

Enhanced Binary Genetic Algorithm as a Feature Selection to Predict Student Performance

Hamza Turabieh (✉ turabieh@gmail.com)

Taif University <https://orcid.org/0000-0002-8103-563X>

Research Article

Keywords: Genetic algorithm, Feature selection, Students performance, Electromagnetic-like mechanism

Posted Date: May 10th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-172454/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Students' performance prediction systems play a vital role in enhancing the educational performance inside universities, schools, and training centers. Big data can come from different resources such as exam centers, virtual courses, registration departments, e-learning systems, and so on. Extracting meaningful knowledge from educational data is a complex task, so, reducing the data dimensionality is needed. In this paper, we proposed an enhanced binary genetic algorithm (EBGA) as a wrapper feature selection algorithm. Novel hybrid selection mechanism based on a k-means algorithm and Electromagnetic-like mechanism (EM) method is proposed. K-means will cluster the population into a set of clusters, while EM will determine a value called a total force (TF) for each solution. Each cluster has an accumulated total force (ATF) (i.e., adding all TFs together). Selection process will select two solutions with the highest TF from the cluster, which has the highest ATF. We employed a hybrid machine learning approach between the proposed EBGA and five different classifiers (i.e., k-Nearest Neighbors (k-NN), Decision Trees (DT), Naive Bayes (NB), Support Vector Machine (SVM), and Linear Discriminant Analysis (LDA)). Two real case studies obtained from UCI Machine Learning Repository are used in this paper. Obtained results showed the ability of the proposed approach to enhance the performance of the binary genetic algorithm. Moreover, the performances of all classifiers are improved between 1% to 11%.

Full Text

This preprint is available for [download as a PDF](#).

Figures

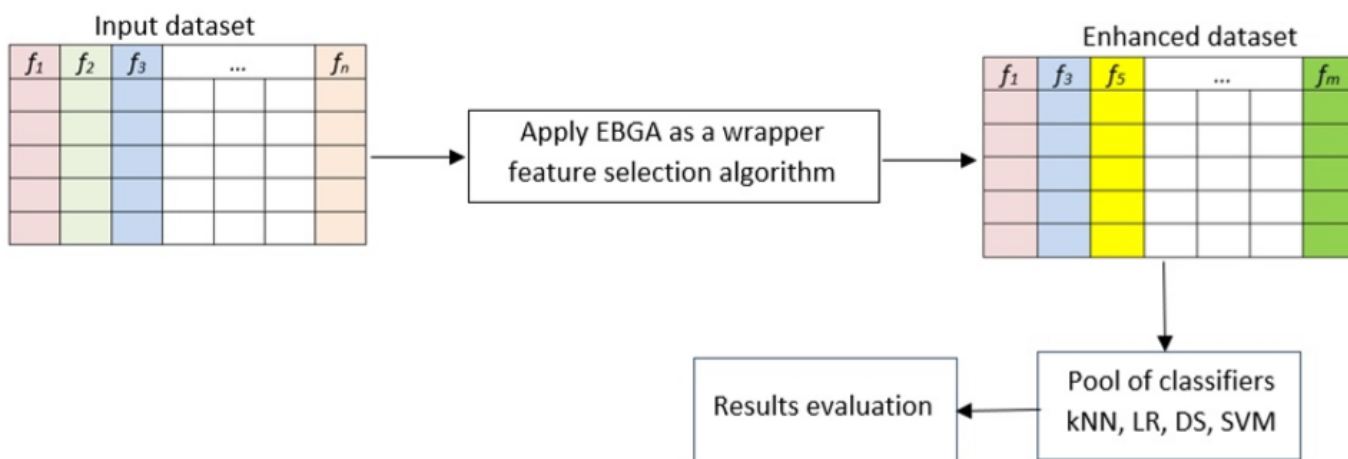


Figure 1

Proposed approach.

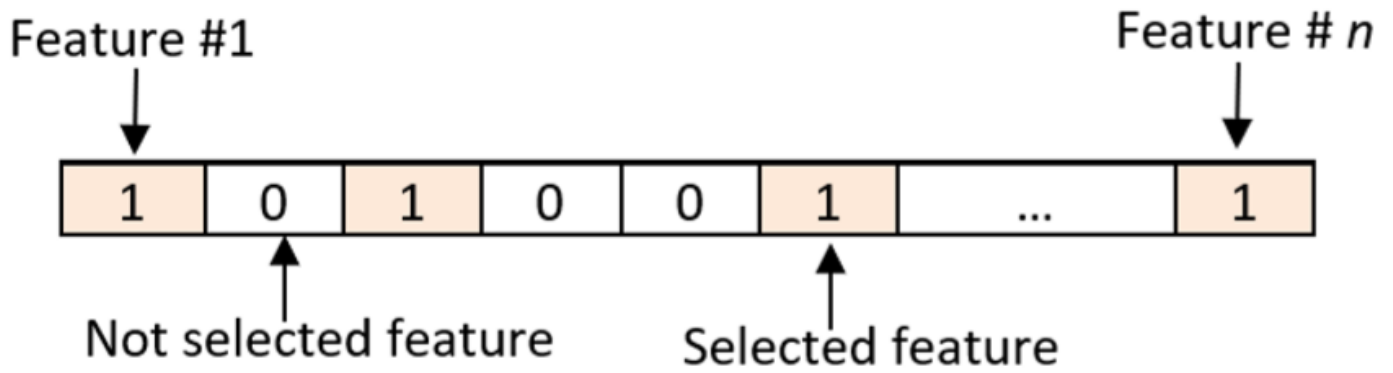


Figure 2

Individual presentation as a binary vector.

