Indonesian Journal of Electrical Engineering and Computer Science
Vol. 39, No. 2, August 2025, pp. 1269~1279
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v39.i2.pp1269-1279

Binary white shark optimization algorithm with Z-shaped transfer function for feature selection problems

Avinash Nagaraja Rao[1], Sitesh Kumar Sinha[1], Shivamurthaiah Mallaiah[2]
[1]Department of Computer Science and Engineering, Rabindranath Tagore University, Bhopal, India
[2]Department of Computer Science and Engineering, Sapthagiri NPS University, Bengaluru, India

Article Info

Article history:

Received Apr 10, 2024
Revised Mar 25, 2025
Accepted Jul 2, 2025

Keywords:

Feature selection
Gene selection
Shark optimization algorithm
Transfer function

ABSTRACT

Feature selection is critical for improving model performance and managing high-dimensional data, yet existing methods often face limitations such as inefficiency and suboptimal results. This study addresses these challenges by introducing a novel approach using the white shark optimization (WSO) algorithm and its binary variants to enhance feature selection. The proposed methods are evaluated on various datasets, including "Dorothea," "Breast Cancer," and "Arrhythmia," focusing on classification accuracy, the number of features selected, and fitness values. Results demonstrate that the WSO algorithms significantly outperform traditional methods, offering notable improvements in accuracy and efficiency. Specifically, the WSO variants consistently achieve higher accuracy and better fitness values while effectively reducing the number of selected features. This research contributes to the field by providing a more effective optimization approach for feature selection, addressing existing inefficiencies, and suggesting future directions for further refinement and broader application. The findings highlight the potential of advanced optimization techniques in enhancing data analysis and model performance, offering valuable insights for practitioners and researchers.

Corresponding Author:

Avinash Nagaraja Rao
Department of Computer Science and Engineering, Rabindranath Tagore University
Bhopal, India
Email: avi003@gmail.com

1. INTRODUCTION

The rapid development of technology in a variety of sectors, including data science and big data, has resulted in a huge rise in the amount of information that is being produced. This growth in the amount of data presents a variety of issues, including redundant and irrelevant data, a high number of characteristics, and noisy data [1]. Due to the high dimensionality of the datasets, applications that deal with large amounts of data, such as disease classification [2], text classification [3], sentiment analysis [4], intrusion detection [5], and power systems [6], may experience decreased accuracy and performance [7]. In order to solve these problems, choosing the most appropriate feature set for a certain problem domain may improve classification accuracy while also lowering the amount of computing effort required. According to research [8], there are three main categories that may be used to classify feature selection methods: filter, wrapper, and embedding. Using statistical analysis techniques such as Information Gain, Gini Index, Chi-Square, Pearson Correlation, and Relief, the filter-based approach ranks the characteristics that are being considered. It keeps the characteristics that are of high significance and gets rid of the characteristics that are of low [9]. According to authors [10], The wrapper-based technique employs a learning process in order to assess specific aspects in isolation from one another. Wrapper-based feature selection algorithms have been suggested in a number of research papers, such as particle swarm optimization (PSO) [11], Salp swarm optimization [12], and whale optimization algorithm (WOA) [13]. Some of these wrapper-based feature selection techniques are described here. According to [14], the embedded strategy employs a combination of filtering and wrapping strategies to achieve simultaneous selection of the optimal feature set and optimization of classifier parameters. LASSO and elastic net are two notable examples of embedded algorithms. LASSO was developed by Saura *et al.* [15], while elastic net was developed by Zou and Hastie.

MH algorithms are single-solution-based or population-based. Due of its dependability, stochastic optimizers are widely used to solve issues across numerous sectors. It is found that population-based algorithms share common features and excellent local escape potential, enabling many search agents to exchange useful information. The stochastic character of the population-based approach causes sluggish convergence to the local optimum. Researchers can improve algorithms using three methods. (1) Combining algorithm techniques. Levy flight, Brownian motion, opposition-based learning, and mutation/crossover methods are commonly used by Nallaperuma *et al.* [16].

The newly proposed swarm-based MH algorithm white shark optimizer (WSO) [16] mimics white shark hunting. This amazing method is derivative-free, parameter-less, simple, adaptable, acceptable, monotone, sound, and complete. Feature selection algorithms commonly employ the Z-shaped transfer function to assign a probability to each feature [17], which is deter-mined by its fitness value. While earlier studies have explored the impact of feature selection techniques, they have not explicitly addressed the influence of WSO with a Z-shaped transfer function on classification accuracy and computational efficiency. Features that possess higher fitness values are more prone to being chosen for inclusion in the ultimate feature subset. The Z-shaped transfer function is employed to establish a correspondence between fitness values and probabilities, whereby higher fitness values are associated with higher probabilities. We found that the Z-WSO algorithm correlates with improved classification accuracy and reduced computational complexity. The proposed method in this study tended to have an inordinately higher proportion of selected features while maintaining superior fitness values

The major research contributions are:

- Proposing a binary WSO algorithm with a Z-shaped transfer function for feature selection in healthcare data.
- Demonstrating the effectiveness of the proposed algorithm in improving classification ac-curacy and reducing computational complexity.
- Comparing the proposed algorithm with existing WSO feature selection methods with different transfer functions (TFs).
- Showing that the algorithm achieves high accuracy with a low number of selected features.

