ABSTRACT

Imbalanced data in machine learning presents a challenge, particularly in binary classification, where one class is significantly underrepresented compared to the other. Conventional classification algorithms typically prioritize the majority class and neglect the minority class, ultimately affecting overall predictive accuracy. Current approaches cannot guarantee achieving a global optimum for all problems. Therefore, this study introduces the Komodo Mlipir Algorithm-based Undersampling (KMAUS) to address this issue. Tested on 100 datasets, KMAUS outperforms baseline methods (such as NS, RUS, TL, NM, and ENN) and benchmark methods (US-GA, US-DE, US-ABC, US-PSO). KMAUS achieved the highest performance in Gmean and BA-Score metrics, with the Random Forest classifier achieving 0.8622 for Gmean and 0.8752 for BA-Score, and the Decision Tree classifier achieving 0.8229 for Gmean and 0.8406 for BA-Score. KMAUS also achieved the highest performance compared to other benchmark methods, with average Gmean values of 0.9534. This approach demonstrates that KMAUS is more effective in maintaining the balance between majority and minority classes, making it suitable for addressing imbalanced data issues, particularly in classification tasks.

Keywords: Komodo Mlipir Algorithm, Undersampling, Imbalanced Data, Evolutionary Computation, Machine Learning