A Novel Swarm Intelligent Based Feature Selection Method for Stream Clustering

Suman R. Tiwari [1], Prof. Kaushik K. Rana [2]
[1]Research Scholar, Gujarat Technological University, Ahmedabad, Gujarat, India., and
[1] Lecturer, R.C. Technical Institute, Gujarat, Ahmedabad, India
[2]Associate Professor, L. D. College of Engineering, Ahmedabad, Gujarat, India.
*Corresponding Email: *rsumantiwari@gmail.com*

Abstract

Stream clustering faces challenges due to data dynamics, high dimensionality, and absence of labels. Existing feature selection methods rely on labeled data and threshold values, which limit their effectiveness on large, unstructured streaming data. To overcome these issues, we propose a swarm intelligence-based feature selection method for streaming environments, inspired by the problem-solving mechanism of a swarm. The Grey Wolf Algorithm is swarm intelligent based method known for its simplicity, minimal parameter derivation, and hunting behavior, which balances exploration and exploitation in complex search spaces. Our enhancements—including dynamic scaling, separate search phases, and retention of elite solutions—address premature convergence, prevent trapping in local optima, and reduce parameter sensitivity in the standard Grey Wolf Optimization algorithm. Our approach effectively selects semantically relevant features, achieving high clustering quality as evidenced by the Dunn Index (39.45–42.26), CH Index (0.678–0.893), and Davies-Bouldin Index (0.678–0.927) on different standard dataset used for clustering. The method proves to be robust, efficient, and suitable for real-time, unsupervised streaming data analysis, representing a significant advancement in dynamic feature selection.

Keywords: Big Data, Drift, Feature Selection, Grey Wolf Optimization, Swarm Intelligent

1. INTRODUCTION

In the modern digital landscape, the production of data generation has reached record levels, giving rise to the phenomenon known as Big Data. This vast volume of data is produced continuously from diverse sources such as social media platforms, sensor networks, financial transactions, and online user activities. Unlike traditional static datasets, Big Data is characterized by its velocity, volume, and variety, requiring real-time or near-real-time processing to extract meaningful insights. Among the various forms of Big Data, stream data has gained significant attention due to its continuous flow and rapid generation. Stream data refers to real-time, high-velocity data streams that arrive sequentially from distributed sources, which requires the immediate analysis to support timely decision-making processes in applications like fraud detection, network security, financial market analysis, and social media monitoring.

One of the key challenges associated with stream data is its dynamic nature[1]. Unlike static datasets, stream data is subject to constant change, making it difficult for traditional data analysis models to remain effective over time. This phenomenon, known as concept drift[2], involves gradual or sudden shifts in the underlying data distribution, which can significantly degrade the performance of predictive models if not properly managed. Concept drift requires continuous model adaptation and updating to ensure sustained accuracy, presenting unique challenges in data mining and machine learning tasks[2], [3].

Feature optimization is an essential component of data preprocessing, aiming to identify and select the most relevant features for analysis. Effective feature optimization reduces the dimensionality of large datasets, minimizes redundant or irrelevant information, improves computational efficiency, and enhances the interpretability of models[4], [5], [6].

In the context of stream data, feature selection becomes even more critical due to the need for rapid processing and adaptability. However, conventional feature selection techniques are often designed for static datasets and lack the capability to adapt dynamically to evolving data streams and concept drift. This limitation highlights the need for advanced, adaptive feature optimization methods specifically suited for streaming environments.To address these challenges, the field of swarm intelligence has emerged as a promising paradigm for optimization problems. Inspired by natural collective behaviors—such as bird flocking, ant colony foraging, and bee swarming—swarm intelligence techniques mimic these behaviors to explore complex search spaces efficiently. These methods are inherently distributed, scalable, and adaptable, making them well-suited for real-time decision-making in dynamic environments like data streams. Algorithms such as Particle Swarm Optimization (PSO)[7], [8], Ant Colony Optimization (ACO)[9], [10], and Grey Wolf Optimization (GWO)[4], [11] have shown promising results in various optimization tasks, including feature optimizations.

Motivated by these considerations, this research proposes a novel feature selection method based on an improved swarm intelligence technique—specifically, an enhanced binary Grey Wolf Optimization (GWO) algorithm. The proposed approach introduces several modifications, including dynamic parameter adjustment, an elite solution guiding mechanism, and separate exploration and exploitation phases used for streaming data. These enhancements enable the algorithm to adaptively search for and select the most relevant features in the presence of data drift, thereby maintaining high clustering and classification performance while reducing computational overhead.

The paper is structured as follows: Section II provides background information on feature selection and Grey Wolf Optimization. Section III reviews related work carried out in this area, highlighting existing approaches and findings. The proposed methodology is presented in Section IV, detailing our approach and contributions. Section V discusses the experimental results and their implications. Finally, Section VI offers concluding remarks and potential directions for future research

2. BACKGROUND

2.1. Feature Selection

Several approaches have been explored in the literature for feature optimization, broadly categorized into four main types[4], [5], [12], [13]: filter-based, wrapper-based, embedded, and swarm intelligence-based methods.

