UNSUPERVISED MACHINE LEARNING FOR STRATIFYING PATIENT SUBGROUPS IN AUTOIMMUNE LIVER DISEASE

Abstract No. 113

Authors

E. Ben George [1] ; Abraham Varghese [1] ; Gnana Rajesh [2] ; Samuel Giftson [1] ; Raja Al-Hinai [3] ; Amira Al-Nasseri [4] ; C. G. Raji [5] ; Teresa K. George [1] ; Jeba Rosline [1] ; Giri Ramadoss [1] ; Fatima Al-Junaibi [1]

1 - College of Computing and Information Sciences, University of Technology and Applied Sciences - Muscat, Muscat 133, Sultanate of Oman ; 2 - College of Computing and Information Sciences, University of Technology and Applied Sciences, Al Musannah 314, Sultanate of Oman ; 3 - Hepatology Unit, Royal Hospital, Muscat 113, Sultanate of Oman ; 4 - Nutrition and Dietetics Unit, Royal Hospital, Muscat 113, Sultanate of Oman ; 5 - Department of Computer Science, Assumption College Autonomous, Kottayam, Kerala 686101, India

Project Aims & Objectives

Autoimmune liver disease is a complex and heterogeneous condition that challenges effective diagnosis and management. Manual grouping of patients based on disease grade and clinical parameters is a tedious task and makes it challenging to derive meaningful insights from each group. This study aimed to use machine learning based clustering techniques to identify distinct and clinically relevant patient subgroups from the patient data, enhancing the understanding of disease heterogeneity and supporting more personalized care strategies

Materials And Methods

The study analyzed a comprehensive dataset containing clinical, biochemical, immunological, and demographic variables from patients diagnosed with autoimmune liver disease. Eight unsupervised clustering algorithms were applied to the data. The elbow method was used to identify the optimal number of clusters, and clustering performance was evaluated using Silhouette Score, Davies-Bouldin Index, and Calinski-Harabasz Index. A novel Combined Performance Score was used to compare model performance. The features were assessed to determine the most significant ones driving cluster formation.