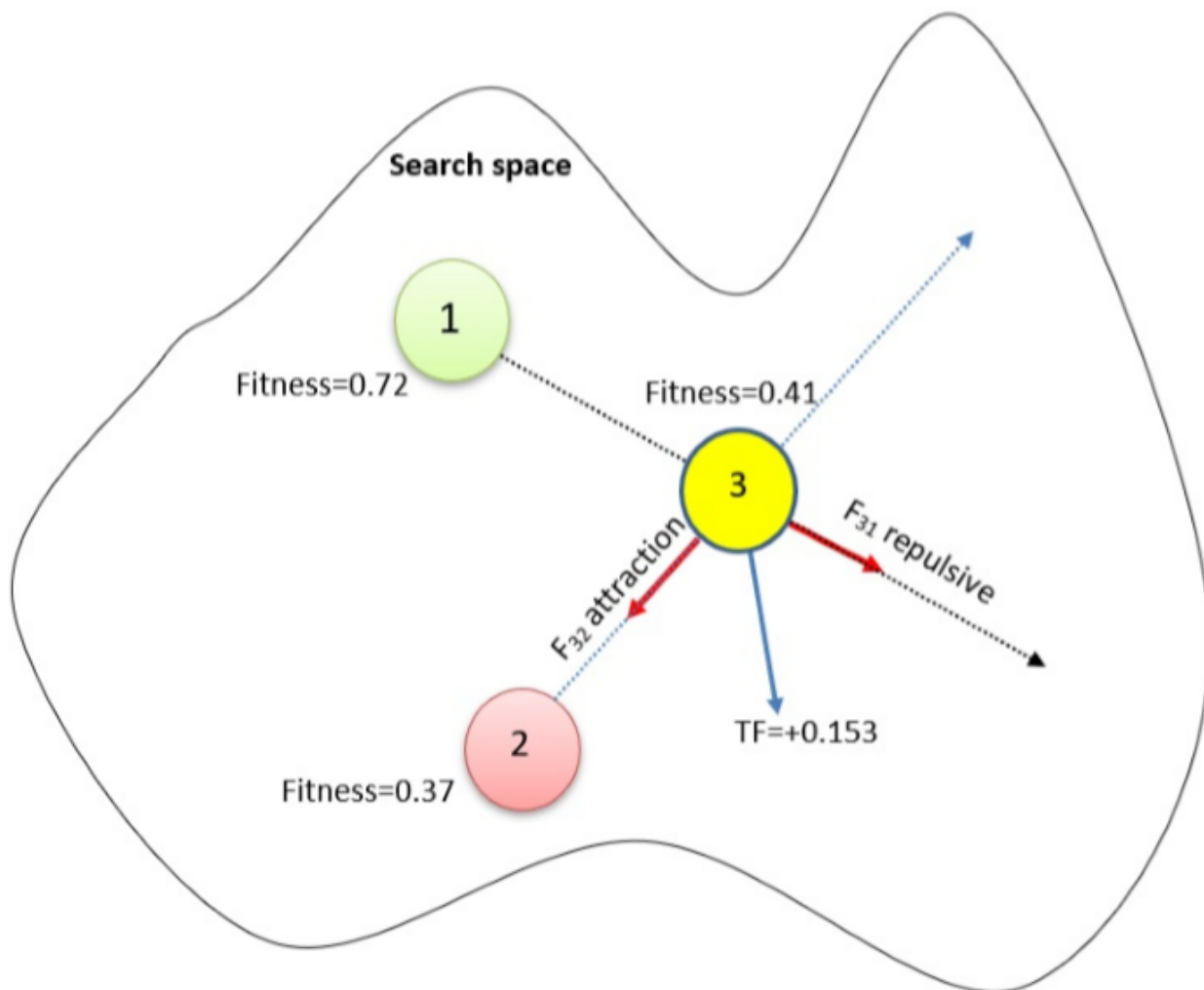


Figure 3

Attraction-repulsion effect on solution 3.

