Utilizing the Bat Algorithm-Ensemble Approach for Ulcerative Colitis Detection from Gene Expression Data

Fahri Sunarya[1], Isman Kurniawan[2]

[1,2]Fakultas Informatika, Universitas Telkom, Bandung
[1]fahrisunarya@students.telkomuniversity.ac.id,   [2]ismankrn@telkomuniversity.ac.id

Abstract

Ulcerative Colitis (UC) is a chronic inflammatory bowel disease that affects the colon, resulting in recurrent periods of mucosal inflammation. The incidence and prevalence of UC increased from 29.81 per 100,000 people in 1990 to 39.37 per 100,000 people in 2019, with the highest rates among all continents. Clinical studies have underscored the critical importance of early prevention and detection in patients with UC due to its uncertain etiology. Traditional methods for detecting UC rely on a comprehensive analysis of clinical manifestations, laboratory tests, and findings from endoscopic, histological, and radiological examinations. However, these methods lack sufficient sensitivity or specificity, often presenting ambiguous clinical pictures and a lack of specific laboratory parameters for detecting UC. Therefore, there is a need for a method that can overcome these deficiencies to detect UC accurately. This study aims to enhance the predictive capabilities and diagnostic accuracy for UC using gene expression data by employing a Bat Algorithm combined with three ensemble models, i.e., Random Forest (RF), Adaptive Boosting (AdaBoost), and Extreme Gradient Boosting (XGBoost). In our study results, we found that the performance of the AdaBoost method surpassed the other models, achieving accuracy and F1-score values of 0.80 and 0.85, respectively. The findings suggest that the AdaBoost model, in conjunction with the Bat Algorithm, can significantly enhance the diagnostic accuracy for UC, potentially leading to better early detection and management of the disease.

Keywords : ulcerative colitis, gene expression, ensemble methods, bat algorithm, feature selection, hyperparameter tuning