Filter-Based Methods evaluate features independently using relevance scores like mutual information or chi-square, and select features based on thresholds. They are fast and scalable but may result in suboptimal subsets, as they do not consider classifier performance or set arbitrary thresholds. Wrapper-Based Methods evaluate feature subsets by training classifiers directly, often through techniques like forward or backward selection. They are more accurate but computationally intensive due to repetitive training and evaluation. Embedded Methods incorporate feature selection within the classifier's training process, such as decision trees, offering efficiency but limited to specific algorithms.

Swarm Intelligence-Based Methods are inspired by natural decentralized systems such as ant colonies, bird flocking, and fish schooling. These techniques leverage a population of agents that follow simple rules and interact locally to collectively find near-optimal solutions. Swarm algorithms like Particle Swarm Optimization and Grey Wolf Optimization explore the solution space by iteratively updating agent positions based on fitness evaluations, demonstrating effectiveness in handling complex, high-dimensional feature selection problems.

2.2. Grey wolf optimizer

Gray Wolf Optimization (GWO) is a relatively recent metaheuristic inspired by the social structure and hunting behavior of gray wolves, introduced by Seyedali Mirjalili in 2014 [9]. One of its main advantages over other swarm intelligence techniques is its simple design, requiring minimal external parameters and only a few derivatives, which makes it easier to implement and tune. GWO mimics the hierarchical structure of a wolf pack [9], [10], where each wolf symbolizes a potential solution to an optimization problem.

The pack is organized into a hierarchy. The topmost wolf is the Alpha, followed by the Beta, and Delta, with Omega wolves at the lowest rank. The Alpha is considered the most dominant and powerful member, representing the current best solution, while the Beta is the second most influential, assisting in decision-making and taking charge when the Alpha is absent. The Delta and Omega wolves are viewed as weaker members, serving roles such as scouts, elders, or caretakers within the pack. In the GWO algorithm, each wolf updates its position based on the locations of the alpha, beta, and delta wolves, following three primary steps: encircling the prey, hunting, and attacking the prey. These steps simulate the pack's natural hunting strategy and guide the exploration of the solution space towards optimal or near-optimal solutions. This hierarchical and behavior-inspired approach enables GWO to effectively balance exploration and exploitation during the optimization process.

3. RELEATED WORK

Numerous research efforts have been directed toward developing effective feature optimization techniques for evolving data streams. These methods aim to address the challenges of high dimensionality, concept drift, and real-time processing inherent in streaming data environments.

One notable work is the Velocity-Guided Grey Wolf Optimization Algorithm with Adaptive Weights and Laplace Operators proposed in [14]. This method introduces a new weighting mechanism within the Grey Wolf Optimization (GWO) framework to enhance feature selection for data classification tasks. The control parameters in this algorithm change non-linearly to facilitate a seamless transition between exploration and exploitation phases, improving search efficiency. However, it has been evaluated on static datasets without incorporating drift detection mechanisms, which limits its applicability in real-time stream scenarios where concept drift is prevalent. In [15], a Streaming Feature Selection approach for multi-label datasets is presented. This method employs a dynamic sliding window and feature repulsion loss to select relevant features adaptively. The algorithm integrates adaptive windowing (ADWIN) for drift handling and uses both online and offline decision tree classifiers. The approach stores the best feature subsets within fixed sliding windows to ensure robustness against changing data distributions. The method effectively addresses multi-label classification in streaming environments, although it is primarily designed for fixed or dynamically updated window contexts, not explicitly for concept drift detection.

Another significant contribution is [16], where a general framework based on dynamic multi-objective evolutionary algorithms for handling feature drift in data streams. This filter-based approach operates on sliding windows and employs evolving multi-objective optimization to adapt feature subsets over time. While the framework facilitates handling feature drift effectively, the process tends to incur higher processing times, and integration of drift detection mechanisms is an area identified for future improvement. Additionally, in [10], an Ant Colony Optimization (ACO)-based streaming feature selection method is applied to medical image diagnosis. This hybrid approach uses symmetric uncertainty to guide feature selection but operates in an offline manner and does not incorporate explicit drift detection strategies. While effective for static datasets or batch processing, its lack of online capabilities limits deployment in streaming data applications where real-time adaptation is crucial. Lastly, in the context of clustering data streams, [16] proposes an evolutionary algorithm for dynamic clustering with a variable number of clusters. The approach employs the PHT (probabilistic heuristic tree) to detect concept drift and automatically estimate the optimal number of clusters in an online fashion. Incoming data points are assigned to appropriate clusters dynamically, with the process continually updating based on quality metrics. Although effective in automatically adjusting to data distribution changes, the focus is primarily on clustering rather than feature optimization.