Charts and Figures

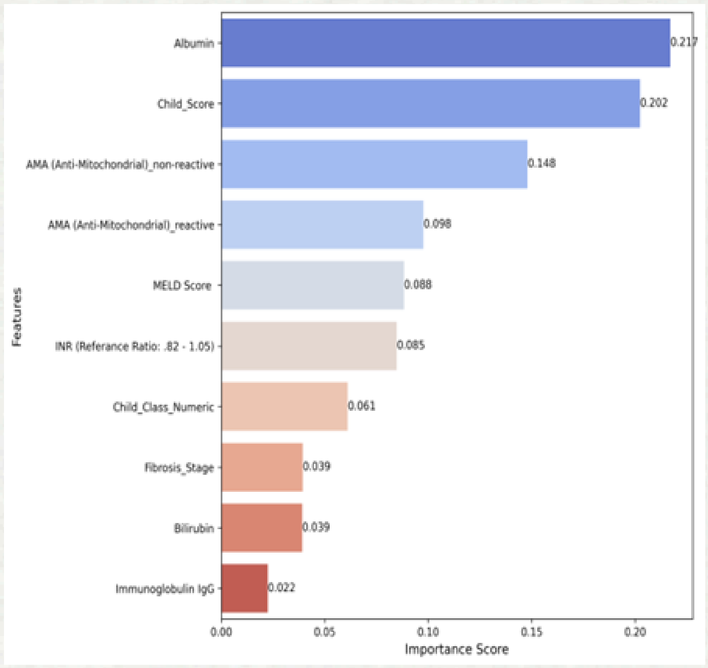

Fig. 1. Feature Importance Plot

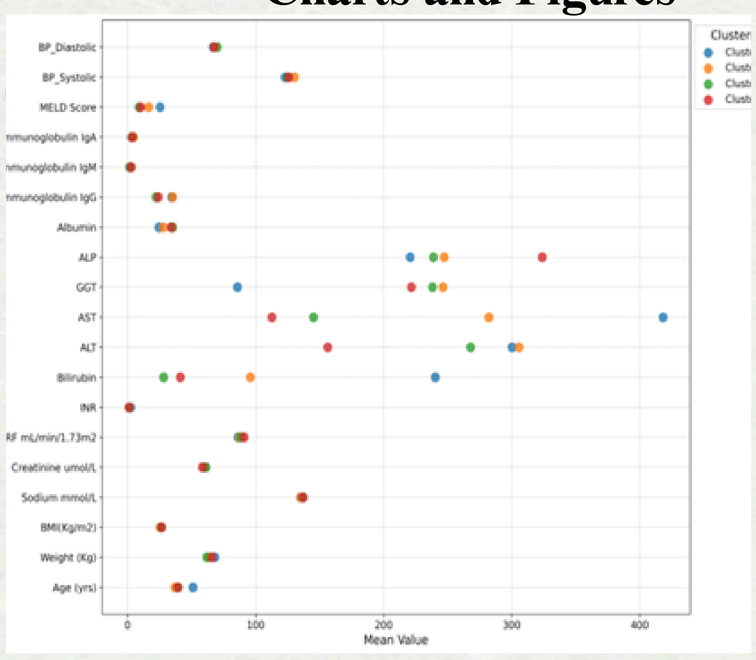

Fig. 2. Distribution of mean values of each cluster

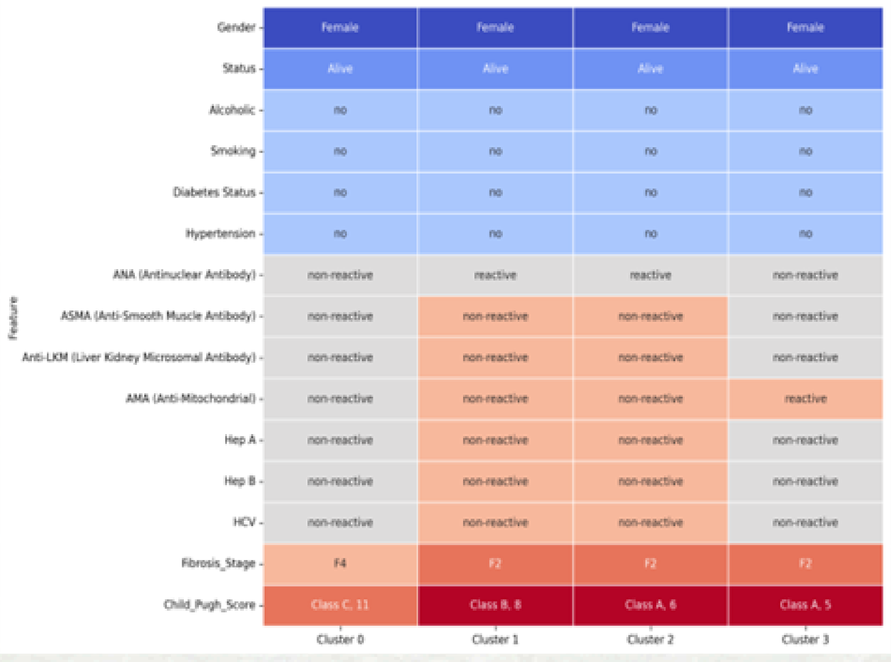

Fig. 3. Categorical feature analysis of clusters

Analysis and Result

KMeans achieved the highest combined performance score of 0.969 among the eight clustering algorithms. The elbow method identified four optimal clusters. Feature importance analysis found albumin, the Child–Pugh score, and key immune markers as crucial parameters relevant to cluster formation. Comprehensive cluster analyses revealed distinct patterns of liver disease progression and severity. The identified clusters revealed a spectrum of disease states, ranging from severe to mild, with intermediate groups showing active immune involvement. These subgroup patterns aligned with varying stages of disease progression and severity

Conslusion

This study demonstrates the use of machine learning based clustering in identifying distinct patient subgroups in autoimmune liver disease. This clustering framework underscores the potential of clustering techniques to cluster patients to enhance the understanding of disease heterogeneity, improving patient outcomes and providing personalized treatments. This work contributes to biomedical research by effectively utilizing machine learning based clustering to uncover autoimmune liver disease patterns and support more accurate predictions.

Acknowledgment

The authors thank the University of Technology and Applied Sciences, Muscat, for supporting this research and Royal Hospital of Oman for providing the data required for this research

Contact email:

ben.george@utas.edu.om