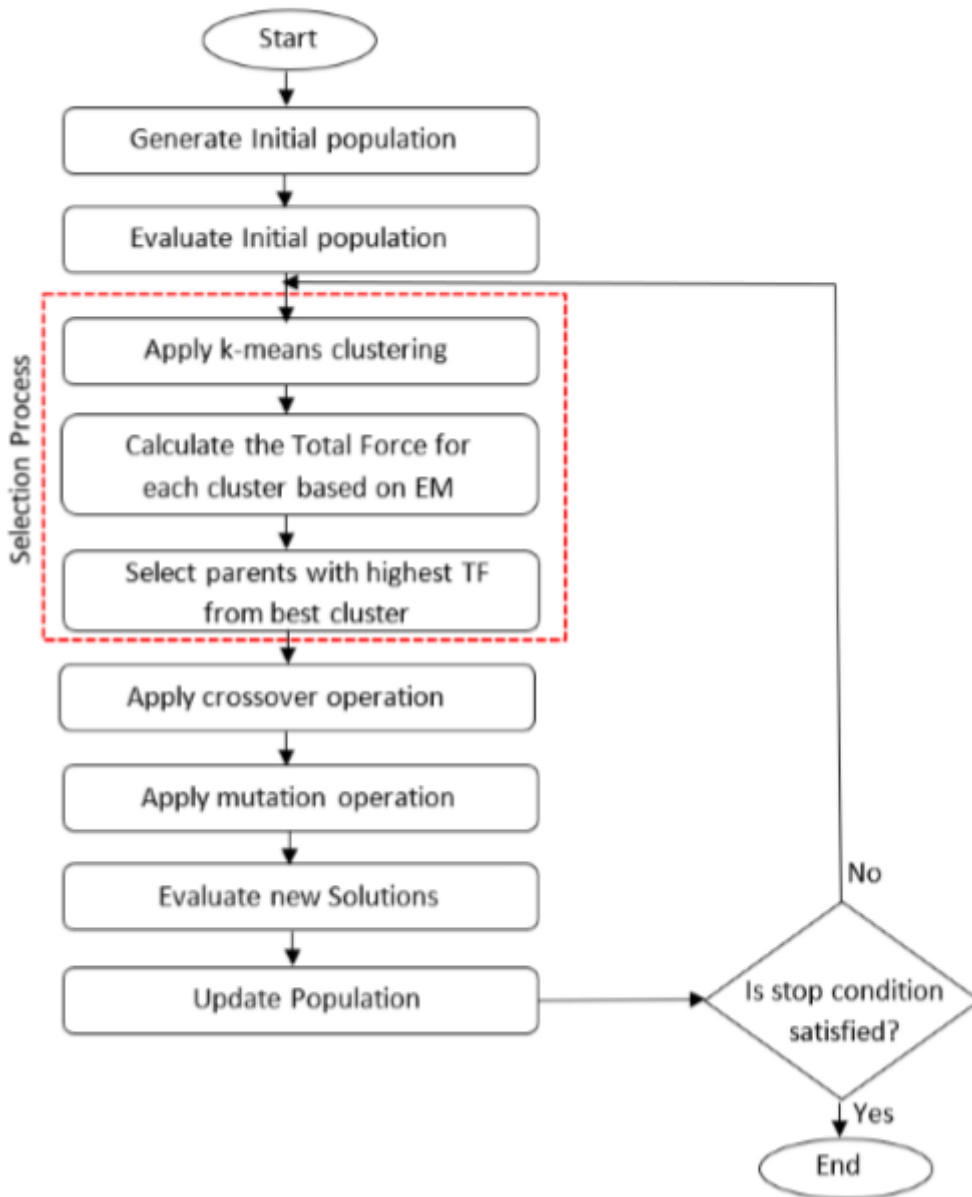


Figure 4

Flow chart of the proposed approach.

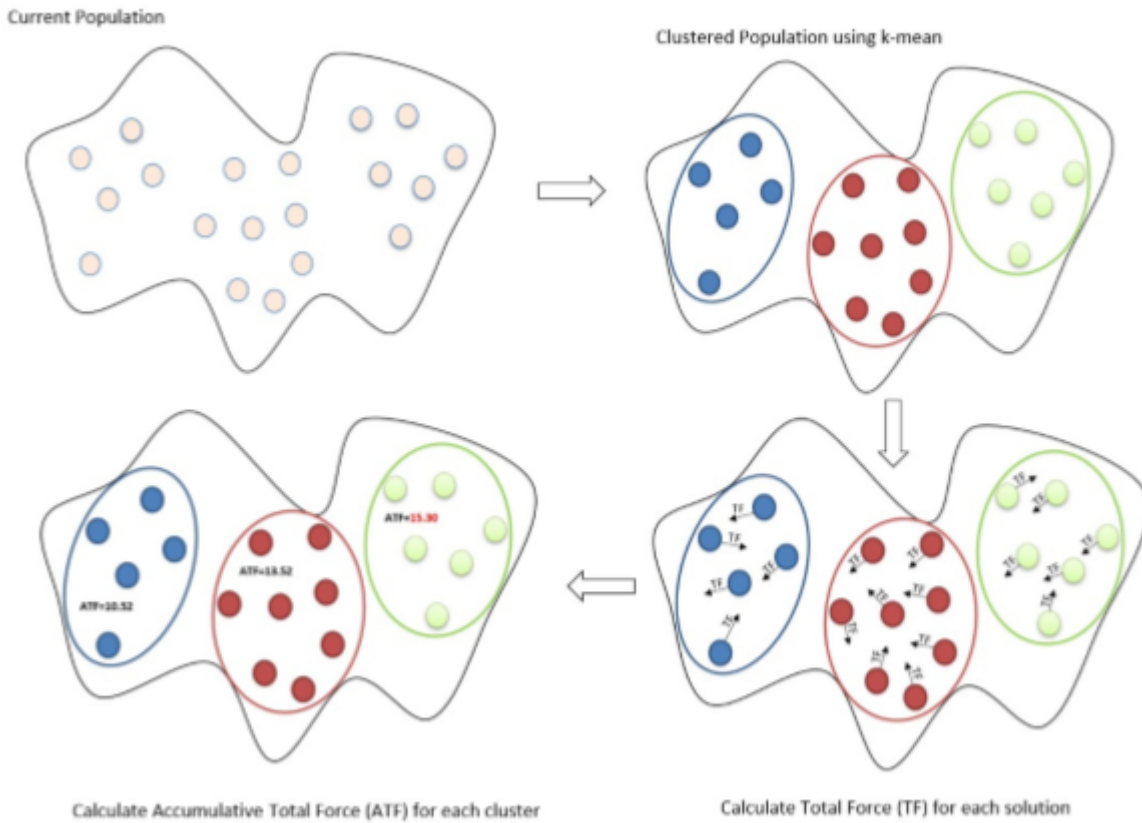


Figure 5

Proposed Selection method.