4. PROPOSED SYSTEM

The materials and methods section should contain sufficient detail so that all procedures can be repeated. It may be divided into headed subsections if several methods are described.

4.1. Overall Proposed Work

This work introduces a method called MBGWOSFS – modified binary grey wolf optimization for stream feature selection, to select important features from streaming data that changes over time. The process starts by taking raw, unstructured data, such as text, and cleaning it. This includes converting all text to lowercase, removing special symbols, digits, and unnecessary words to reduce noise, and filtering out very short or very long words. The cleaned data is then converted into numerical form using Word2Vec word embedding method[17], which makes it easier for the computer to understand. Since data arrives continuously, the system uses a sliding window approach, dividing the data stream into batches based on time. The first batch is used to train a model, and this initial training helps set a performance baseline. Clustering is performed on this first batch to group similar data points, and the model is saved for future comparisons. For each new batch, the saved model is re-evaluated to check if it still performs well. The system uses certain metrics to measure the model's accuracy. If the performance drops below a set threshold, it indicates that the data is changing, and the model needs to be retrained to stay accurate.

Because the number of clusters in the data can change over time, the system dynamically determines the best number of clusters for each batch using a scoring method called the Silhouette score. This allows the clustering to adapt to new trends or patterns in the data. Overall, this approach helps keep the feature selection and predictive model accurate and efficient, even as the data continuously changes, by regularly updating and adjusting to new information in real-time.

4.2. Proposed Feature Optimization

We have proposed a significant enhancement to the conventional binary Grey Wolf Optimization (GWO) algorithm for feature selection in stream clustering applications. As shown in Figure-1, although GWO is known for its simplicity, having relatively few control parameters, and maintaining a good balance between exploration and exploitation, it still exhibits certain limitations that can restrict its effectiveness in highly dynamic and complex environments such as data streams. These limitations include a tendency to prematurely converge to local optima, a decline in solution diversity over the course of iterations, restricted exploration capabilities particularly in high-dimensional search spaces that may not adapt well to the rapid changes and evolving patterns typical in streaming data scenarios.
To overcome these challenges and improve the robustness, flexibility, and overall performance of GWO for feature selection in stream clustering, we introduce a comprehensive set of modifications. These enhancements focus on key aspects of the optimization process that are particularly critical in streaming environments. The proposed method introduces several key enhancements to the existing Grey Wolf Optimization (GWO) algorithm for feature selection in streaming data environments. First, it

incorporates a dynamic scale factor assigned to each wolf, which is randomly initialized and updated in each iteration. This approach enables individual wolves to explore different regions of the search space independently, thereby potentially increasing coverage during the exploration phase.Secondly, the algorithm introduces the concept of an elite wolf—an individual that stores the best feature subset found so far. During the exploitation phase, wolves are guided towards this elite wolf, ensuring that valuable features are effectively exploited and preserved. This focus accelerates convergence and enhances the quality of the solutions. Finally, the algorithm explicitly separates exploration and exploitation phases. In the initial stage, the wide-ranging search explores various promising regions, while the subsequent focused refinement hones in on the elite solution. This structured approach effectively identifies important features eliminating noisy and unimportant features which improves the performance of feature selection method in stream environment. Proposed method reduces reliance on fixed parameters by dynamically adjusting the control parameter ‘a’. This adaptability maintains a balanced exploration-exploitation trade-off throughout the streaming process, improving robustness and responsiveness to data changes.The fitness function Eq.(3) , is a key part of the method, guiding the search toward the best solutions by evaluating each candidate, or wolf, based on two main factors. First, it measures clustering validity Eq.(1), using the Dunn Index, which assesses how well-separated and cohesive the clusters are created. Higher Dunn Index values indicate better clustering, Second, it penalizes larger feature subsets to avoid overly complex models that can lead to overfitting Eq.(2). The interaction between these factors ensures that better clustering performance results in lower penalties, promoting solutions with strong, meaningful clusters and concise feature sets. Overall, this balance helps identify optimal features for effective, interpretable streaming data clustering.

$$Fitness_{Score} = 1 - \frac{Dunn_{index} - Dunn_{min}}{Dunn_{max} - Dunn_{min}} - \frac{Feature_{selected}}{Feature_{max}}, \quad if\ dunn_{max} \neq dunn_{min}\ else\ 0 \quad (1)$$

