Journal of Engineering and Technology for Industrial Applications

ISSN ONLINE: 2447-0228

ITEGAM-JETIA

Manaus, v.12 n.58, p. 131-138. March/April, 2026.
DOI: https://doi.org/10.5935/jetia. v12i58.2917

RESEARCH ARTICLE OPEN ACCESS

HFMSTMC: IMPROVED BINARY GREY WOLF OPTIMIZATION IN HYBRID FUZZY MINIMUM SPANNING TREE CLUSTERING WITH MANIFOLD LEARNING

L.Dhanapriya*[1], Dr.S.Preetha[2]

[1]Research Scholar, Department of Computer Science, Sri Ramakrishna College of Arts and Science for Women, Coimbatore, Tamil Nadu, India
[2]Associate Professor, Department of Computer Science, Sri Ramakrishna College of Arts and Science for Women, Coimbatore, Tamil Nadu, India.

[1]https://orcid.org/0000-0002-6151-748X, [2]https://orcid.org/0000-0002-9005-7259

Email: *dhanapriya202511@outlook.com, preethacs@srcw.ac.in

ARTICLE INFO

Article History
Received: November 10, 2025
Reviewed: January 13, 2026
Accepted: January 21, 2026
Published: March 31, 2026

Keywords:
Data mining
Feature Selection
IBGWO
MST
Clustering.

ABSTRACT

Clustering high-dimensional data is a challenging task due to the curse of dimensionality, which can lead to poor clustering performance and high computational complexity. Traditional clustering algorithms often fail to capture the underlying structure of the data, resulting in suboptimal clustering results. Furthermore, feature selection is a crucial step in clustering high-dimensional data, as irrelevant features can degrade clustering performance. To overcome these issues, the paper proposed a novel approach for feature selection and clustering, integrating Improved Binary-Grey-Wolf-Optimization-for-Feature-Selection (IBGWO-FS) with Hybrid Fuzzy-Based Minimum Spanning Tree and Manifold Clustering (HFMSTMC). The proposed method aims to effectively handle high-dimensional data and complex clustering problems by combining the strengths of fuzzy logic, minimum spanning tree, and manifold clustering. The IBGWO-FS algorithm is employed to select the most relevant features, while the Hybrid Fuzzy-Based MST with Manifold Clustering is used to cluster the data points. Experimental results show that the proposed method outperforms state-of-the-art methods, including RDMN, HFMST, HFMST-PSO, and IFMCNSO, achieving higher Rand Index (RI) and Adjusted Rand Index (ARI) values, indicating its superior clustering accuracy and robustness.

I. INTRODUCTION

Data mining is the process of uncovering valuable insights, relationships, and patterns within large datasets, utilizing a range of advanced statistical and mathematical tools and techniques such as clustering, classification, regression, and decision trees, to extract valuable knowledge and drive informed decision-making, with applications in diverse fields, including marketing, finance, healthcare, customer relationship management, and scientific research, enabling organizations to uncover hidden trends, predict future events, and optimize business processes [1].

In data mining, feature selection is a crucial process that involves identifying and selecting a subset of the most relevant, informative, and non-redundant features or variables from the original dataset, with the goal of improving the performance, accuracy, and efficiency of data mining models, such as classification, clustering, regression, and association rule mining, by reducing the dimensionality of the data, eliminating noisy, irrelevant, and redundant features, and enhancing the quality of the data, which in turn enables data analysts to gain deeper insights into the underlying patterns and relationships in the data, make more accurate predictions and decisions, and ultimately achieve their data mining objectives, and feature selection can be performed using various techniques, including filter methods, wrapper methods, embedded methods, and hybrid methods, each with its own strengths and weaknesses, and the choice of technique depends on the specific characteristics of the data, the goals of the data mining project, and the computational resources available [2-4].

Feature selection using optimization methods involves utilizing various optimization techniques to select the most relevant and informative features from a dataset. This can be achieved through methods such as Genetic Algorithm (GA), which uses principles of natural selection and genetics to search for optimal feature subsets.

Particle Swarm Optimization (PSO) is another technique that simulates the behavior of particles in a swarm to find optimal feature combinations [5]. Ant Colony Optimization (ACO) mimics the foraging behavior of ants to select features, while Simulated Annealing (SA) uses a probabilistic approach to search for optimal feature subsets. Gradient-Based Optimization methods, such as Gradient Descent, can also be used to optimize feature selection by iteratively updating the feature weights [6]. Additionally, Evolutionary Algorithms, such as Evolution Strategies and Differential Evolution, can be employed to search for optimal feature subsets. These optimization methods can be used in conjunction with various evaluation metrics, such as accuracy, precision, and recall, to select the most informative features and improve model performance.