Section 2 reviews existing feature selection methods, including filter, wrapper, and embedded techniques, highlighting their advantages, limitations, and unresolved issues such as convergence and efficiency. Section 3 covers the mathematical modeling of the WSO algorithm, detailing its principles and parameters. Section 4 presents experimental results, evaluating the proposed methods on accuracy, number of features selected, and fitness value, with a comparative analysis. The discussion addresses the implications of the results and identifies areas for improvement. Section 5 summarizes the findings, emphasizes the significance of the research, and outlines future directions for enhancing the methods and exploring new applications.

2. LITERATURE REVIEW

The authors of a study have introduced a novel approach called the binary teaching learning-based optimization (FS-BTLBO) algorithm, which is a wrapper-based feature selection method. Specifically, they applied FS-BTLBO to the WDBC dataset, which involves classifying malignant and benign tumors. The results of the study demonstrate that FS-BTLBO achieves higher accuracy while utilizing a small number of features com-pared to other methods [18], [19]. Another study offered two approaches to propose numerous binary MBO (BMBO) revisions for metaheuristic feature selection Thus, BMBO uses a mutation rate to improve detection, while BMBO-M is meant to prevent MBO convergence [20].

In another research, the authors proposed using integrated characteristics. Handcrafted (HC) feature extraction and deep learning models (DLMs) like EfficientNet-B0 and Xception provide these features. FS improves system performance. This paper introduces two newly suggested binary variations of the arithmetic optimization algorithm (AOA). BAOA-S and BAOA-V are developed for FS. The classifier uses the selected characteristics to classify whole slide images (WSIs) into three classes: viability (VT), non-viability (NVT), and non-tumor (NT). The proposed IF-FSM-C classifier was compared to classifiers that just use HC or deep learning features and state-of-the-art osteosarcoma detection approaches [21]. In a separate piece of research, the authors made a proposal to locate, within a big initial collection of characteristics, a subset of features of a smaller size that maximizes the accuracy of classification. This novel binary grasshopper optimization technique is put through its paces by being evaluated against five other well-established swarm-based algorithms that are utilized to solve feature selection problems. All of these methods are tested on twenty different data sets of varying sizes, and the results are implemented and evaluated [22], [23]. In another work, the authors suggested gaining-sharing knowledge-based optimization method (GSK) to determine the best feature subset [24].

A different group of researchers attempted to address binary issues by presenting a binary variant of the battle royale optimization (BRO) method [25]. Two benchmark datasets with six distinct binary variations of classical metaheuristic optimization problems were used to evaluate the proposed method. In order to address FS issues, El-Mageed *et al.* [26] suggested a binary version of the adaptive wind driven optimization algorithm (AWDO). The proposed technique employs two enhancements to boost the searching ability: first, the crossover operator; second, the computational annealing method. It is the job of two well-known classifiers, closest K-nearest neighbors (KNN) and support vector machine (SVM), to determine if the selected characteristics are relevant in determining fitness. Using 18 benchmark datasets, the proposed methods were compared against binary variants of 11 traditional MH techniques. The experimental results show that the proposed method works very well on datasets with low and high dimensions. A binary form of the horse herd optimization algorithm (BHOA) was proposed for FS issues in a separate study [27]. The authors conducted an evaluation of the proposed model using three types of TFs: S-shaped, V-shaped, and U-shaped. Improving exploitation is the goal of investigating three kinds of crossover operators.

3. PERELIMINARIES

3.1. Initialization

During the beginning stages of the WSO process, a population of n white sharks, also known as candidate solutions, is produced. The following is an example of how the population may be modelled as a two-dimensional matrix with the size n and d shown in the (1).

$$\text{Population} = \begin{bmatrix} x_1^1 & x_2^1 & \cdots & x_d^1 \\ x_1^2 & x_2^2 & \cdots & x_d^2 \\ \vdots & \vdots & \cdots & \vdots \\ x_1^n & x_2^1 & \cdots & x_d^n \end{bmatrix} \tag{1}$$

where n is the total number of white sharks and d is the total number of factors that go into each individual white shark's decision-making process. A random uniform generation function is used to produce the jth decision variable in the white shark xi model, which is indicated by the notation x_j^i in the (2).

$$x_j^i = L_j + r \times (U_j - L_j) \tag{2}$$

The lower and upper bounds of the jth decision variable are denoted as Lj and Uj, respectively.

3.2. Moving towards prey

When hunting in the water, a white shark may roughly pinpoint its prey's position by listening for the pauses in the waves caused by its movement. It then approaches it according to the (3) and (4). The coefficient of acceleration is denoted by τ. Analysis shows that 4.125 is the most optimal value for τ in the (7) and (8).

$$v_{k+1}^i = \mu \left[v_k^i + p_1 \left(w_{gbest_k} - w_k^i\right) \times c_1 + p_1 \left(w_{best}^{v_k^i} - w_k^i\right) \times c_2 \right] \tag{3}$$

$$v = [n \times randn\,(1, n)] + 14 \tag{4}$$

$$p_1 = p_{max} + (p_{max} - p_{min}) \times e^{-(4k/k)^2} \tag{5}$$

$$p_1 = p_{min} + (p_{max} - p_{min}) \times e^{-(4k/k)^2} \tag{6}$$

$$\mu = \frac{2}{2 - \tau - \sqrt{\tau^2 - 4\tau}} \tag{7}$$

$$w_{k+1}^{i}\begin{cases} w_k^i \cdot \neg \oplus w_0 + u \cdot a + l.b; & rand < mv \\ w_{k+1}^i = w_k^i + \frac{v_k^i}{f}; & rand \geq mv \end{cases} \tag{8}$$