Fig. 1: Proposed Modification in BGWO

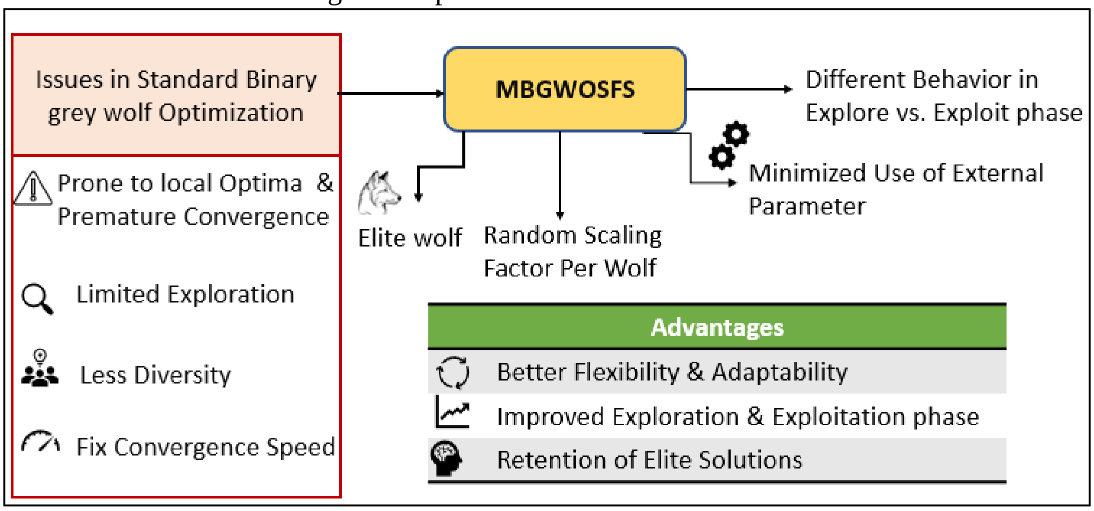
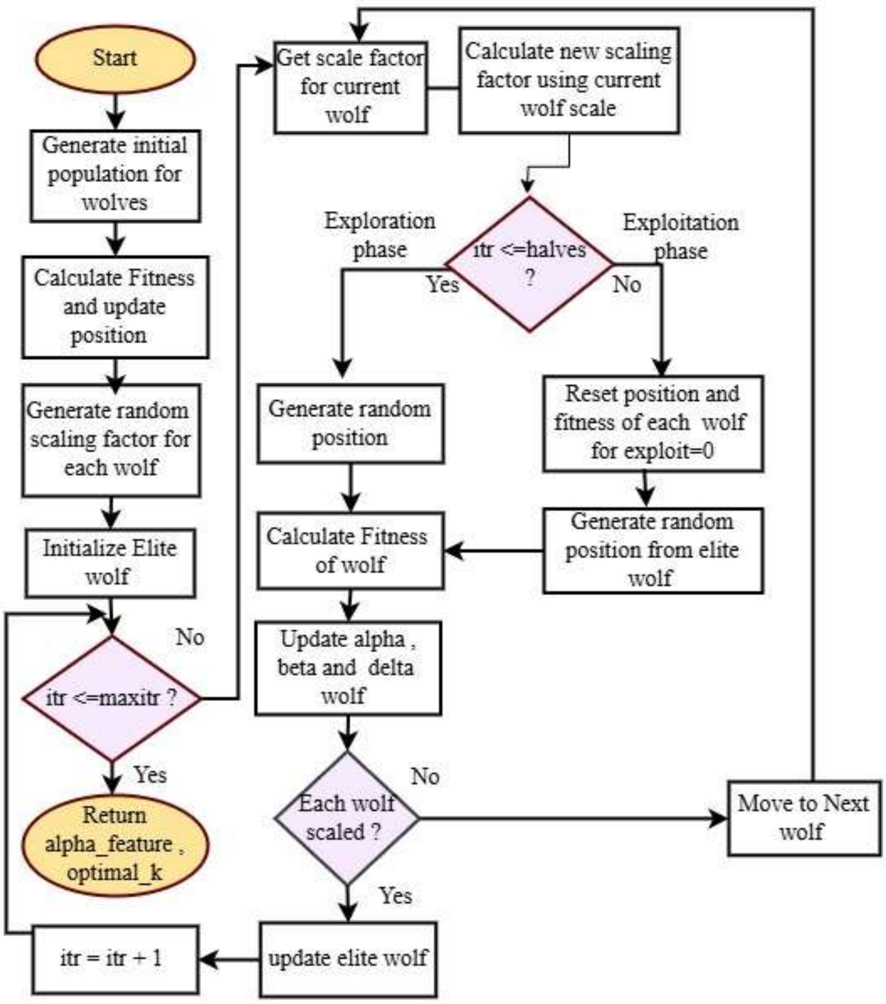


Fig. 2: Flow of Propose Feature Selection Method



The overall flow of the proposed method is illustrated in Figure 2. The process begins with initializing the population of wolves. Based on their fitness values, the alpha, beta, and delta wolves are identified. Each wolf is assigned a random dynamic scaling factor, which is updated in a controlled manner during each iteration.

