EVALUATING VARIOUS FEATURE SELECTION ALGORITHMS FOR ENHANCED PREDICTIVE MODELING

Sricharan Kotala, Dr. Simone Ludwig, Aaron Mackenzie

Abstract

Since the rise of Machine Learning in the late 1980s, it has brought more formalized methods for feature selections with the aim to preprocess data in a different number of ways. There are hundreds of different techniques and methods which can typically be classified underneath three different methods: embedded, filter, and wrapper methods. Specifically, in this paper what will be tested is ANOVA (filter) and Elastic Net (embedded), and Binary Particle Swarm Optimization (PSO). Each has their own advantages and disadvantages in how they preprocess data and it is crucial that we evaluate all of its pros and cons along with the different types of use cases dependent on the computational problem that we are trying to solve. The best way to tell the differences between each and evaluate its performance is by performing highly extensive tests on a large amount of varying dimensional datasets. From there, an evaluation of each feature selection can be done from which will be compared amongst each other.

Introduction

In computational science, handling high dimensional data is crucial for extracting meaningful insights [1]. Feature selection, a machine learning process, identifies the most relevant features for classification, enhancing model performance by reducing data dimensionality, preventing overfitting, and lowering computational costs. Despite numerous feature selection methods since the 1980s, no single method works reliably across all datasets [2], posing a challenge in choosing the best method. This study conducts a comparative analysis of three feature selection methods, evaluating their performance metrics on high-dimensional datasets and using statistical tests like Friedman with Nemenyi to assess differences. By comparing these methods, the research aims to help practitioners select the most effective feature selection methods, thereby improving model performance and contributing to the understanding of feature selection effectiveness in machine learning.

Methodology

The training and results of the tests will be obtained using North Dakota State University's supercomputer, the Center for Computationally Assisted Science and Technology (CCAST). CCAST provides advanced cyberinfrastructure for research and education at NDSU and beyond [4]. Given the computationally intensive tasks of initializing particles and creating subsets of every feature in a dataset, using a powerful machine was necessary. All programs will be run on 16 CPU cores and 2 GPUs. Every algorithm was run on 12 medical orientated datasets, which have been normalized and contain a binary classification. Every dataset has approximately 2,000 rows and 40,000 features.