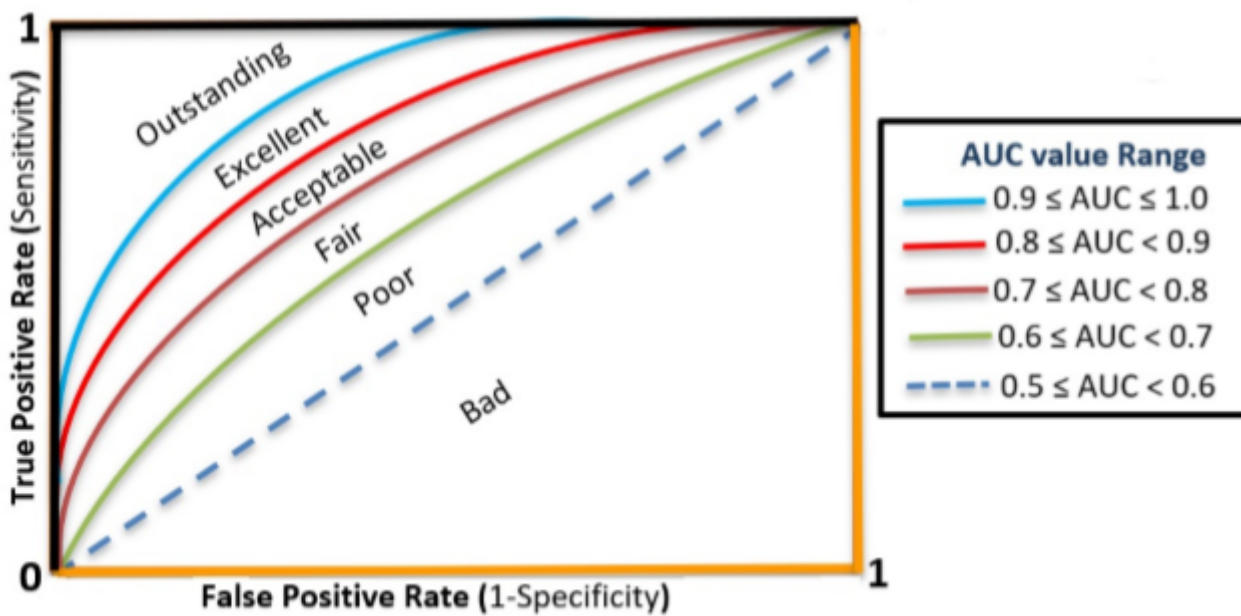


Figure 6

Pictorial diagram for the ROC curves and AUC values.

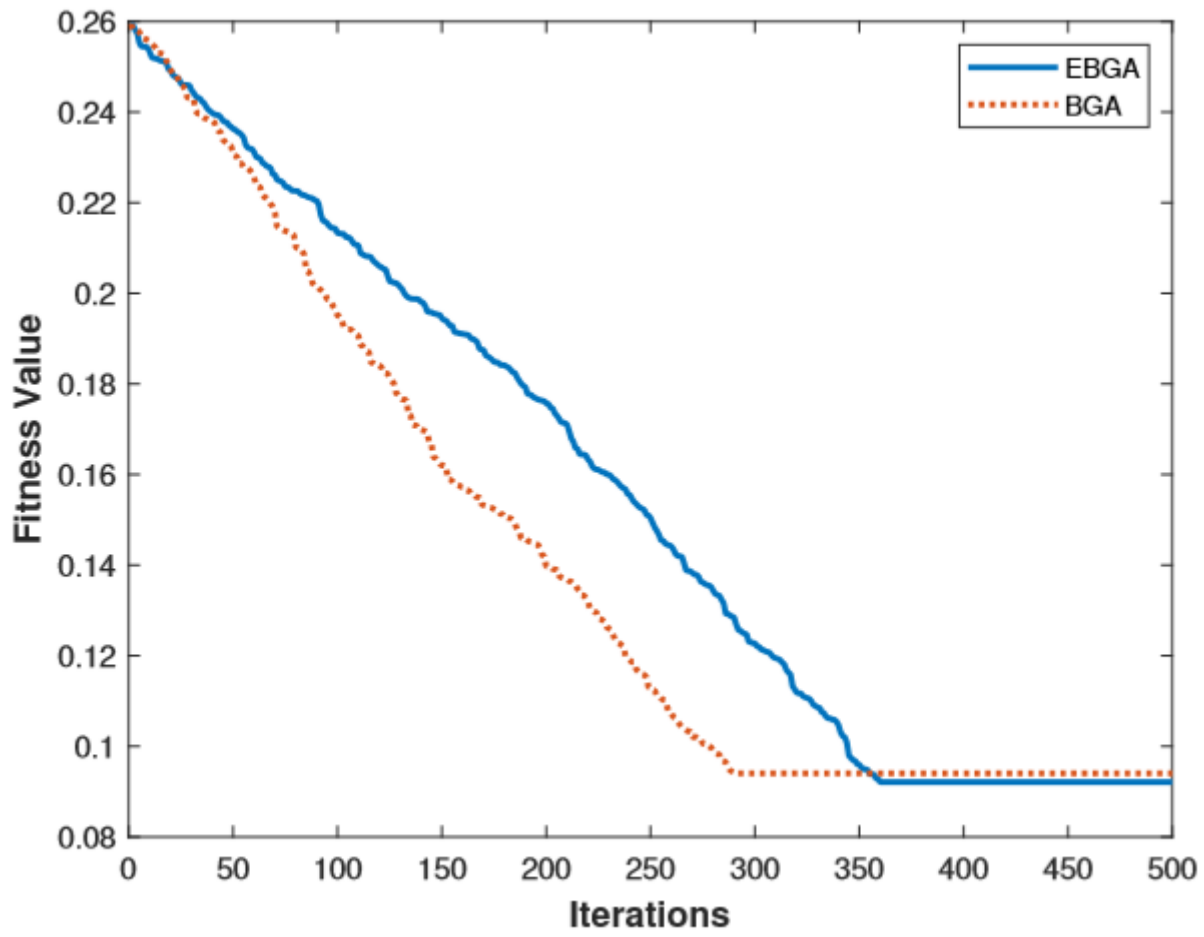


Figure 7

Convergence curve for BGA and EBGA for dataset 1.