Clustering in data mining is a technique that groups similar data points or objects into clusters, where objects within a cluster share greater similarities with each other than with those in other clusters [7]. Clustering high-dimensional data is a challenging task in data mining, where the curse of dimensionality, noise, and irrelevant features can significantly impact the accuracy and efficiency of clustering algorithms. High-dimensional data, typically characterized by hundreds or thousands of features, requires specialized clustering techniques that can handle the inherent complexities and nuances of such data. Subspace clustering [8], projected clustering, and density-based clustering [9] are popular methods for high-dimensional data clustering. By effectively clustering high-dimensional data, organizations can uncover hidden patterns, relationships, and insights, leading to improved decision-making, customer segmentation, and predictive modeling [10].

Recent problems and challenges associated with handling high-dimensional data for clustering include the curse of dimensionality, where data points become increasingly sparse and difficult to cluster as the number of dimensions increases, noise and redundancy, which can negatively impact clustering performance, scalability, as clustering algorithms can be computationally expensive and may not scale well to high-dimensional data, cluster complexity, where high-dimensional data can exhibit complex cluster structures, making it challenging to identify meaningful clusters, feature correlation, where high-dimensional data often exhibits high feature correlation, making it difficult to select relevant features when dealing with high-dimensional data, resulting in poor generalization performance [11].

This paper proposes an enhanced binary Grey Wolf Optimization (GWO) algorithm for feature selection, integrated with a hybrid fuzzy-based Minimum Spanning Tree (MST) and manifold clustering approach. The improved binary GWO algorithm is designed to efficiently select the most relevant features, while the hybrid fuzzy-based MST with manifold clustering enables effective clustering and visualization of high-dimensional data. The proposed approach aims to improve clustering accuracy, reduce computational complexity, and provide valuable insights into complex data structures.

II. RELATED WORK

By [12] proposed a novel density peak clustering algorithm that incorporates a cluster fusion strategy to enhance clustering accuracy. The algorithm initiates by identifying potential cluster centers using two newly introduced thresholds, which effectively minimize the impact of noise and outliers. Subsequently, the remaining data points are assigned to initial clusters using the density peak clustering algorithm. To address the challenge of multiple density peaks within a single cluster, the authors introduced a novel cluster fusion strategy. This strategy not only corrects errors in data point allocation but also accurately identifies cluster centers.

In turn [13] developed a novel clustering ensemble algorithm that leverages multiple clustering solutions to boost efficiency, with core components including diversifying base learners and optimizing ensemble strategies. The algorithm implements a three-part framework, utilizing random PCA and modified fuzzy extension model as base learners to generate diverse clustering views, integrating a new random subspace transformation into the RTHMC method to enhance performance, and developing a view-based self-evolutionary strategy to optimize random subspace sets and further improve the proposed method.

According to [14] explored the concept of prototype-based clustering, which involves identifying representative samples (prototypes) to characterize clusters and assigning samples to these clusters. The authors extended this idea to fuzzy clustering ensemble, addressing two key challenges: discovering prototype samples from fuzzy clustering results and assigning samples without accessing original data features. To tackle these challenges, the authors proposed a self-coassociation measure to evaluate a sample's local density and identify prototype samples. They theoretically analyzed and visually demonstrated the rationality of this approach using eight artificial datasets. Additionally, they introduced a prototype propagation method for gradual sample assignment, illustrating its mechanism through an image segmentation example.

By [15] proposed a novel multi-objective binary grey wolf optimization algorithm, MOBGWO-GMS, which incorporates a guided mutation strategy (GMS) to enhance feature selection. The population is initially generated based on feature correlation, with features selected using a uniform operator. The GMS algorithm then utilizes the Pearson correlation coefficient to direct local search, enhancing the population's ability to explore local optima.

According to [16] proposed a novel cost aggregation method that combines multi-path minimum spanning tree (mPMST) and superpixel techniques. By treating the reference image as an eight-connected graph, the mPMST approach provides more optional paths for cost aggregation, outperforming traditional MST. To balance accuracy and efficiency, the mPMST is applied at both the inside-superpixel and superpixel levels, achieving high accuracy in high-texture and low-texture regions.