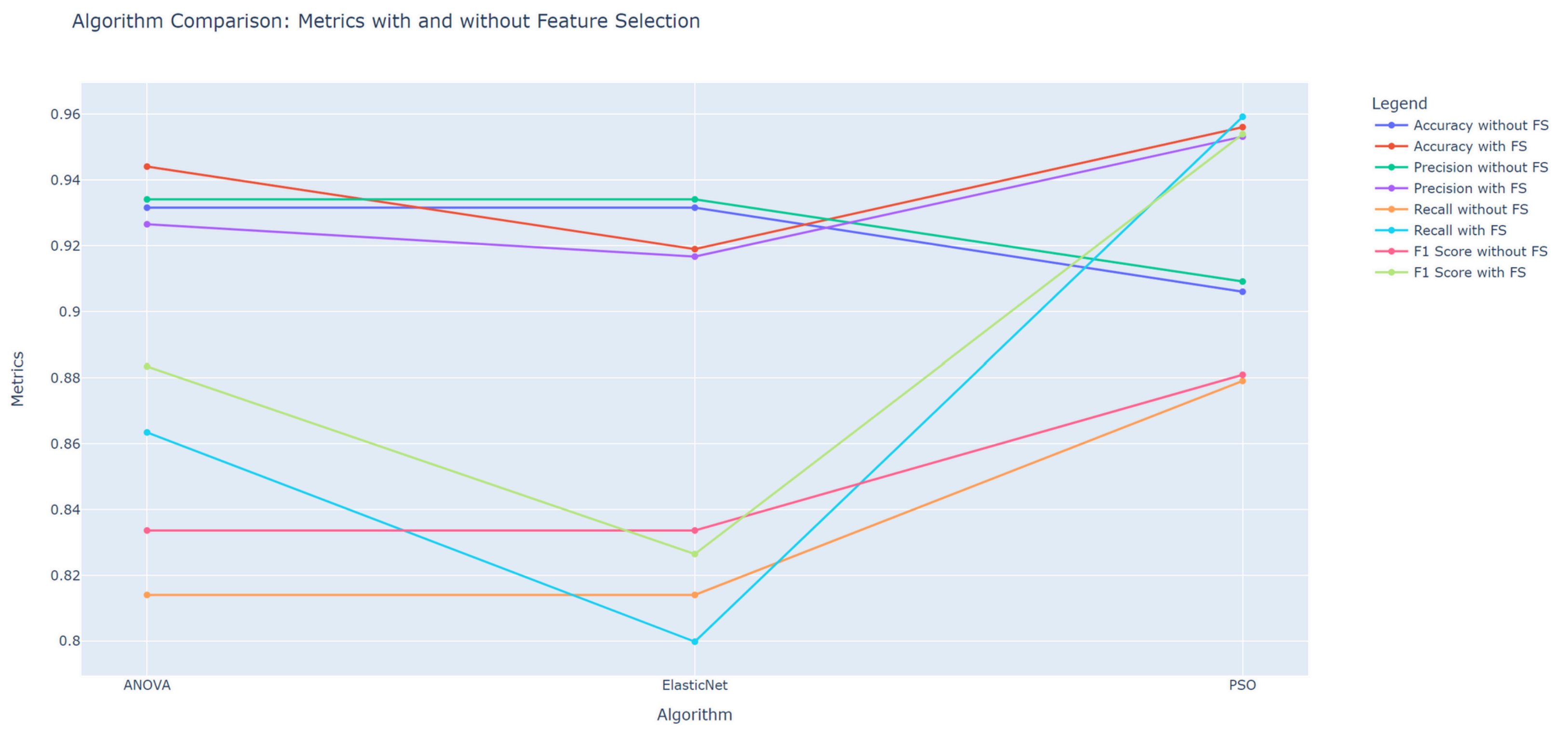


Figure 1: Average of 25 runs per algorithm on 12 datasets

1). ANOVA is a statistical method used to identify significant differences among group means by comparing the variability within groups to the variability between groups. In feature selection, the ANOVA F-test identifies features with significant differences across classes, helping select relevant features for model performance improvement. It evaluates the impact of feature selection on a K-Nearest Neighbors (KNN) classifier by selecting the top 10% of features based on statistical significance, retaining important features while reducing dimensionality.

2). Elastic Net is an embedded method that combines L1 (lasso) and L2 (ridge) regularization methods to enhance regression model performance, particularly when predictors are numerous or highly correlated. By combining the penalties, Elastic Net encourages sparsity (like lasso) for variable selection and stabilizes the solution (like ridge), improving prediction accuracy. It is particularly useful in scenarios with more predictors than observations or highly correlated predictors. PSO is a metaheuristic wrapper method that uses particles representing potential solutions (subsets of features) to search the solution space. Each particle adjusts its position based on its own experience and the experience of neighboring particles, converging towards the global best solution. PSO is particularly useful for feature selection in high-dimensional datasets and is favored for its ability to explore and exploit the search space to create optimal subsets of features, leading to better model convergence and performance.

3). Particle Swarm Optimization (PSO) is a metaheuristic wrapper method that works to search a solution space in a variety of ways using particles which are all influenced by their best search position along with their own personal search position. In the code, each particle represents a potential solution, a subset of features and these particles iteratively adjust their positions based on their own experience and the experience of neighboring particles, converging towards the global best solution. PSO is partially useful for feature selection in higher dimensional dataset and is often favored for its ability to create subsets of different features (both important and unimportant) to determine a better calculated position that would lead towards convergence [7].

Results

The study aimed to compare the performances of each algorithm and determine where their strengths and weaknesses lie. Overall, by looking at the overall averages and the individual performances, PSO outperforms ANOVA and ElasticNet due to its better accuracy, precision, recall, and F1 scores in feature selection. This is an indicator of how robust PSO is at optimizing classification performance. This makes sense from a computational standpoint, simply because PSO is superior to the other algorithms in efficiently being able to explore and exploit a search space because each particle represents a subset of features, a potential solution. This allows the particles to efficiently search for optimal feature subsets, a convergence point, thus enhancing model performance. This is the reason why PSO is left with more features after Feature Selection than any of the other two algorithms. Once more, PSO's adaptability and convergence properties make it an extremely vital tool for handling complex, high-dimensional data in feature selection tasks [8].