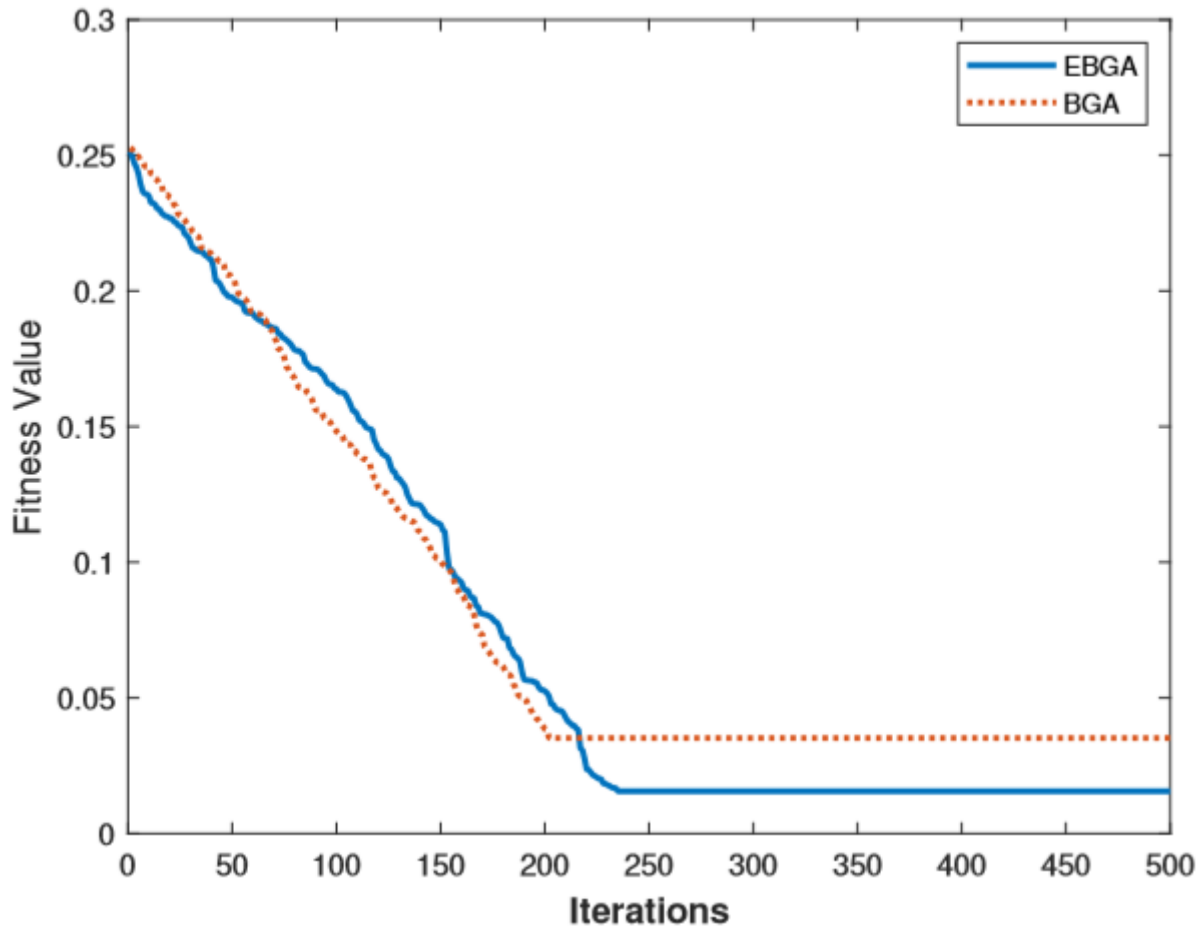


Figure 8

Convergence curve for BGA and EBGA for dataset 2.

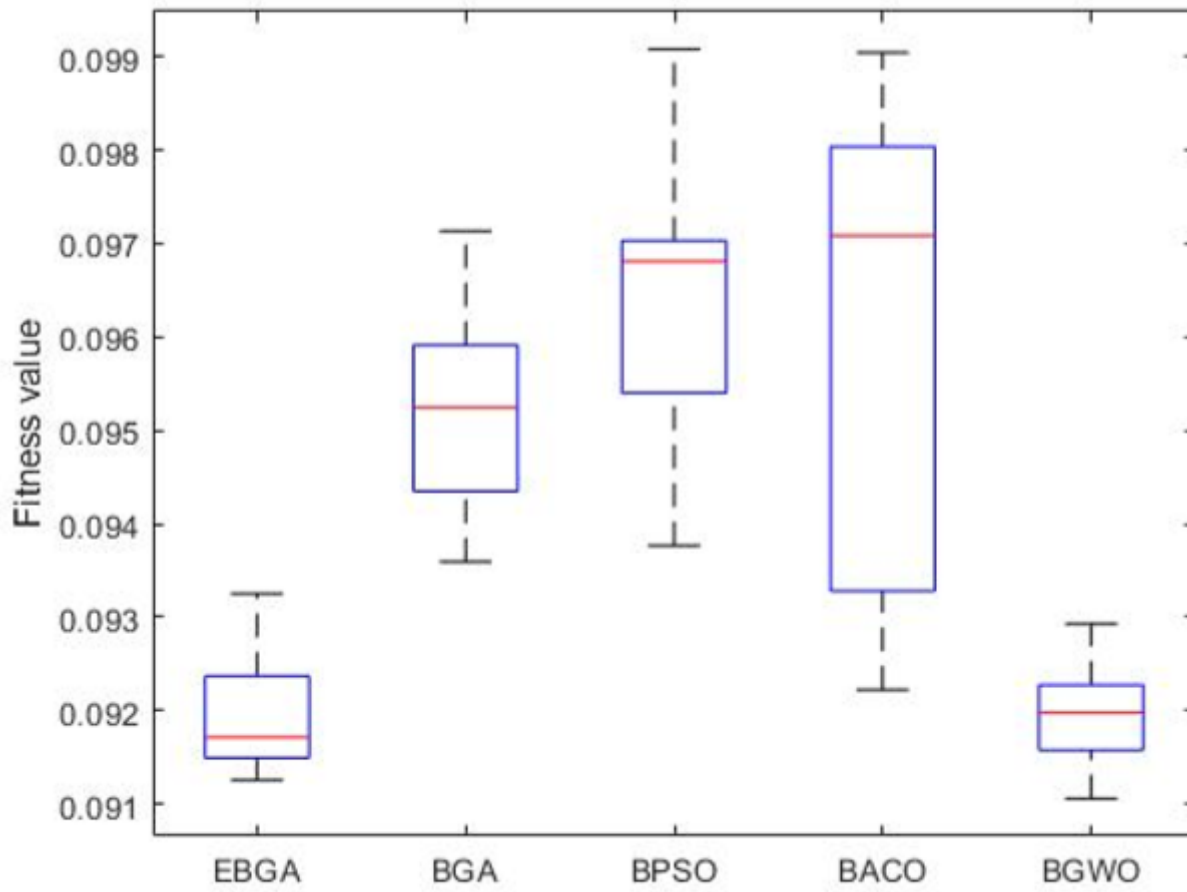


Figure 9

Boxplot diagram for dataset 1 using different feature selection algorithms.