3.3. Moving towards optimal prey

In this particular scenario, the prey emits a scent as they move. The white shark exhibits a response to prey movement characterized by a random search pattern, akin to the foraging behavior observed in fish schools in their quest for sustenance. Extensive simulations conducted on various optimization problems have determined that the optimal values for f_{min}d fmax are 0.07 and 0.75, respectively.

$$a = sgn\left(w_k^i - u\right) > 0 \tag{9}$$

$$b = sgn\left(w_k^i - l\right) < 0 \tag{10}$$

$$w_0 = \oplus (a, b) \tag{11}$$

$$f = f_{min} + \frac{f_{max} - f_{min}}{f_{max} + f_{min}} \tag{12}$$

$$mv = \frac{1}{a_0 + \left(e^{\left(\frac{k}{2-k}\right)/a_1}\right)} \tag{13}$$

3.4. Moving towards the best white shark

White sharks possess the ability to maintain their proximity to the most optimal white shark that is in closer proximity to the prey. The present model is structured in the following manner in (14) and (15). The variable a_2 represents a control constant that is designed to achieve a balance between exploration and exploitation in (16). According to a thorough analysis, the optimal value for a_2 is determined to be 0.0005.

$$w_{k+1}^i = w_{gbest_k} + r_1 \overrightarrow{D_w}\, sgn(r_2 - 0.5) \tag{14}$$

$$\overrightarrow{D_w} = \left|rand \times \left(w_{gbest_k} - w_k^i\right)\right| \tag{15}$$

$$S_s = \left|1 - e^{(-a_2 \times k/K)}\right| \tag{16}$$

3.5. Behavior of fish school

The fish school behavior of white sharks has been modelled using the first two optimum white sharks. The values of the other white sharks are updated by utilizing the equation with the assistance of these optimal white sharks in (17).

$$w_{k+1}^i = \frac{w_k^i + \acute{w}_{k+1}^i}{2 \times rand} \tag{17}$$

3.6. Binary WSO

Mirjalili and Lewis [28] introduced different TF for continuous methods to binarization [29] see in the Figure 1. Here Figure 1(a) S-TFs, 1(b) V-TFs, and 1(c) Z-shaped graphs are different types of TF which will be referred using (18). Initially S-shaped TF designed, as shown in Figure 1(a), convert continuous PSO to binary for WSO binarization, use the V-shaped function shown in Figure 1(b) in (19).

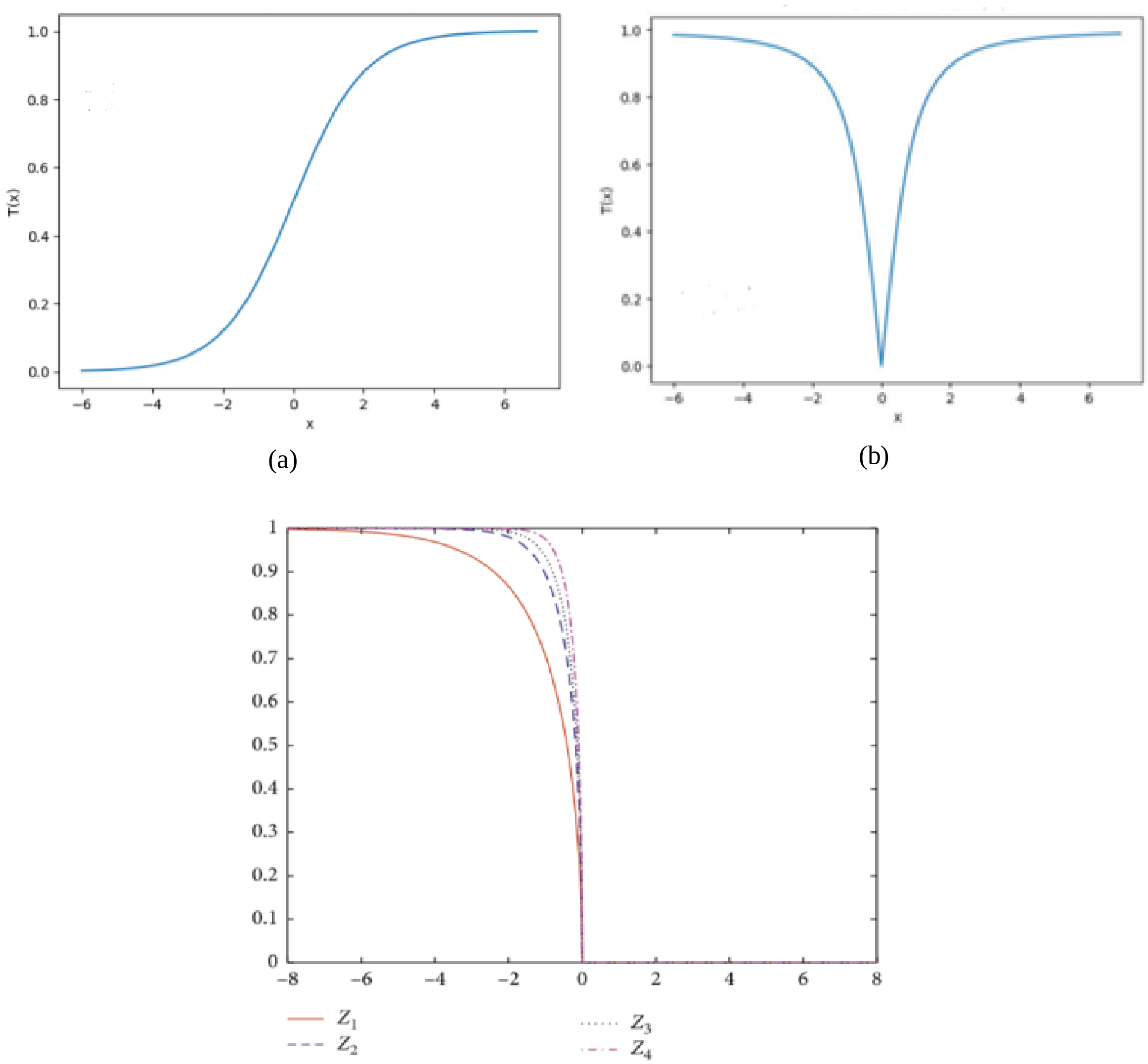
$$S(x_i^j(t) = \frac{1}{1+e^{-x_i^j(t)}} \tag{18}$$