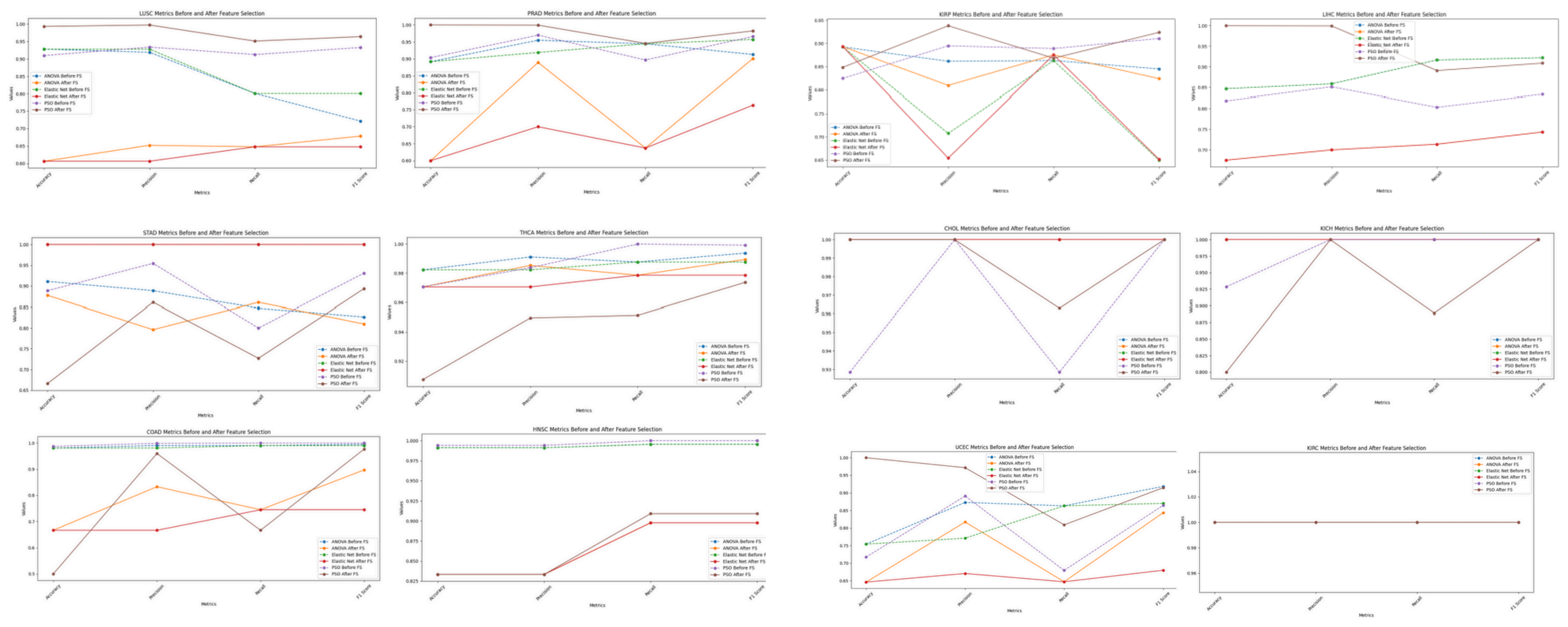


Figure 2: Graphs of 12 Datasets with metrics of Accuracy, Precision, Recall, and F1 Score records for both before and after Feature Selection with their respective algorithms

To determine how statistically different the results of the algorithm are from each other, a Fridman with Nemenyi Test was used. The results highlight that PSO is significantly better than both ANOVA and ElasticNet when looking at recall and F1 score when feature selection is applied. However, there are no significant differences between any combination of algorithms for accuracy with feature selection.

