COMMUNICATION SCIENCE ET TECHNOLOGIE
Vol 23, N°2, pp 36, 43
ENPO – Maurice AUDIN
ASJP/ ESSN 2773-3483

A Comparative Study of Bio-Inspired Algorithms for Feature Selection in Biomedical Data

B. Khellat Kihel[1] and S. Chouraqui[2]

[1] Department of Economics, Faculty of Economics, Commercial and Management Sciences Mohamed BenAhmed Oran2 University Oran, Algeria

[2] Department of Informatics, Faculty of Mathematic and Informatics Université des Sciences et de la Technologie d'Oran Mohamed-Boudiaf Oran, Algeria

ABSTRACT

Feature selection is a critical step in biomedical data analysis due to the high dimensionality and noise commonly present in datasets such as medical images and patient records. This paper presents a comparison of several bio-inspired optimization algorithms: Genetic Algorithm (GA), Artificial Immune System (AIS), Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), Artificial Bee Colony (ABC), Firefly Algorithm (FFA), Grey Wolf Optimizer (GWO), and Black Widow Optimization (BWO) for feature selection in biomedical classification tasks. These algorithms are evaluated on their ability to reduce feature dimensionality while maintaining or enhancing predictive performance.

Experiments are conducted on benchmark biomedical datasets using k nearest neighbour (Knn) classifier. Our results highlight trade-offs between exploration and exploitation in different algorithms, revealing that FFA and GWO achieve higher accuracy with fewer features compared to swarm-based methods like PSO and ABC.

Keywords: Metaheuristics; Feature Selection; Disease diagnosis; Medical Data; Bio-inspired Algorithms.

Author Correspondence, e-mail: khellat_badra@yahoo.fr

https://www.asjp.cerist.dz/en/PresentationRevue/42

1. INTRODUCTION

The exponential growth of biomedical data has created unprecedented opportunities for data-driven healthcare innovations. However, the high dimensionality, noise, and heterogeneity inherent to these datasets pose significant challenges for machine learning models. Feature selection, the process of identifying the most discriminative variables while eliminating redundancies, is critical to improving model performance, interpretability, and computational efficiency in biomedical applications [1]. Traditional filter and wrapper methods often struggle with scalability and non-linear relationships, prompting the adoption of bio-inspired metaheuristic algorithms as robust alternatives [2].

Bio-inspired optimization techniques, modeled after natural phenomena such as evolution, swarm intelligence, and immune systems, excel at navigating complex search spaces. Algorithms like Genetic Algorithms (GA), Particle Swarm Optimization (PSO), and Ant Colony Optimization (ACO) have demonstrated success in tasks ranging from diabetes diagnosis to cancer detection from MRI images [3, 4]. More recent approaches, including the Grey Wolf Optimizer (GWO) and Black Widow Optimization (BWO), introduce novel mechanisms to balance exploration and exploitation, yet their comparative efficacy in biomedical feature selection remains underexplored [5].

This paper presents a systematic comparative study of several bio-inspired algorithms: GA, AIS, PSO, ACO, ABC, FFA, GWO, and BWO for feature selection across diverse biomedical datasets. The remainder of this paper is organized as follows: Section 2 reviews related work, Section 3 details the feature selection process, Section 4 presents experimental results, and Section 5 discusses implications and future directions.

2. RELATED WORK

Feature selection has been widely recognized as a critical step in high-dimensional biomedical data analysis. It reduces computational cost, improves model performance, and enhances interpretability. Traditional feature selection methods are commonly categorized into three groups: filter, wrapper, and embedded approaches. Filter methods rank features independently of any learning algorithm. They are computationally efficient but may ignore feature interactions. Wrapper methods use a predictive model to evaluate the performance of different feature subsets. While accurate, they tend to be computationally intensive. Embedded methods, perform feature selection during the model construction process but are often model-specific [6].