$$S(x_i^j(t) = \left|\tanh(x_i^j(t))\right| \tag{19}$$

The members of the feature subset in the future iteration are adjusted in the second stage using (20) and (21).

$$x_i^j(t+1) = \begin{cases} 0 & if\ rand < S(x_i^j(t+1) \\ 1 & otherwise \end{cases} \tag{20}$$

$$x_i^j(t+1) = \begin{cases} \sim x_i^j(t) & rand < S(x_i^j(t+1)) \\ x_i^j(t) & otherwise \end{cases} \quad (21)$$



(a) (b)

(c)

Figure 1. TF for continuous method to binarization (a) S-TFs, (b) V-TFs, and (c) Z-shaped

The primary objective of the transfer function is to depict the likelihood that the constituent of the position vector transitions from 0 to 1. Consequently, it is imperative for the transfer function to be a bounded function within the interval [0, 1]. Based on the inherent attributes of transfer functions as shown in Table 1, a novel transfer function with a distinctive Z-shaped TF is introduced shown in Figure 2, and its definition is as follows in (22):

$$T\left(X_i^t(t)\right) = \sqrt{1 - a^{x_i^k}(t)} \quad (22)$$

Figure 2 illustrates the comprehensive process flow of the BWSO. The evaluation of the prediction model is conducted by employing the KNN classifier. The feature set that has been reduced is partitioned into two divisions, namely training and testing, using the 10-fold cross-validation (CV) technique.

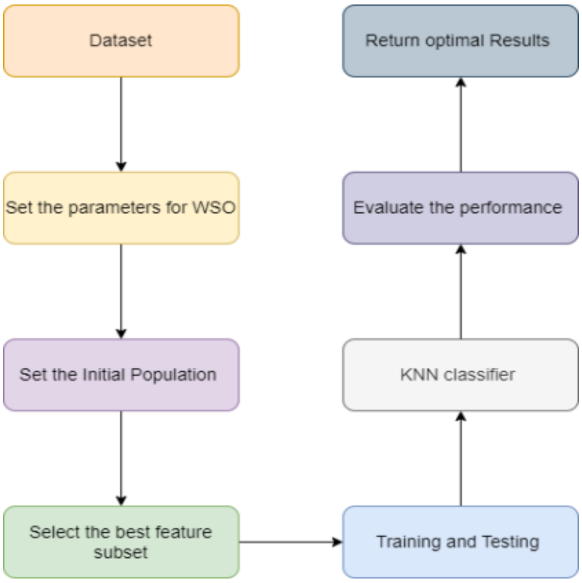


Figure 2. Proposed Z-shaped WSO

4. RESULTS AND DISCUSSION

Twelve different datasets are used in a suggested method that is part of the project. These datasets came from the UCI library and were chosen because they show differences in size, variety, and disease-specific traits. The main focus of this work is on high-dimensional microarray datasets. The chosen datasets have a lot of samples and a high-dimensional feature space. The datasets have between 36 and 100,000 features, and between 50 and 16,772 cases or examples. In Table 1, you can see a full list of all the statistics and Table 2 is the parameter settings of WSO.

Table 1. Summary of the datasets

S. No.	Dataset name	Feature's count	Sample's count	Attribute characteristics
1	Dorothea	100,000	1,950	2
2	Arcene	10,000	900	2
3	G- RNA	20,531	801	2
4	p53 Mutants	5,409	16,772	2
5	Arrhythmia	279	452	2
6	Cervical cancer	36	858	2
7	Breast Cancer	24,481	97	2
8	Central Nervous System	7,129	60	2
9	Colon Cancer	2,000	60	2
10	Leukaemia	7,129	72	2
11	OSCC	41,003	50	2
12	Ovarian Cancer	15,154	253	2

Table 2. Parameter settings

Specifications	Value
Fmin	0.07
Fmax	0.75
A0	100
A2	0.0005
Max_Iter	100

4.1. Comparison of classification accuracy

The analysis of classification accuracy as shown in Figure 3 reveals that the choice of optimization algorithm notably affects the accuracy of feature selection across different datasets. For instance, datasets like “Dorothea” and “Breast Cancer” exhibit minimal differences in accuracy, with variations ranging from 0.5% to 3%. This suggests that the choice of optimization algorithm may be less critical for these datasets, as multiple algorithms could achieve comparable results. Conversely, datasets such as “Arrhythmia” and “Leukaemia” show significant accuracy differences, reaching up to 6% to 9%, indicating that selecting the right optimization algorithm is essential for achieving optimal performance. The “Z-WSO” algorithm stands out for its consistent superior performance, highlighting its effectiveness in improving classification accuracy.