During the exploration phase, new positions for the wolves are generated randomly based on their respective scaling factors, allowing diverse search directions across the entire search space. The alpha, beta, and delta wolves are updated accordingly to reflect the current best solutions. In the exploitation phase, the positions of the wolves are re-initialized, and their search is guided by the elite (best) wolf. The scaling factors influence the search, which is now focused within a reduced search space cantered around the elite, promoting efficient local refinement. This structured flow enables effective feature selection by balancing broad exploration with focused exploitation.

5. RESULT ANALYSIS AND DISCUSSION

The experiments were conducted in a streaming environment using Apache Spark 3.2.2 and Python 3.10.14 on a system with 8GB of RAM and a 500GB HDD. Spark was selected for its ability to efficiently handle real-time data processing. A 5-millisecond time-based sliding window was used to continuously collect data, from which the first 500 features were used from each window.

The comprehensive results across various large-scale datasets—namely 20Newsgroup, BBC News, AGNews, and Reuters-21578—demonstrate the robustness and efficiency of the proposed MBGWOFS method. We have proposed our method using internal validation matrix – Dunn Index, Calinski-Harabasz Index, Davies Bouldin Index, silhouette score, compactness(CP), and separation(SP).

On the 20Newsgroup dataset, MBGWOFS, which utilizes only 71 features, consistently achieved high clustering quality with a Dunn Index of approximately 42.26. This performance is notably close to that of other advanced methods that employ over 130 features, emphasizing the method's ability to maintain effective cluster separation despite a substantial reduction in feature dimensionality. Similarly, on the BBC News dataset, MBGWOFS selected only 103 features but attained a Dunn Index around 41.42, indicating good cluster cohesion and separation while significantly decreasing the computational load associated with managing large feature sets. The results on AGNews further reinforce this trend, with MBGWOFS using merely 54 features to produce a Dunn Index of approximately 39.45. Although slightly lower than methods using larger feature subsets such as BWOA and BCSO, it still offers acceptable clustering performance, demonstrating the method's ability to balance feature sparsity and clustering efficacy effectively. On the Reuters-21578 dataset, which is known for its complexity and high feature dimensionality, MBGWOFS maintained high clustering quality with a Dunn Index around 39.45 despite only selecting 54 features. This balance between reducing feature dimensions and preserving high cluster validity highlights the strength of the proposed approach in large-scale, high-dimensional streaming data environments. Furthermore, while some methods like BWOA and BCSO outperform MBGWOFS in raw clustering metrics, they do so at the cost of utilizing substantially larger feature subsets—often three to four times more. MBGWOFS consistently offers a favorable trade-off by achieving competitive clustering performance with a minimal number of features, making it inherently more efficient. This efficiency is particularly advantageous for real-time, scalable applications where computational resources are limited, and rapid processing is crucial. Overall, the extensive evaluation across these datasets conclusively demonstrates that MBGWOFS is capable of delivering high-quality clustering, strongly supporting its suitability for dynamic, large-scale data environments where reducing feature complexity without sacrificing accuracy is paramount.

The results clearly indicate that the proposed feature selection method consistently outperforms existing approaches across multiple datasets, including 20Newsgroups, BBC News, AGNews, and Reuters-21578. It achieves comparable or superior clustering quality—measured by indices such as Dunn Index, CH Index, and Davies-Bouldin Index—while using significantly fewer features. This demonstrates that our approach not only enhances clustering performance but also improves efficiency and effectiveness in high-dimensional, streaming data environments.

The manual analysis of the 20Newsgroup dataset shows that the proposed feature selection method effectively groups semantically related features. The top 15 features per cluster, as shown in Figure 7., highlight meaningful themes. Figure 8 shows the distribution of features among the cluster to check how cluster is formed . Cluster 0 includes terms like "ron," "clark," and "norris," indicating personal or celebrity topics; Cluster 1 contains words like "outplaying" and "undefeated," related to sports; and Cluster 2 features "mathematics," "university," and "waterloo," related to education and academia.

The feature distribution across clusters is balanced, with no small or noisy groups, demonstrating the method's ability to filter irrelevant features and produce cohesive, meaningful clusters for streaming data environments, Figure 8.Overall, proposed MBGWOSFS with relatively small feature set demonstrates competitive performance across datasets. Its ability to produce meaningful clusters with fewer features underscores its effectiveness in capturing semantic nuances, providing a good balance between feature dimensionality and clustering quality.