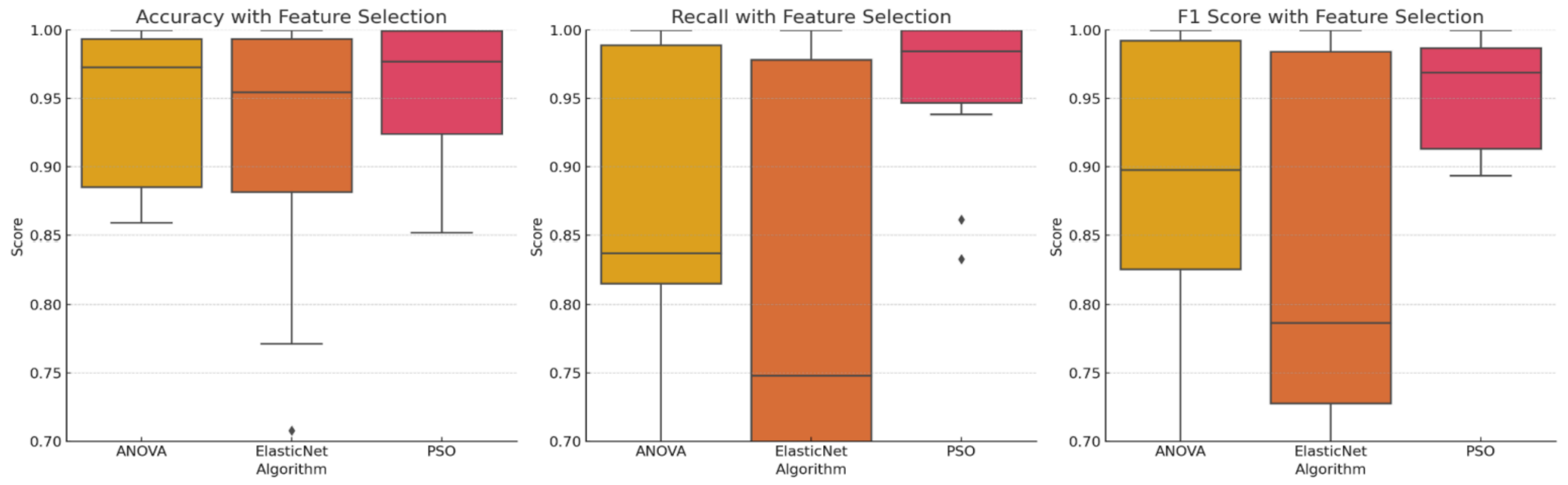
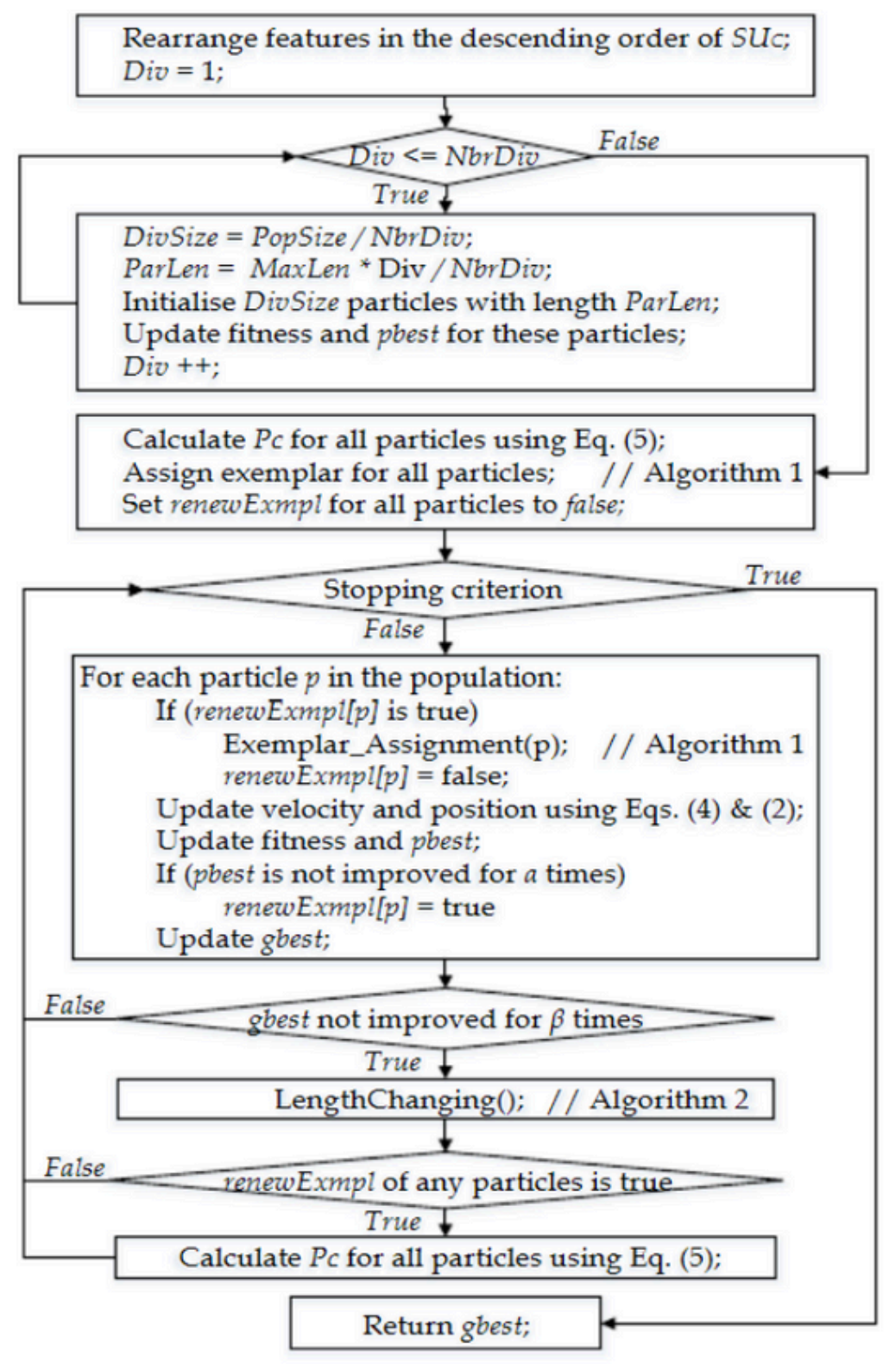