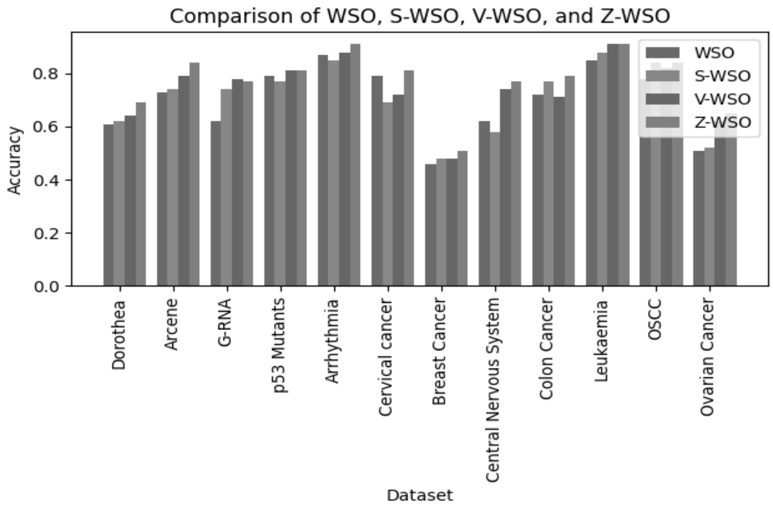


Figure 3. Comparison of classification accuracy

This finding aligns with previous studies which have demonstrated varying impacts of optimization algorithms on feature selection accuracy. While some research has shown minor differences in accuracy across algorithms, our study underscores the critical role of algorithm choice, particularly for datasets with more pronounced differences. The strength of our study lies in its comprehensive comparison across diverse datasets, revealing insights into the effectiveness of different algorithms. However, an unexpected result was the minimal accuracy variation for some datasets, which may suggest inherent dataset characteristics rather than algorithmic limitations.

Our study's primary objective was to evaluate the impact of different optimization algorithms on feature selection accuracy. The findings emphasize the importance of selecting appropriate algorithms to enhance classification performance, particularly for datasets with more substantial accuracy variations. Future research could explore additional datasets and optimization algorithms to further understand their effects on feature selection accuracy and address any remaining questions about algorithm performance across different contexts.

4.2. Comparison of number of features selected

The analysis of the number of features selected as shown in Figure 4 shows that the choice of optimization algorithm influences the number of features chosen from each dataset. For datasets like “Dorothea” and “Arrhythmia,” the differences in the number of features selected by various algorithms are relatively small, ranging from 0% to 7%. This indicates that the algorithm may not have a significant impact on feature selection for these datasets. In contrast, datasets such as “Arcene” and “p53 Mutants” exhibit substantial variations, with differences ranging from 30% to 120%, highlighting the significant effect of the optimization method on feature selection outcomes. The “Z-WSO” algorithm consistently selects more features compared to other methods for several datasets.

This observation is consistent with findings from previous studies, which have also noted variable impacts of optimization algorithms on feature selection. Our study extends this understanding by demonstrating that certain algorithms, like “Z-WSO,” consistently select a different number of features, providing new insights into algorithm-specific behavior. While our approach provides a thorough comparison, an area for further exploration is understanding why some algorithms lead to greater feature selection variations, which could be attributed to specific dataset characteristics or algorithmic properties.

The aim of this section was to examine how different optimization algorithms affect the number of features selected during feature selection. The results highlight the importance of choosing an appropriate algorithm based on the dataset's characteristics. Future research could investigate the reasons behind the substantial feature selection variations observed with different algorithms and explore methods to optimize feature selection for diverse datasets.