In recent years, bio-inspired metaheuristics have gained attention as promising tools for feature selection due to their ability to effectively search large, complex solution spaces. For instance, Chen et al. proposed a rough set-based FS approach using the Fish Swarm Algorithm (FSA) [7], which mimics fish behaviors such as swarming, following, and random movement. Similarly, Mirjalili introduced a wrapper-based FS method combining the Binary Dragonfly Algorithm (as a search strategy) with a K-Nearest Neighbors (KNN) classifier for

evaluation [8]. Another recent optimizer, the Ant Lion Optimizer (ALO), inspired by the hunting behavior of antlions, has been applied as a wrapper FS technique in [9, 10]. Additionally, the Firefly Algorithm (FFA) was hybridized with an Artificial Immune System (AIS) in [11] to reduce dimensionality in large datasets while avoiding local optima.

Among swarm intelligence techniques, the Grey Wolf Optimizer (GWO) has gained significant attention for solving FS problems [12, 13]. In [12], GWO was employed to select optimal features for coronary artery disease detection, with fitness evaluation performed using a Support Vector Machine (SVM) classifier. GWO, modeled after grey wolf hunting behavior, has also been widely explored in hybrid metaheuristics. For example, [13] proposed a multi-objective GWO variant with an innovative leader selection mechanism, while [14] developed an evolutionary GWO combined with a genetic algorithm for distributed generation allocation. Other studies hybridized GWO with the Artificial Bee Colony (ABC) algorithm to enhance complex system performance [15]. Further improvements include a binary PSO-GWO hybrid for FS [16], a PSO-GWO fusion to accelerate convergence [17], and a chaos-enhanced GWO for faster global optimization [18]. These studies demonstrate that hybrid metaheuristics often outperform conventional optimization methods.

Table 1. Summary of bio-inspired algorithms for feature selection

Algorithm	Inspiration	Key Mechanism	Strengths	Limitations	Biomedical Applications	Key References
Genetic Algorithms (GA)	Darwinian evolution	Crossover, mutation, and selection operations	- Effective global search - Works with non-linear data	- High computational cost - Premature convergence risk	Gene selection for cancer classification	[19], [2]
Artificial Immune Systems (AIS)	Human immune system	Antibody-antigen interactions	- Strong anomaly detection - Good adaptability	- Complex parameter tuning	Rare disease biomarker identification	[20] [11]
Particle Swarm Optimization (PSO)	Bird flocking/swarming	Particles collaboratively explore search space	- Fast convergence - Simple implementation	- Premature convergence - Parameter sensitivity	Medical imaging feature selection	[21], [22] [4]
Ant Colony Optimization (ACO)	Ant foraging behaviour	Pheromone-based path building	- Handles heterogeneous data well - Adaptive search	- Slow for large datasets	Multi-omics data analysis	[23]
Firefly Algorithm (FFA)	Firefly bioluminescence	Attraction based on light intensity	- Automatic search space division - Good for noisy data	- Sensitive to parameter settings	EEG signal classification	[11] [20]

Grey Wolf Optimizer (GWO)	Wolf pack hierarchy	Social hierarchy and hunting behavior	- Fast convergence - Few parameters	- Weak in high-dimensional spaces	Medical diagnosis systems	[22], [18], [24]
Black Widow Optimization (BWO)	Spider mating behaviour	Unique reproduction operator	- Escapes local optima - High accuracy	- Computationally intensive	Genomics and rare disease prediction	[5]

3. FEATURE SUBSET SELECTION

All The identification of useful and informative attributes for a given dataset, broadly referred to as Feature Selection (FS), is an attractive and challenging research topic for several domains including predictive data mining, pattern recognition, machine learning and information retrieval. One of the fundamental motivations for feature selection is to reduce the dimensionality. In fact, the presence of useless features may not only deteriorate the performance of learning algorithms but also obscure information behind data. Considered as a fundamental problem in machine learning, the role of FS is critical, especially in a context deemed with irrelevant features [6].

