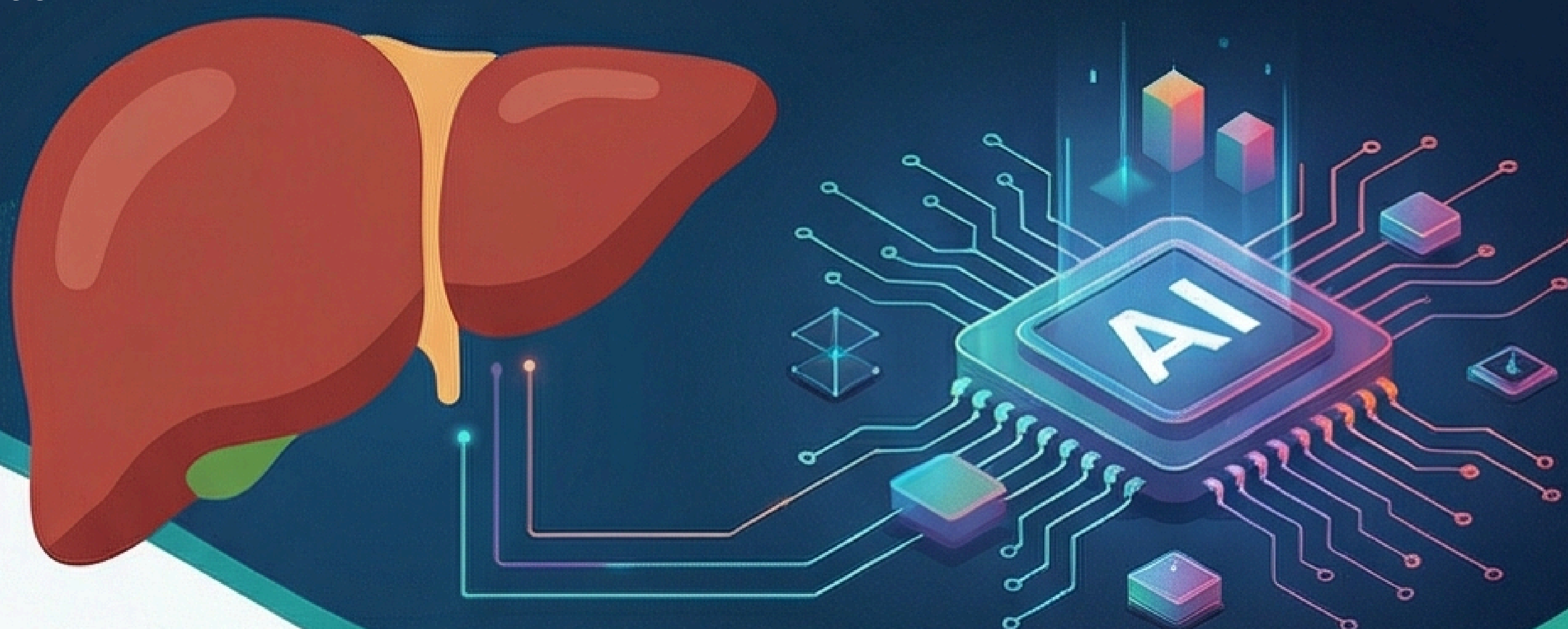


Authors: D.J. BOARDMAN GONZÁLEZ, M.I. MARTIN NIÑO, E. GARCÍA LÓPEZ, B. MARTÍNEZ RUIZ, G. PICAZO SANCHIZ, L. MARTÍNEZ VALDIVIESO.

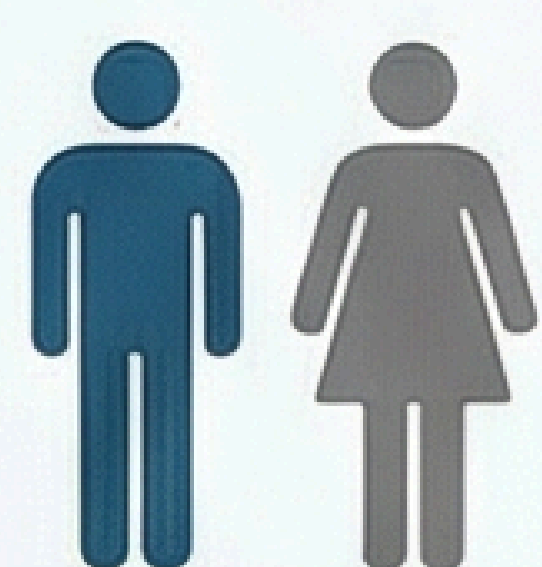
Early diagnosis of Hepatitis C Virus (HCV) is critical to preventing advanced liver diseases. This study evaluates a Random Forest machine learning model applied to a clinical dataset of 589 individuals, analyzing specific liver biochemical parameters to classify patients into blood donors, suspected cases, hepatitis, fibrosis, and cirrhosis categories, achieving high diagnostic accuracy through k-fold cross-validation.

Contact data: djboardmanglez@gmail.com

4CPS-305



METHODOLOGY



DEMOGRAPHIC PROFILE:
MEDIAN AGE 47 YEARS,
MAJORITY (59%) MALE

CLINICAL CLASSIFICATION (PATIENT GROUP):



RESULTS

IDENTIFYING CRITICAL BIOMARKERS:

Out of all variables tested, six specific liver markers showed statistically significant differences ($p < 0.05$) between patient groups.

TOP-TIER SIGNIFICANCE IN CHOLINESTERASE:

Cholinesterase showed the most significant variance among the groups with a p-value of **0.0004**.

BIOMARKER	MEDIAN VALUE	P-VALUE
Cholinesterase	5.45 U/mL	0.0004 ★
Cholesterol	5.4 mmol/L	0.0081
Albumin	42 g/L	0.01
Creatinine	78 μ mol/L	0.011
Total Proteins	71.2 g/L	0.016
GGT	21.3 U/L	0.020

CONCLUSION AND RELEVANCE



This high AUC value indicates the model has an excellent ability to distinguish between the different stages of liver disease.



The model recorded a perfect sensitivity rate, resulting in zero false negatives during the cross-validation process.



The high specificity demonstrates the model's reliability in correctly identifying those without the specific disease stages.