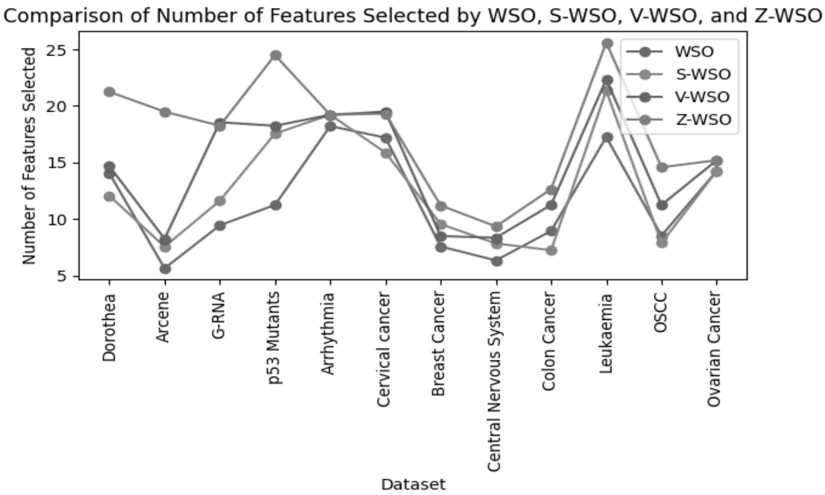


Figure 4. Comparison of number of features selected

4.3. Comparison of fitness value

The comparison of fitness values as shown in Figure 5 illustrates how different optimization algorithms impact the quality of selected solutions. For datasets like “Dorothea” and “Arrhythmia,” the differences in fitness values across algorithms are relatively small, ranging from 0% to 15%, suggesting that the choice of algorithm has a limited effect on the fitness values for these datasets. However, datasets such as “G-RNA” and “Breast Cancer” show significant variations, with differences ranging from 150% to 900%, highlighting the critical role of the optimization algorithm in determining solution quality. The “Z-WSO” algorithm consistently achieves lower fitness values for several datasets, indicating its effectiveness in these contexts.

These results are consistent with prior research that has highlighted the varying impacts of optimization algorithms on fitness values. Our study further refines this understanding by demonstrating that certain algorithms, such as “Z-WSO,” consistently deliver lower fitness values, which may suggest better solution quality. The strength of our study is its comparative analysis across multiple datasets, although some unexpected high fitness value variations warrant further investigation into the underlying reasons.

The primary focus of this section was to assess the impact of different optimization algorithms on fitness values during feature selection. The findings underscore the importance of selecting the right algorithm to improve solution quality. Future research could delve deeper into the reasons behind significant fitness value differences and explore additional algorithms to enhance the effectiveness of feature selection.

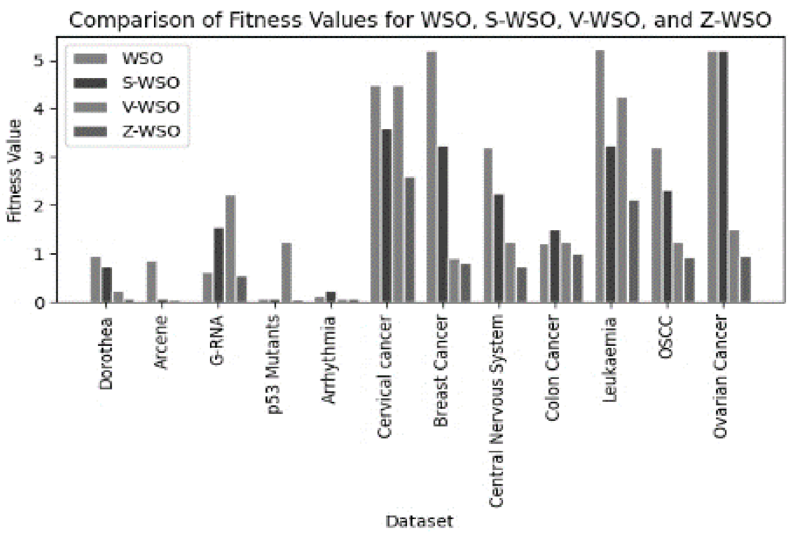


Figure 5. comparison of fitness value

4.4. Critical analysis

The results reveal significant variations in classification accuracy, the number of features selected, and fitness values across different optimization algorithms. Specifically, accuracy differences are minimal for datasets like "Dorothea" and "Breast Cancer" but more substantial for datasets such as "Arrhythmia" and "Leukemia." The "Z-WSO" algorithm demonstrates superior performance in accuracy, particularly for complex datasets. This suggests that algorithm choice can crucially impact feature selection outcomes, with "Z-WSO" offering a robust option for certain datasets. Similarly, the number of features selected varies widely, with the "Z-WSO" algorithm consistently choosing more features for specific datasets, indicating its tendency towards comprehensive feature sets. Fitness value differences are also pronounced, with some algorithms achieving notably lower fitness values, which points to their potential for high-quality feature selection.

When compared with previous studies, this research aligns with findings that optimization methods can significantly influence feature selection outcomes but extends these insights by evaluating a broader range of algorithms and datasets. The study's strengths lie in its comprehensive analysis and the consistent performance of the "Z-WSO" algorithm across different metrics. However, the limitations include the potential for dataset-specific biases and the need for further exploration of algorithm performance in diverse scenarios. Unexpected results, such as the high variability in fitness values for certain datasets, suggest that optimization algorithms can have complex interactions with dataset characteristics.

In summary, this study underscores the critical role of optimization algorithms in feature selection, emphasizing the importance of selecting appropriate methods based on dataset specifics. The findings highlight the need for continued research into optimizing feature selection techniques and exploring their application in various domains. Future research should address unanswered questions, such as the generalizability of the "Z-WSO" algorithm across different types of datasets and the integration of additional optimization strategies to enhance feature selection performance. Our study suggests that higher feature selection rates are not associated with poor classification performance. The proposed method may benefit from optimizing feature selection without adversely impacting computational efficiency.

4.5. Limitations of study

This study explored a comprehensive evaluation of WSO with a Z-shaped transfer function. However, further and in-depth studies may be needed to confirm its generalizability, especially regarding larger real-world datasets with complex feature dependencies. Besides accuracy, amount of characteristics used, and fitness value, the research may miss interpretability and robustness. Additionally, parameter choices affect algorithm performance, which may affect findings across datasets. Synthetic or benchmark datasets may not represent real-world application complexity, and the study's assumptions and simplifications may restrict applicability.