In the literature, authors identify three main objectives of using feature selection for classification that are:

- Feature extraction reduction task;
- Classification precision improvement;
- Performance estimation reliability improvement [4].

Figure 1 depicts the generic process that could summarize the different steps by any feature selection approach.

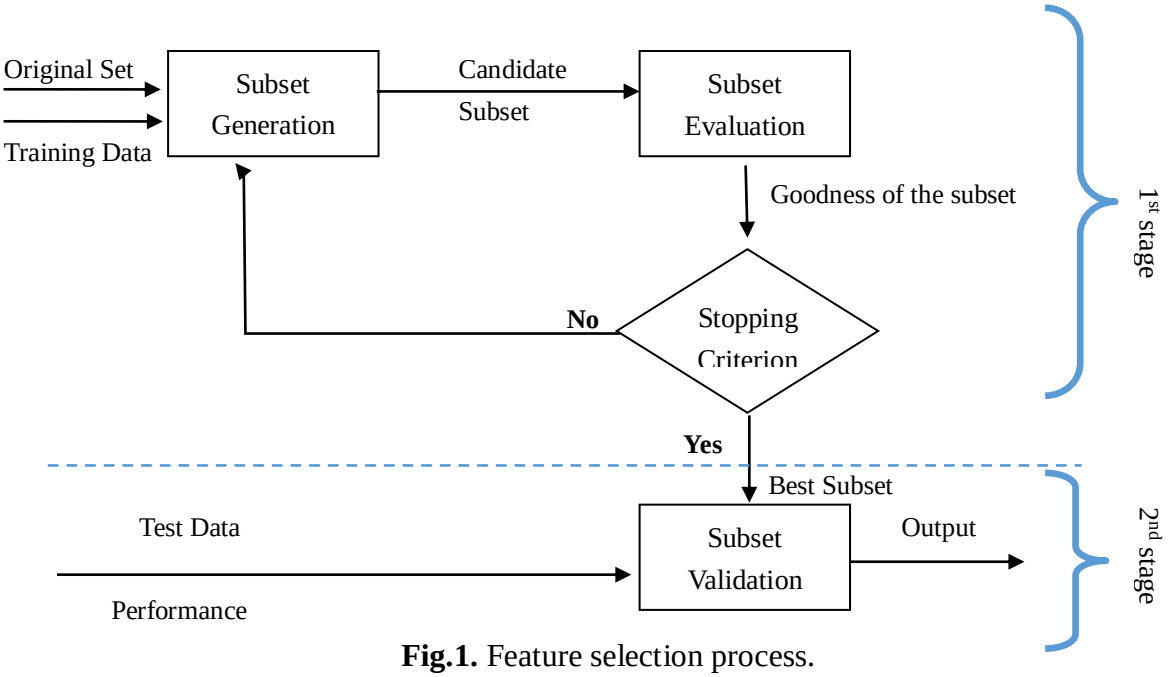


Fig.1. Feature selection process.

As input the feature selection process requires the dataset for which the relevant features will be identified. The outcome should include the retained features.
Generally, such process consists of two stages: search and validation. Each candidate subset is evaluated according to certain criterion and compared to the best solution found. Generation and evaluation is repeated until a given stopping criterion is satisfied. The best subset resulting from the first stage is provided as input for the second stage where it is usually, validated on a different data set [4].

4. EXPERIMENTAL RESULTS

We first describe our datasets and the number of features they contain. Then we give the results of the evaluation showing the relationship between the number of selected features and error rates.

4.1 Datasets used

In order to make our evaluation results comparable to most published results in feature selection evaluations, we have chosen datasets from the UCI machine learning repository.

Table 2. Selected datasets

Data set	Number of Features	Number of Classes	Number of instances
Spect Heart (Binary)	23	02	187
Spect Heart	44	02	187
Parkinson's Disease	22	02	96
Parkinson 2	26	02	1040
Breast tissue	09	06	106
Word Breast Cancer	09	02	699
Word Breast Cancer Diagnostic	31	02	569

4.2 Results and discussion

In this section we present the experimental results obtained for the different approaches on datasets.