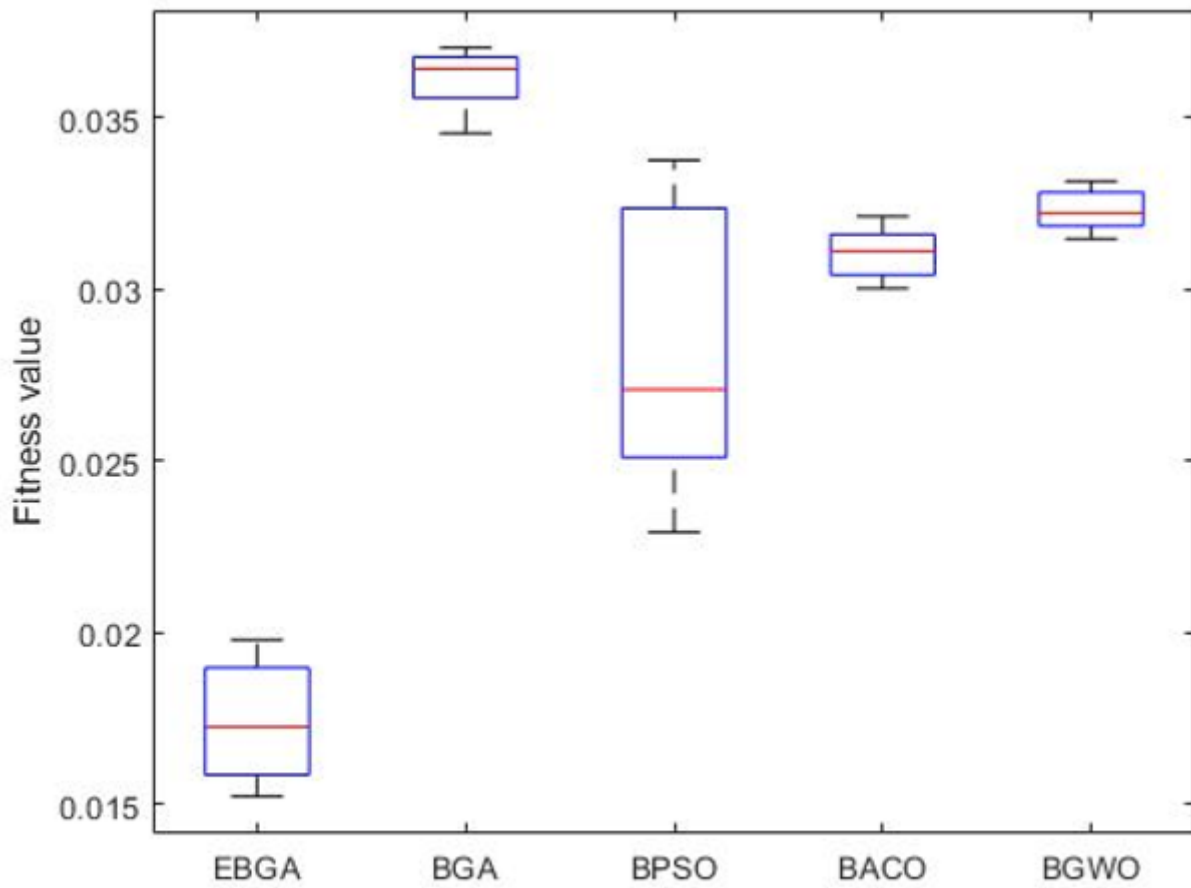


Figure 10

Boxplot diagram for dataset 2 using different feature selection algorithms.