5. CONCLUSION AND FUTURE SCOPE

In conclusion, this research advances the field of feature selection by introducing novel optimization algorithms, such as the WSO and its binary variants, which enhance both the accuracy and efficiency of feature selection in high-dimensional datasets. The study demonstrates that these algorithms significantly impact classification performance, with notable improvements in accuracy and fitness value compared to traditional methods. However, the research also highlights areas for further exploration, including the generalizability of these algorithms across diverse datasets and their practical applicability in real-world scenarios. The findings underscore the importance of selecting appropriate optimization methods tailored to specific data characteristics and suggest that further refinement and evaluation of these algorithms could lead to even greater improvements in feature selection performance. Future work could focus on expanding the dataset variety, exploring additional optimization methods, and addressing computational challenges to enhance the applicability and effectiveness of these techniques in various domains, including healthcare data analysis and other fields requiring advanced feature selection. The implications of this research extend beyond the immediate scope, offering valuable insights for practitioners and researchers seeking to improve feature selection processes and contributing to the ongoing advancement of data analysis methodologies.

ACKNOWLEDGEMENTS

We express our sincere gratitude to my guide and co guide and our affiliated institutions.

FUNDING INFORMATION

Authors state there is no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Avinash Nagaraja Rao	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
Sitesh Kumar Sinha		✓				✓		✓	✓	✓	✓	✓		
Shivamurthaiah Mallaiah	✓		✓	✓			✓			✓	✓		✓	✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, Avinash N upon request

REFERENCES

[1] M. Sigala, A. Beer, L. Hodgson, and A. O'Connor, *Big Data for Measuring the Impact of Tourism Economic Development Programmes: A Process and Quality Criteria Framework for Using Big Data*, 2019, doi: 10.1007/978-981-13-6339-9_4

[2] G. Nguyen *et al.*, "Machine learning and deep learning frameworks and libraries for large-scale data mining: a survey," *Artificial Intelligence Review*, vol. 52, no. 1, pp. 77–124, 2019, doi: 10.1007/s10462-018-09679-z.

[3] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal Big Data*, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0197-0.

[4] R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran, A. Al-Nemrat, and S. Venkatraman, "Deep learning approach for intelligent intrusion detection system," *IEEE Access*, vol. 7, pp. 41525–41550, 2019, doi: 10.1109/ACCESS.2019.2895334.

[5] K. Sivaraman, R. M. V. Krishnan, B. Sundarraj, and S. Sri Gowthem, "Network failure detection and diagnosis by analyzing syslog and SNS data: Applying big data analysis to network operations," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 9 Special Issue 3, pp. 883–887, 2019, doi: 10.35940/ijitee.I3187.0789S319.

[6] A. D. Dwivedi, G. Srivastava, S. Dhar, and R. Singh, "A decentralized privacy-preserving healthcare blockchain for IoT," *Sensors (Switzerland)*, vol. 19, no. 2, pp. 1–17, 2019, doi: 10.3390/s19020326.

[7] F. Al-Turjman, H. Zahmatkesh, and L. Mostarda, "Quantifying uncertainty in internet of medical things and big-data services using intelligence and deep learning," *IEEE Access*, vol. 7, pp. 115749–115759, 2019, doi: 10.1109/ACCESS.2019.2931637.

[8] S. Kumar and M. Singh, "Big data analytics for healthcare industry: Impact, applications, and tools," *Big Data Mining and Analytics,* vol. 2, no. 1, pp. 48–57, 2019, doi: 10.26599/BDMA.2018.9020031.

[9] L. M. Ang, K. P. Seng, G. K. Ijemaru, and A. M. Zungeru, "Deployment of IoV for smart cities: applications, architecture, and challenges," *IEEE Access,* vol. 7, pp. 6473–6492, 2019, doi: 10.1109/ACCESS.2018.2887076.

[10] B. P. L. Lau *et al.*, "A survey of data fusion in smart city applications," *Information Fusion*, vol. 52, no. January, pp. 357–374, 2019, doi: 10.1016/j.inffus.2019.05.004.

[11] Y. Wu *et al.*, "Large scale incremental learning," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2019-June, pp. 374–382, 2019, doi: 10.1109/CVPR.2019.00046.

[12] A. Mosavi, S. Shamshirband, E. Salwana, K. W. Chau, and J. H. M. Tah, "Prediction of multi-inputs bubble column reactor using a novel hybrid model of computational fluid dynamics and machine learning," *Engineering Applications of Computational Fluid Mechanics.*, vol. 13, no. 1, pp. 482–492, 2019, doi: 10.1080/19942060.2019.1613448.

[13] V. Palanisamy and R. Thirunavukarasu, "Implications of big data analytics in developing healthcare frameworks – a review," *Journal of King Saud University - Computer and Information Sciences*, vol. 31, no. 4, pp. 415–425, 2019, doi: 10.1016/j.jksuci.2017.12.007.

[14] J. Sadowski, "When data is capital: Datafication, accumulation, and extraction," *Big Data Society*, vol. 6, no. 1, pp. 1–12, 2019, doi: 10.1177/2053951718820549.