Table 3. Experimental results

Dataset	GA (#Features/ Err)	AIS (#Features/ Err)	PSO (#Features/ Err)	ABC (#Features/ Err)	ACO (#Features/ Err)	FFA (#Features/ Err)	BWO (#Features/ Err)	GWO (#Features/ Err)
Spect Heart (Binary)	9 / 0.2875	9 / 0.212	12 / 0.0991	9 / 0.102	19 / 0.111	15 / 0.088	8 / 0.175	10 / 0.325
Spect Heart	22 / 0.1500	21 / 0.187	20 / 0.0615	34 / 0.062	34 / 0.080	19 / 0.056	3 / 0.1625	16 / 0.150
Parkinson's Disease	11 / 0.0104	12 / 0.083	12 / 0.0259	15 / 0.022	14 / 0.023	13 / 0.027	4 / 0.09375	11 / 0.020
Parkinson 2	17 / 0.3356	13 / 0.337	16 / 0.1713	16 / 0.169	20 / 0.170	17 / 0.166	6 / 0.34808	14 / 0.341
Breast tissue	6 / 0.2736	5 / 0.311	5 / 0.3067	7 / 0.332	9 / 0.510	7 / 0.290	4 / 0.37736	4 / 0.264
Wbc	6 / 0.0272	7 / 0.027	7 / 0.0191	7 / 0.017	8 / 0.019	7 / 0.016	7 / 0.021459	6 / 0.025
Wbcd	16 / 0.0615	16 / 0.049	17 / 0.0105	21 / 0.010	26 / 0.010	18 / 0.009	8 / 0.043937	7 / 0.042