Figure 3: Box and Whisker plot of P scores from the Friedman with Nemenyi Test

Related Works

As previously mentioned, PSO is highly effective for feature selection in high-dimensional datasets due to its ability to produce robust metrics on average. Its strength lies in creating subsets of different features to identify the most suitable subset for classification. Tren et al. (2018) discovered a method to enhance PSO by varying the lengths of particle initialization, which allows particles to have different and shorter lengths. This adjustment defines a smaller search space and consequently improves PSO's performance. In their research, the initialization process was more advanced compared to regular Binary PSO because the features were ranked using a measure of Symmetric Uncertainty (Smith et al., 2018).



Rearrange features in the descending order of *SUc*;
Div = 1;

Div <= *NbrDiv*
True
False

DivSize = *PopSize* / *NbrDiv*;
ParLen = *MaxLen* * Div / *NbrDiv*;
Initialise *DivSize* particles with length *ParLen*;
Update fitness and *pbest* for these particles;
Div ++;

Calculate *Pc* for all particles using Eq. (5);
Assign exemplar for all particles;   // Algorithm 1
Set *renewExmpl* for all particles to *false*;

Stopping criterion
True
False

```
For each particle p in the population:
    If (renewExmpl[p] is true)
        Exemplar_Assignment(p);    // Algorithm 1
        renewExmpl[p] = false;
    Update velocity and position using Eqs. (4) & (2);
    Update fitness and pbest;
    If (pbest is not improved for a times)
        renewExmpl[p] = true
    Update gbest;
```

False
gbest not improved for β times
True

LengthChanging(); // Algorithm 2

False
renewExmpl of any particles is true
True

Calculate *Pc* for all particles using Eq. (5);

Return *gbest*;

Figure 4: Pseudocode for the Variable Length PSO (VLPSO) [8]

Acknowledgements

The authors acknowledge the support of the NDSU Department of Computer Science sponsored by NSF Directorate for Engineering and EPSCoR Co-Funding through NSF grant EEC2050175. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation

References

[1] "References," [Online]. Available: https://sciendo.com/article/10.2478/cait-2019-0001. [Accessed: Jul. 24, 2024]. [2] "Access Wayback," [Online]. Available: https://scholar.archive.org/work/3oy23kpqrre6jhq7crjsgeszce/access/wayback/https://www.iieta.org/download /file/fid/51911. [Accessed: Jul. 24, 2024]. [3] "Particle Swarm Optimization: Development, Applications, and Resources," [Online]. Available: https://www.researchgate.net/publication/3903911_Particle_swarm_optimization_Development_applications_and_resources. [Accessed: Jul. 24, 2024]. [4] "NDSU IT Help - CCAST," [Online]. Available: https://www.ndsu.edu/it/help/ccast/. [Accessed: Jul. 24, 2024]. [5] "ANOVA Analysis of Variance," [Online]. Available: https://www.analyticsvidhya.com/blog/2018/01/anova-analysis-of-variance/#:~:text=ANOVA%20calculates%20an%20F%2Dstatistic,compare%20means%20across%20multiple%20groups. [Accessed: Jul. 24, 2024]. [6] "Journal Article," [Online]. Available: https://academic.oup.com/jrsssb/article/67/2/301/7109482. [Accessed: Jul. 24, 2024]. [7] "IEEE Explore Document," [Online]. Available: https://ieeexplore.ieee.org/document/4433821. [Accessed: Jul. 24, 2024]. [8] "Comparison of three evolutionary algorithms: GA, PSO, and DE," [Online]. Available: https://www.researchgate.net/publication/260393066_Comparison_of_three_evolutionary_algorithms_GA_PSO_and_DE. [Accessed: Jul. 22, 2024]. [9] "Variable-Length Particle Swarm Optimisation for Feature Selection on High-Dimensional Classification," [Online]. Available: https://www.researchgate.net/publication/327767830_Variable-Length_Particle_Swarm_Optimisation_for_Feature_Selection_on_High-Dimensional_Classification. [Accessed: Jul. 24, 2024].