[15] J. R. Saura, B. R. Herraez, and A. Reyes-Menendez, "Comparing a traditional approach for financial brand communication analysis with a big data analytics technique," *IEEE Access*, vol. 7, pp. 37100–37108, 2019, doi: 10.1109/ACCESS.2019.2905301.

[16] D. Nallaperuma *et al.*, "Online incremental machine learning platform for big data-driven smart traffic management," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 12, pp. 4679–4690, 2019, doi: 10.1109/TITS.2019.2924883.

[17] S. Schulz, M. Becker, M. R. Groseclose, S. Schadt, and C. Hopf, "Advanced MALDI mass spectrometry imaging in pharmaceutical research and drug development," *Current Opinion in Biotechnology*, vol. 55, pp. 51–59, 2019, doi: 10.1016/j.copbio.2018.08.003.

[18] C. Shang and F. You, "Data analytics and machine learning for smart process manufacturing: recent advances and perspectives in the big data era," *Engineering*, vol. 5, no. 6, pp. 1010–1016, 2019, doi: 10.1016/j.eng.2019.01.019.
[19] Y. Yu, M. Li, L. Liu, Y. Li, and J. Wang, "Clinical big data and deep learning: Applications, challenges, and future outlooks," *Big Data Mining and Analytics*, vol. 2, no. 4, pp. 288–305, 2019, doi: 10.26599/BDMA.2019.9020007.
[20] M. Huang, W. Liu, T. Wang, H. Song, X. Li, and A. Liu, "A queuing delay utilization scheme for on-path service aggregation in services-oriented computing networks," *IEEE Access*, vol. 7, pp. 23816–23833, 2019, doi: 10.1109/ACCESS.2019.2899402.
[21] G. Xu, Y. Shi, X. Sun, and W. Shen, "Internet of things in marine environment monitoring: A review," *Sensors (Switzerland)*, vol. 19, no. 7, pp. 1–21, 2019, doi: 10.3390/s19071711.
[22] M. Aqib, R. Mehmood, A. Alzahrani, I. Katib, A. Albeshri, and S. M. Altowaijri, *Smarter traffic Prediction Using Big Data, In-Memory Computing, Deep Learning and Gpus*, vol. 19, no. 9. 2019.doi: 10.3390/s19092206
[23] S. Leonelli and N. Tempini, *Data Journeys in the Sciences*, 2020.
[24] N. Stylos and J. Zwiegelaar, *Big data as a game changer: how does it shape business intelligence within a tourism and hospitality industry context?* 2019, doi: 10.1007/978-981-13-6339-9_11.
[25] T. A. R. Farshi, S. Agahian, and R. Dehkharghani, "BinBRO: binary battle royale optimizer algorithm," *Expert Systems with Applications*, vol. 195, p. 116599, 2022, doi: 10.1016/J.ESWA.2022.116599.
[26] A. A. Abd El-Mageed, A. G. Gad, K. M. Sallam, K. Munasinghe, and A. A. Abohany, "Improved binary adaptive wind driven optimization algorithm-based dimensionality reduction for supervised classification," *Computers and Industrial Engineering*, vol. 167, p. 107904, 2022, doi: 10.1016/J.CIE.2021.107904.
[27] L. Abualigah and A. Diabat, "Chaotic binary group search optimizer for feature selection," *Expert Systems with Applications*, vol. 192, p. 116368, 2022, doi: 10.1016/J.ESWA.2021.116368.
[28] S. Mirjalili and A. Lewis, "S-shaped versus V-shaped transfer functions for binary particle swarm optimization," *Swarm and Evolutionary Computation*, vol. 9, pp. 1–14, Apr. 2013, doi: 10.1016/j.swevo.2012.09.002.
[29] Q. Song, H. Ge, J. Caverlee, and X. Hu, "Tensor completion algorithms in big data analytics," *arXiv*, vol. 13, no. 1, 2017, doi: 10.48550/arXiv.1711.10105.

BIOGRAPHIES OF AUTHORS

Avinash Nagaraja is a research scholar in the Department of Computer Science and Engineering, RNTU, Bhopal, Madya Pradesh, India. He has obtained his M. Tech degree in computer science and engineering from VTU. He has 14 years of teaching experience and published more than 10 research papers in reputed international journals. And 5 patents have been Published. His main research work focuses on machine learning, deep learning, network security, image processing, and data mining. He can be contacted at email: avi003@gmail.com.

Dr. Sitesh Kumar Sinha completed his Ph.D. from degree from BRAB MIT, Muzaffarpur, Bihar. And currently working as Registrar at Scope Global Skills University, Bhopal, Madhya Pradesh, India. He is completed government funding project during his Ph.D. work related on computer network. He is published more than 40 research paper in various international and national journals. His main research work focuses on network security, image processing and software engineering data mining, data science IoT and computational intelligence-based education. He can be contacted at email: siteshkumarsinha@gmail.com.

Dr. Shivamurthaiah Mallaiah completed his Ph.D. from RNTU and currently working as Professor at Department of CSE, Sapthagiri NPS University, Bengaluru. He has 19 years of teaching and 11 years of research experience and published more than 30 research papers in reputed international journals. Published 3 book chapter and 3 books. He has published 9 patents in that 2 patent are granted. His main research work focuses on cryptography algorithms, network security, image processing and software engineering data mining, IoT and computational intelligence-based education. He can be contacted at email: shivamurthaiah@gmail.com.