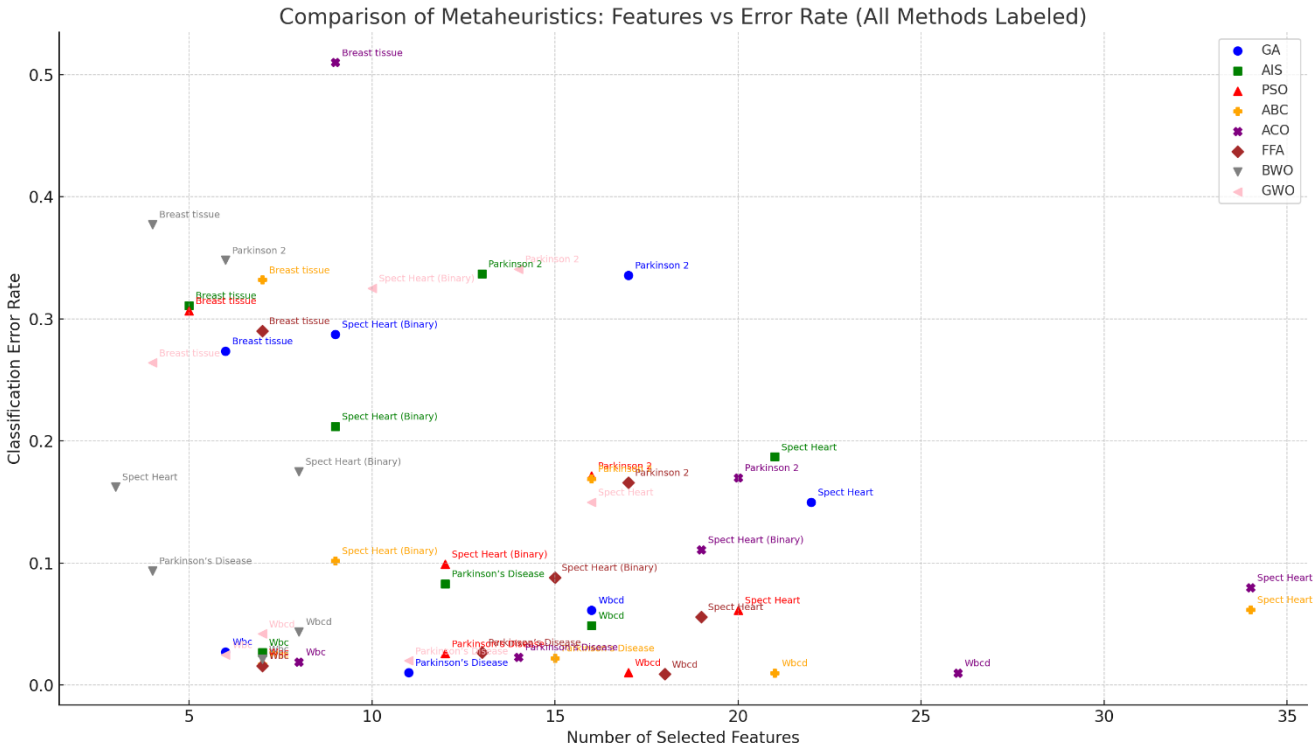


Fig.2. A Comparison of metaheuristics: Number of Features vs Error rate.

Following the comparative table, the analysis reveals several key insights. The Firefly Algorithm (FFA) stands out as the top performer, achieving the lowest error rates (e.g., 0.009 on Wbcd) while maintaining a balanced number of features. ABC and PSO also demonstrate strong performance, offering competitive accuracy (e.g., ABC: 0.010 on Wbcd) with reasonable feature reduction. However, BWO excels in aggressively minimizing features (e.g., just 3 for Spect Heart) but at the cost of higher error rates (0.377 on Breast tissue), suggesting potential oversimplification. Meanwhile, GA and AIS show inconsistent results, with GA ranging from highly accurate (0.0104 on Parkinson's) to poor (0.2875 on Spect Heart). GWO delivers mixed outcomes, performing well in some cases (0.020 on Parkinson's) but poorly in others (0.341 on Parkinson 2). These trends highlight clear trade-offs: while FFA and ABC prioritize precision, BWO favors extreme feature reduction. No single algorithm dominates all

scenarios, reinforcing the need for context-driven selection. For biomedical applications, FFA or ABC are ideal for diagnostics, PSO/ACO for high-dimensional data, and BWO for resource-constrained systems. Future work could explore hybrid models to merge these strengths.

5. CONCLUSION

This study highlights the critical trade-offs between accuracy and feature reduction in bio-inspired algorithms for biomedical data. The Firefly Algorithm (FFA) consistently delivers the lowest error rates, proving ideal for precision-sensitive tasks like diagnostics. In contrast, Black Widow Optimization (BWO) achieves remarkable feature reduction but struggles with accuracy, making it better suited for resource-constrained scenarios. ABC and PSO offer balanced performance, combining acceptable error rates with moderate dimensionality reduction. Notably, no single algorithm dominates across all datasets, reinforcing the importance of context-specific selection. The inconsistent performance of GA and AIS suggests they may require careful parameter tuning. For high-dimensional data, PSO and ACO demonstrate particular promise. Future research should explore hybrid approaches to merge FFA's accuracy with BWO's efficiency, while computational cost analysis remains an open challenge. These findings provide practical guidance for optimizing feature selection in biomedical applications, emphasizing that algorithm choice must align with specific clinical or research objectives. Ultimately, the study underscores the need for flexible, problem-aware solutions in bioinformatics pipelines.

6. REFERENCES

[1] Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J., & Liu, H. (2017). Feature selection: A data perspective. ACM Computing Surveys, 50(6), 1-45.

[2] Xue, B., Zhang, M., & Browne, W. N. (2019). Particle swarm optimization for feature selection in classification: Novel initialization and updating mechanisms. Applied Soft Computing, 81, 105549.

[3] Ahmad, S. R., Bakar, A. A., & Yaakub, M. R. (2020). Metaheuristic algorithms for diabetes classification using biometric features. Journal of Biomedical Informatics, 112, 103606.