Table 1 *: Performance Analysis on 20NewsGroup dataset*

Method	Selected Feature	Dunn index	DB index	CH index	Compactness	Separation
NoFS	--	38.9207	1.1664	0.7362	0.1053	0.0064
BGWO	130	40.4619	1.355	0.5453	0.0221	0.0042
BPSO	137	38.9346	1.4071	0.8868	0.0372	0.0172
Variance	241	40.3072	1.1334	0.7828	0.2871	0.0198
BCSO	248	43.1287	1.2954	0.5967	0.0052	0.0052
BWOA	269	45.93	1.2247	0.6676	0.006	0.006
MBGWOFS	71	42.2649	1.0995	0.8929	0.0051	0.0151

Table 2 : Performance Analysis on BBC News dataset

Method	Selected Feature	Dunn index	DB index	CH index	Compactness	Separation
NoFS	--	40.3521	1.1957	0.7004	0.0452	0.00540
BGWO	98	40.8298	1.3198	0.5748	0.0402	0.0046
BPSO	151	41.2535	1.3573	0.5436	0.0034	0.0047
Variance	257	40.565	1.1251	0.7944	0.0569	0.0195
BCSO	258	43.1287	1.2954	0.5967	0.0052	0.0052
BWOA	250	43.6702	1.2775	0.6136	0.0055	0.0055
MBGWOFS	103	41.4202	1.1368	0.905	0.0144	0.0148

Table 3 : Performance Analysis on AGNews dataset

Method	Selected Feature	Dunn index	DB index	CH index	Compactness	Separation
NoFS	--	39.8373	1.1664	0.7362	0.0097	0.0063
BGWO	154	38.2608	1.1004	0.8307	0.0136	0.0213
BPSO	153	38.0701	1.0588	0.8269	0.0306	0.0038
Variance	262	38.5579	1.1336	0.7825	0.0134	0.0199
BCSO	257	45.215	1.2395	0.6517	0.0054	0.0054
BWOA	232	44.4797	1.2494	0.6415	0.0058	0.0058
MBGWOFS	79	39.3136	1.1334	0.7828	0.0188	0.0198

Table 4 : Performance Analysis on Reuters-21578 dataset

Method	Selected Feature	Dunn index	DB index	CH index	Compactness	Separation
NoFS	----	38.6779	0.7641	0.0069	0.0069	0.0159
BGWO	134	38.8208	0.9524	0.567	0.0314	0.0131
BPSO	236	38.8964	0.8934	0.6448	0.0468	0.0159
Variance	239	38.4306	0.9269	0.5991	0.0371	0.0166
BCSO	253	44.4915	1.0014	0.5112	0.0049	0.0034
BWOA	240	45.3504	1.246	0.645	0.0064	0.0066
MBGWOFS	54	39.4528	0.6779	0.7641	0.0566	0.0569

Fig. 3 : Average Number of Selected Feature

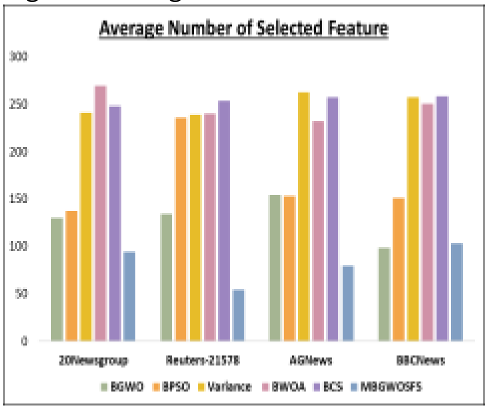


Fig. 6 : C-H index

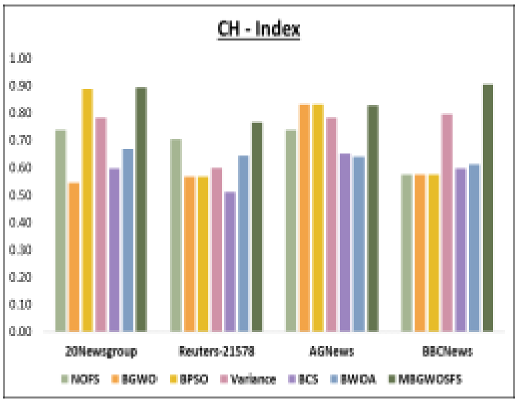


Fig. 4 : Dunn Validity Index

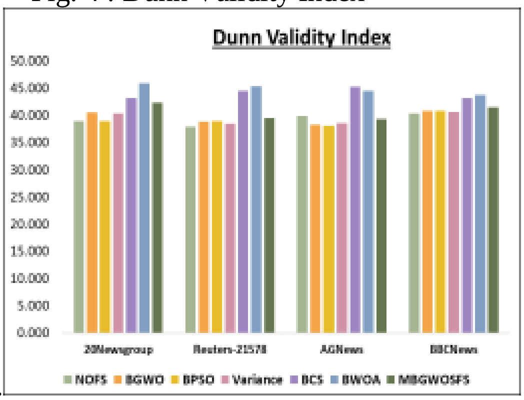
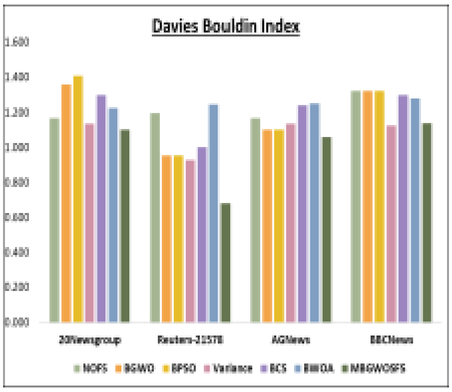