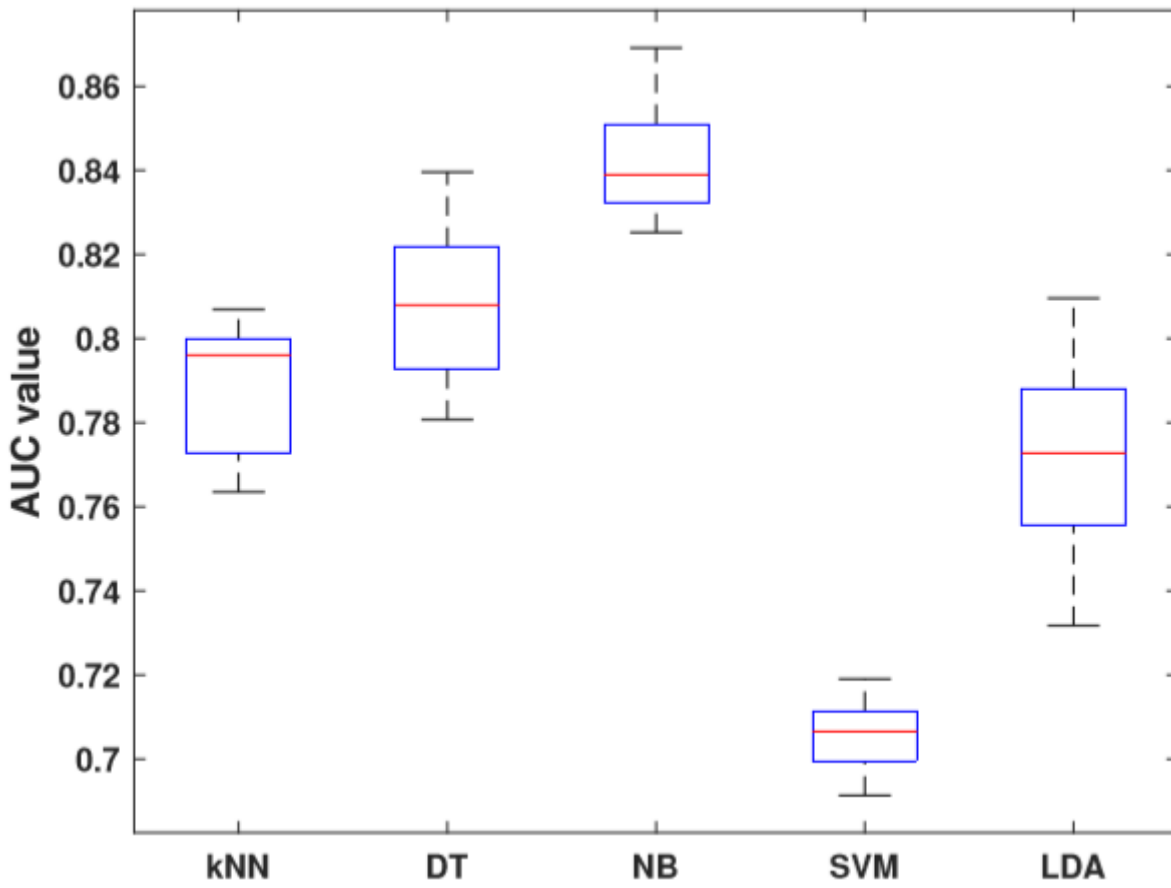


Figure 11

Boxplot diagram for dataset 1 using EBGA with different ML classifiers.

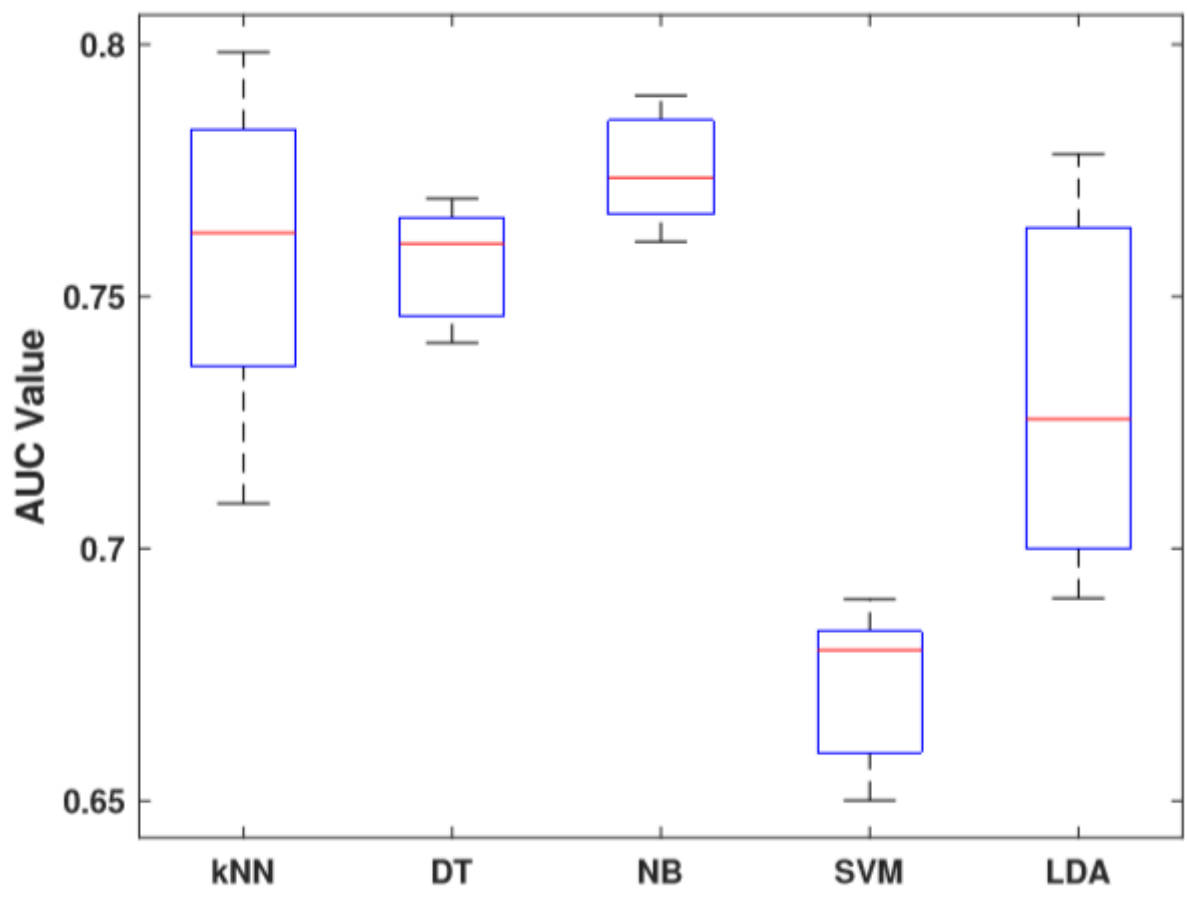


Figure 12

Boxplot diagram for dataset 1 using BGA with different ML classifiers.

