

Hepatitis C Detection from Blood Donor Data Using Hybrid Deep Feature Synthesis and Interpretable Machine Learning.

Author

Safiul Haque Chowdhury

Mohammed Ibrahim Hussain

Mohammed Ibrahim Hussain

Abstract:

Hepatitis is a severe liver inflammation that can lead to chronic disease, liver failure, and death if untreated. It is caused by viral infections, autoimmune disorders, or excessive alcohol consumption, with viral hepatitis (e.g., Hepatitis B and Hepatitis C) being particularly fatal. Early diagnosis and accurate prognosis are crucial for effective treatment. To address this, we employ advanced computational techniques for hepatitis classification and risk assessment, collecting a dataset of 615 samples with 12 biochemical features (e.g., Albumin, Bilirubin, and Cholesterol) obtained from blood donors. We apply the Synthetic Minority Over-sampling Technique (SMOTE) to handle class imbalance and implement Deep Feature Synthesis (DFS) with Aggregation Primitives (AP), Transformation Primitives (TP), and a Hybrid DFS approach to generate three feature-enhanced datasets. Multiple machine learning (ML) models, including Extreme Gradient Boosting (XGB), Random Forest (RF), Gradient Boosting Decision Trees (GBDT), Categorical Boosting (CB), and Adaptive Boosting (AB), are trained with and without DFS. Performance is evaluated using accuracy, precision, recall, F1-score, and specificity, from Confusion Matrix (CM) analysis via 10-fold cross-validation. The GBDT model with Hybrid DFS achieves the highest accuracy of 99.49%, along with 99.56% precision, 99.81% recall, 99.69% F1-score, and 99.12% specificity. To enhance interpretability, we apply Explainable Artificial Intelligence (XAI) techniques, namely Shapley Additive Explanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME), to analyze feature importance and model behaviour. The proposed Hybrid DFS-GBDT approach demonstrates high effectiveness and interpretability, offering a robust framework for hepatitis diagnosis and prognosis.

2025 2nd International Conference on Next-Generation Computing, IoT and Machine Learning, NCIM 2025

Conference Paper2025

EID: 2-s2.0-105017974662

DOI: 10.1109/NCIM65934.2025.1116015

ISBN:979-8-3315-5543-6