Fig. 5 : Davies Bouldin Index

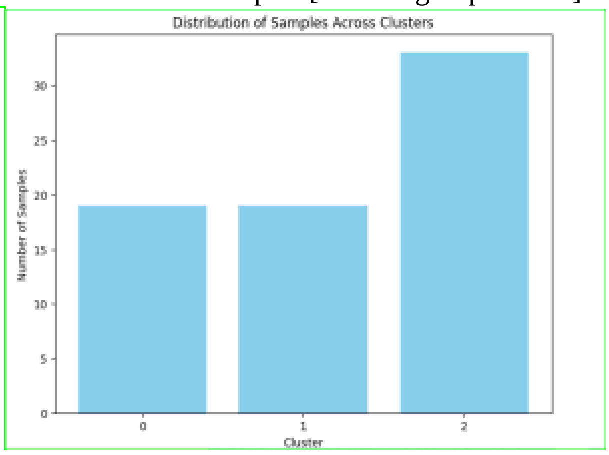


The results clearly indicate that the proposed feature selection method consistently outperforms existing approaches across multiple datasets, including 20Newsgroups, BBC News, AGNews, and Reuters-21578. It achieves comparable or superior clustering quality—measured by indices such as Dunn Index, CH Index, and Davies-Bouldin Index—while using significantly fewer features. This demonstrates that our approach not only enhances clustering performance but also improves efficiency and effectiveness in high-dimensional, streaming data environments.

Fig. 7 : Top -15 Selected feature

words	word_vectors	cluster
[ron]	[-1.1997461115242...	0
[outplaying]	[4.24116005888208...	1
[beg]	[7.20918178558349...	2
[differ]	[-6.0677860165014...	2
[clark]	[1.90078499144874...	0
[deserved]	[-9.0956402709707...	2
[gilmour]	[2.57207866525277...	1
[fast]	[7.98960565589368...	2
[break]	[7.02903838828206...	2
[net]	[-6.4237450715154...	2
[opportunites]	[-8.5208186646923...	1
[reminded]	[4.18436547988676...	0
[old]	[-4.4152857299195...	2
[playing]	[8.00832873210310...	1
[look]	[3.59920493792742...	0

Fig. 8 : distribution of samples [20Newsgroup dataset]



6. CONCLUSION AND FUTURE WORK

This research introduces an efficient and robust method for stream clustering in unstructured data environments using an improved Binary Grey Wolf Optimization (MBGWOSFS). By incorporating features like dynamic scaling, minimal dependence on external parameters, and elite solution retention, the approach effectively handles concept drift and optimizes feature selection without requiring labeled data. Experimental results across multiple datasets demonstrate its ability to produce high-quality, well-separated, and compact clusters while significantly reducing the number of features. Overall, the proposed method proves to be scalable, accurate, and semantically aware, making it suitable for real-time, unstructured data analysis. Future directions include enhancing diversity strategies, exploring advanced embedding techniques, and extending the framework to multi-modal and more complex data streams. Future work could focus on integrating deep learning techniques to further enhance feature representation and selection in streaming environments, enabling more accurate and robust analysis. Expanding the approach to handle multi-modal data—such as combining text, images, and videos—would increase its applicability to diverse real-world scenarios. Additionally, deploying the method within distributed or edge computing frameworks can support real-time processing of large-scale streaming data, making it more scalable and efficient. Finally, applying and testing the proposed approach across various domains like social media analytics, sensor networks, and financial markets can help validate its versatility and effectiveness in different application settings, paving the way for broader adoption in dynamic, high-volume data environments

REFERENCES

[1] N. AlNuaimi, M. M. Masud, M. A. Serhani, and N. Zaki, “Streaming feature selection algorithms for big data: A survey,” *ACI*, vol. 18, no. 1/2, pp. 113–135, Mar. 2022, doi: 10.1016/j.aci.2019.01.001.

[2] D. Zhao and Y. S. Koh, “Feature Drift Detection in Evolving Data Streams,” in *Database and Expert Systems Applications*, vol. 12392, S. Hartmann, J. Küng, G. Kotsis, A. M. Tjoa, and I. Khalil, Eds., in Lecture Notes in Computer Science, vol. 12392. , Cham: Springer International Publishing, 2020, pp. 335–349. doi: 10.1007/978-3-030-59051-2_23.

[3] X. H. Dang, V. C. S. Lee, W. K. Ng, and K. L. Ong, “Incremental and Adaptive Clustering Stream Data over Sliding Window,” in *Database and Expert Systems Applications*, vol. 5690, S. S. Bhowmick, J. Küng, and R. Wagner, Eds., in Lecture Notes in Computer Science, vol. 5690. , Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 660–674. doi: 10.1007/978-3-642-03573-9_55.