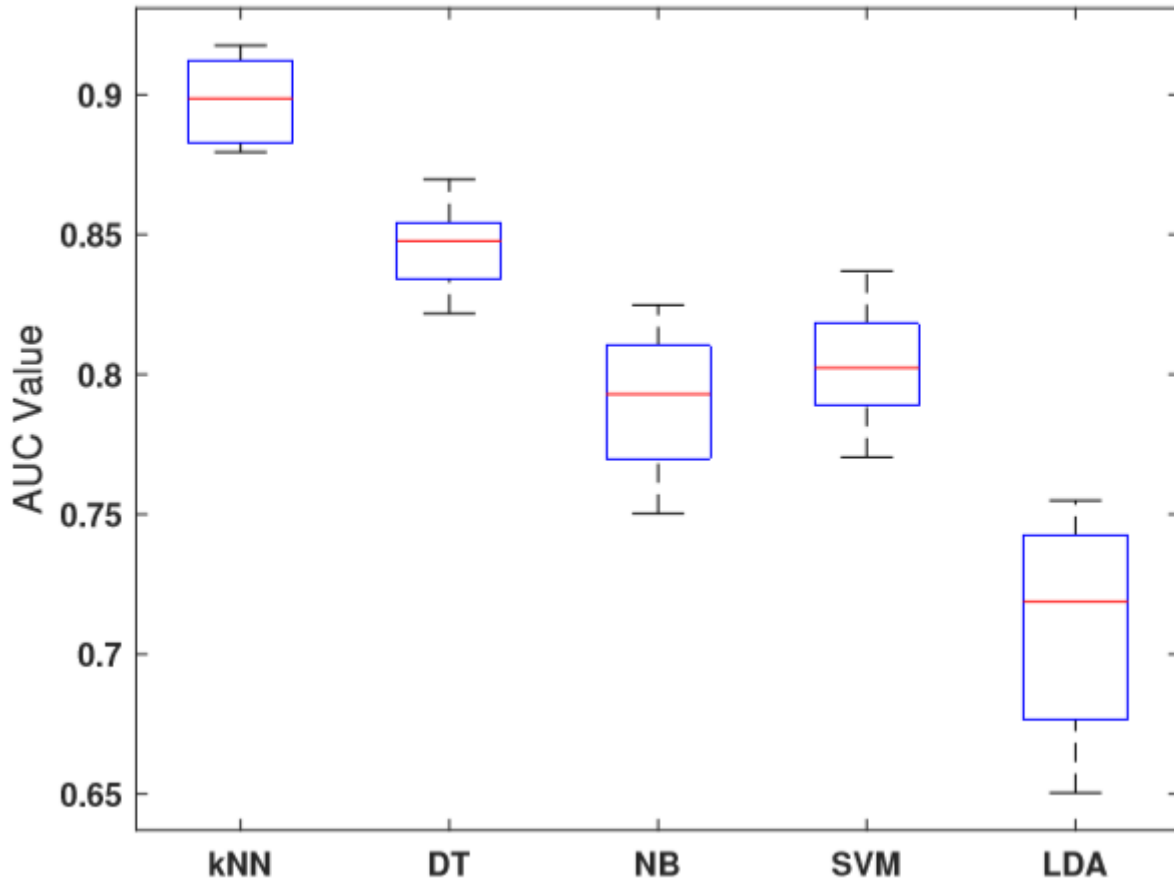


Figure 13

Boxplot diagram for dataset 2 based on selected features using EBGA.

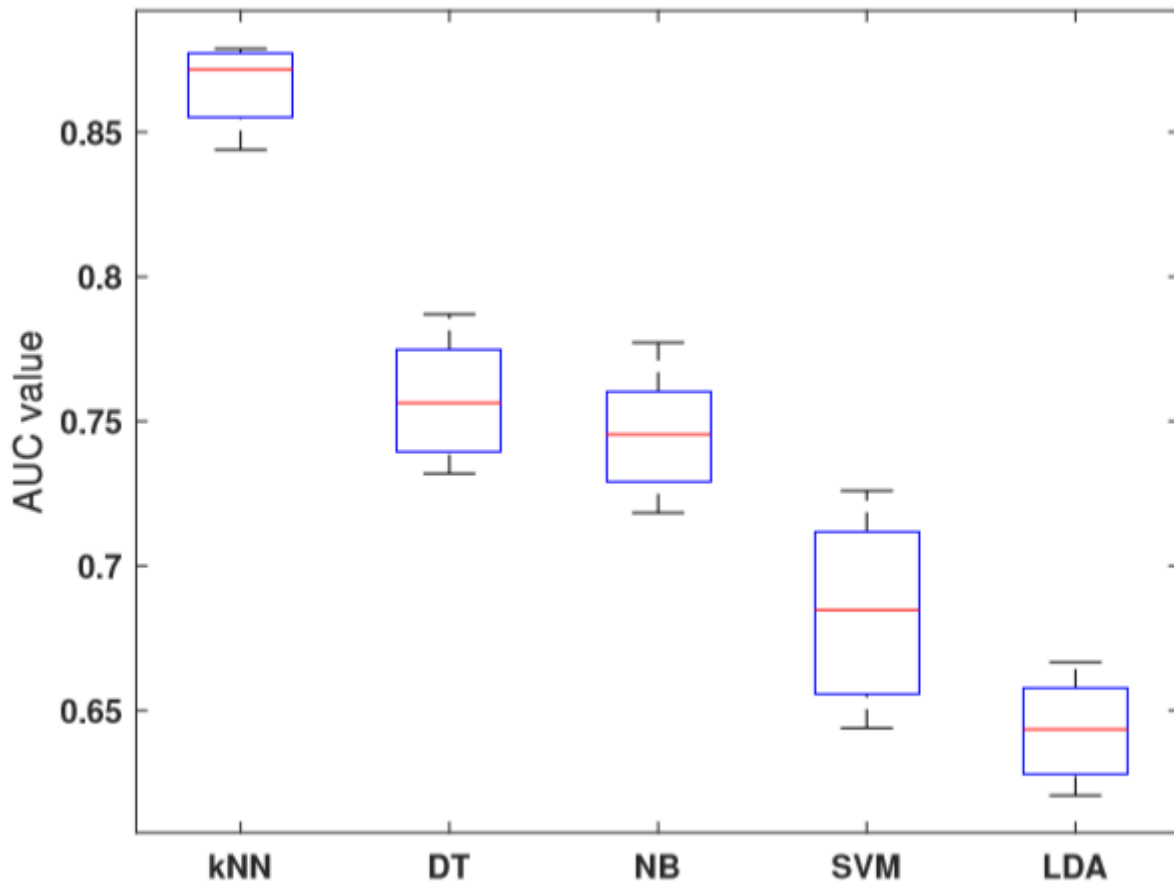


Figure 14

Boxplot diagram for dataset 2 based on selected features using BGA.