[4] Al-Tashi, Q., Rais, H. M., & Jadid, S. (2020). Feature selection method based on grey wolf optimization for coronary artery disease classification. Computers in Biology and Medicine, 122, 103862.

[5] Hayyolalam, V., & Pourhaji Kazem, A. A. (2020). Black Widow Optimization Algorithm: A novel meta-heuristic approach for solving engineering optimization problems. Engineering Applications of Artificial Intelligence, 87, 103249.

[6] Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. Computers & Electrical Engineering, 40(1), 16-28.

[7] Chen, Y., Miao, D., & Wang, R. (2011). A rough set approach to feature selection based on fish swarm algorithm. Journal of Computational Information Systems, 7(5), 1449-1456.

[8] Mirjalili, S. (2016). Dragonfly algorithm: A new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems. Neural Computing and Applications, 27(4), 1053-1073

[9] Emary, E., Zawbaa, H. M., & Hassanien, A. E. (2016). Binary ant lion approaches for feature selection. Neurocomputing, 213, 54-65.

[10] Zawbaa, H. M., Emary, E., & Grosan, C. (2018). Feature selection via chaotic antlion optimization. PLoS ONE, 11(3), e0150652.

[11] B. Khellat and M. Benyettou, "Hybrid Bio-inspired Approach for Feature Subset Selection",International Journal of Applied Engineering

Research, Vol. 13, No.10, pp. 7895-7902, 2018.

[12] Al-Tashi, Q., et al. (2020). Binary grey wolf optimizer for feature selection in coronary artery disease detection. Computers in Biology and Medicine, 122, 103862.

[13] Nadimi-Shahraki, M. H., et al. (2021). An improved grey wolf optimizer for solving engineering problems. Expert Systems with Applications, 166, 113917.

[14] Saha, A., & Saikia, L. C. (2018). Performance analysis of combined GWO-GA optimized TCSC-PSS controlled power system. IET Generation, Transmission & Distribution, 12(14), 3392-3403.

[15] Kumar, M., et al. (2020). A novel hybrid ABC-GWO algorithm for optimization problems. Engineering with Computers, 36(3), 1101-1125.

[16] El-Hasnony, I. M., et al. (2022). Improved feature selection using hybrid PSO-GWO for document clustering. IEEE Access, 10, 5483-5502.

[17] Faris, H., et al. (2018). An efficient binary Salp Swarm Algorithm with crossover scheme for feature selection problems. *Knowledge-Based Systems, 154*, 43-67.

[18] Heidari, A. A., & Pahlavani, P. (2017). An efficient modified grey wolf optimizer with Lévy flight for optimization tasks. Applied Soft Computing, 60, 115-134.

[19] Yang, X.-S., Deb, S., Fong, S., He, X., & Zhao, Y.-X. (2020). Genetic algorithm optimized SVM for medical diagnosis. Journal of Medical Systems, 44(8), 1-9.

[20] Watkins, A., Boggess, L., & Reader, S. (2004). Artificial immune system for feature selection in classification problems. Proceedings of the 2004 IEEE Congress on Evolutionary Computation, 1, 606-613.

[21] Eberhart, R. C., & Shi, Y. (2001). Particle swarm optimization: Developments, applications and resources. Proceedings of the 2001 IEEE Congress on Evolutionary Computation, 1, 81-86.

[22] Mirjalili, S., Mirjalili, S. M., & Lewis, A. (2014). Grey wolf optimizer. Advances in Engineering Software, 69, 46-61.

[23] Solorio-Fernández, S., Carrasco-Ochoa, J. A., & Martínez-Trinidad, J. F. (2020). ACO-based hybrid feature selection method for medical image classification. Applied Soft Computing, 86, 105942.

[24] El-Hasnony, I. M., et al. (2022). Improved feature selection using hybrid PSO-GWO. IEEE Access, 10, 5483-5502.