[4] A. K. Abasi, A. T. Khader, M. A. Al-Betar, S. Naim, S. N. Makhadmeh, and Z. A. A. Alyasseri, “An Improved Text Feature Selection for Clustering Using Binary Grey Wolf Optimizer,” in *Proceedings of the 11th National Technical Seminar on Unmanned System Technology 2019*, vol. 666, Z. Md Zain, H. Ahmad, D. Pebrianti, M. Mustafa, N. R. H. Abdullah, R. Samad, and M. Mat Noh, Eds., in Lecture Notes in Electrical Engineering, vol. 666. , Singapore: Springer Nature Singapore, 2021, pp. 503–516. doi: 10.1007/978-981-15-5281-6_34.

[5] I. Al-Jadir and W. A. Mahmoud, “A Grey Wolf Optimizer Feature Selection method and its Effect on the Performance of Document Classification Problem,” *J.port.sci.res.*, vol. 4, no. 2, pp. 116–122, Dec. 2021, doi: 10.36371/port.2021.2.9.

[6] N. Almusallam, Z. Tari, J. Chan, A. Fahad, A. Alabdulatif, and M. Al-Naeem, “Towards an Unsupervised Feature Selection Method for Effective Dynamic Features,” *IEEE Access*, vol. 9, pp. 77149–77163, 2021, doi: 10.1109/ACCESS.2021.3082755.

[7] S. Fong, R. Wong, and A. V. Vasilakos, “Accelerated PSO Swarm Search Feature Selection for Data Stream Mining Big Data,” *IEEE Trans. Serv. Comput.*, vol. 9, no. 1, pp. 33–45, Jan. 2016, doi: 10.1109/TSC.2015.2439695.

[8] M. Jain, V. Saihjpal, N. Singh, and S. B. Singh, “An Overview of Variants and Advancements of

PSO Algorithm," *Applied Sciences*, vol. 12, no. 17, p. 8392, Aug. 2022, doi: 10.3390/app12178392.

[9] C. Fahy, S. Yang, and M. Gongora, "Ant Colony Stream Clustering: A Fast Density Clustering Algorithm for Dynamic Data Streams," *IEEE Trans. Cybern.*, vol. 49, no. 6, pp. 2215–2228, Jun. 2019, doi: 10.1109/TCYB.2018.2822552.

[10] L. G. Fahad, S. F. Tahir, W. Shahzad, M. Hassan, H. Alquhayz, and R. Hassan, "Ant Colony Optimization-Based Streaming Feature Selection: An Application to the Medical Image Diagnosis," *Scientific Programming*, vol. 2020, pp. 1–10, Oct. 2020, doi: 10.1155/2020/1064934.

[11] E. Emary, H. M. Zawbaa, and A. E. Hassanien, "Binary grey wolf optimization approaches for feature selection," *Neurocomputing*, vol. 172, pp. 371–381, 2016, Accessed: Apr. 26, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231215010504

[12] N. AlNuaimi, M. M. Masud, M. A. Serhani, and N. Zaki, "Streaming feature selection algorithms for big data: A survey," *ACI*, vol. 18, no. 1/2, pp. 113–135, Mar. 2022, doi: 10.1016/j.aci.2019.01.001.

[13] N. Almusallam, Z. Tari, J. Chan, and A. AlHarthi, "UFSSF - An Efficient Unsupervised Feature Selection for Streaming Features," in *Advances in Knowledge Discovery and Data Mining*, vol. 10938, D. Phung, V. S. Tseng, G. I. Webb, B. Ho, M. Ganji, and L. Rashidi, Eds., in Lecture Notes in Computer Science, vol. 10938. , Cham: Springer International Publishing, 2018, pp. 495–507. doi: 10.1007/978-3-319-93037-4_39.

[14] L. Zhang and X. Chen, "A Velocity-Guided Grey Wolf Optimization Algorithm With Adaptive Weights and Laplace Operators for Feature Selection in Data Classification," *IEEE Access*, vol. 12, pp. 39887–39901, 2024, doi: 10.1109/ACCESS.2024.3376235.

[15] Y. Li and Y. Cheng, "Streaming Feature Selection for Multi-Label Data with Dynamic Sliding Windows and Feature Repulsion Loss," *Entropy*, vol. 21, no. 12, p. 1151, Nov. 2019, doi: 10.3390/e21121151.

[16] S. Sahmoud and H. R. Topcuoglu, "A general framework based on dynamic multi-objective evolutionary algorithms for handling feature drifts on data streams," *Future Generation Computer Systems*, vol. 102, pp. 42–52, Jan. 2020, doi: 10.1016/j.future.2019.07.069.

[17] D. S. Asudani, N. K. Nagwani, and P. Singh, "Impact of word embedding models on text analytics in deep learning environment: a review," *Artif Intell Rev*, vol. 56, no. 9, pp. 10345–10425, Sep. 2023, doi: 10.1007/s10462-023-10419-1.