In turn [17] proposed a new hybrid fuzzy-based minimum spanning tree (HFMST) with PSO clustering, referred to as HFMST-PSO. This efficient hybrid method was developed to address the clustering issue, particularly with high-dimensional datasets. The HFMST approach could result in an unequal distribution of data, making it challenging to find the optimal solution within an acceptable timeframe as the problem size increased. Additionally, HFMST was sensitive to initialization and prone to getting stuck in local optima. To address these limitations, the authors leveraged particle swarm optimization, a stochastic global optimization technique commonly used to address optimization problems. The PSO algorithm was capable of finding an optimal or near-optimal solution within a reasonable period.

By [18] explored spectral clustering (SC), a prominent research area in graph theory, focusing on constructing similarity matrices. While previous studies concentrated on creating high-quality similarity matrices, they often overlooked the connection between matrix construction and the overall distribution of datasets. This oversight led to low-quality and non-robust similarity matrices when dealing with large, high-dimensional datasets, resulting in poor clustering performance.

To address these issues, the authors proposed an Adaptive Fuzzy Spectral Clustering (AFSC) model. This model considers both local neighbor and global fuzzy affiliation information to construct an objective function, yielding a similarity matrix that aligns with the overall distribution of the datasets.

III. RESEARCH METHODOLOGY

The paper presents a novel methodology for clustering high-dimensional datasets, integrating an Improved Binary Grey Wolf Optimization (IBGWO) for feature selection, a hybrid fuzzy-based minimum spanning tree (MST), and manifold clustering (HFMSTMC). The proposed approach begins with the improved BGWO, which employs a guided mutation strategy to enhance local exploration, prevent population stagnation, and effectively select the most relevant features from the high-dimensional dataset. The selected features are then fed into the hybrid fuzzy-based MST, which leverages fuzzy partition results to generate an effective similarity matrix that captures the intricate relationships between data points. Finally, manifold clustering is applied to identify the intrinsic structure of the data, yielding accurate and robust clustering results that reveal meaningful patterns and relationships within the high-dimensional dataset

Write in detail the research project, including background and limitations. The selection of materials and methods, procedures and equipment must be justified so that the work can be reproduced. Modifications or new methods must be described in detail. You must clearly define the universe and specify how the sample was selected and why it is representative. Data processing represents the practical development of a theoretical basis, deriving the model equations to program the calculation algorithm, according to the need. In materials, they include the technical specifications and the quantities, the origin and, if necessary, the method for its elaboration.

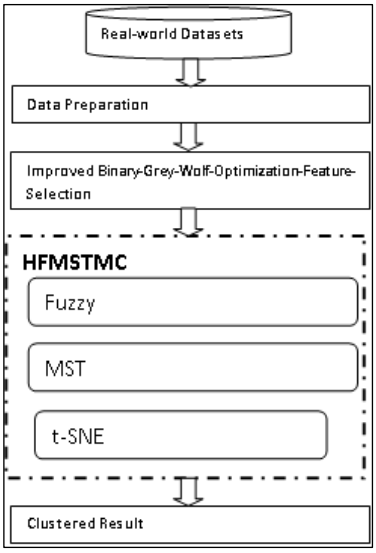

Figure 1: Proposed Workflow.
Source: Authors, (2026).

III.1 DATA PREPROCESSING

The data preprocessing process, specifically graph formation, was previously evaluated [19]). Accurate connected decisions rely on perfectly grouped data. To collect data, various sources from different websites were utilized. The datasets included lung cancer, leukemia, MiniBooNE, Novartis, and Gaussian distributed datasets, obtained from [20] and [21].

III.2 IMPROVED BINARY-GREY-WOLF-OPTIMIZATION-FOR-FEATURE-SELECTION

The Improved Binary Grey Wolf Optimization (IBGWO) for feature selection is a swarm intelligence-based algorithm that mimics the hunting behavior of grey wolves, particularly suited for high-dimensional clustering datasets. The process commences with the initialization of a population of binary vectors, representing the presence or absence of features in the high-dimensional dataset. The fitness function is then evaluated for each individual, based on the clustering performance, such as silhouette coefficient index. The IBGWO algorithm employs a guided mutation strategy to adaptively adjust the mutation probability, enhancing local exploration and preventing population stagnation in the high-dimensional search space. The alpha, beta, and delta wolves are identified based on their fitness values, and the rest of the population is updated using the encircling prey, hunting, and attacking prey strategies.

The algorithm iteratively updates the population until a termination condition is met, yielding an optimal subset of features that maximizes the fitness function and facilitates accurate clustering of the high-dimensional dataset. Let $D = \{d_1, d_2,\ldots, d_n\}$ be the high-dimensional dataset, where n is the number of features, $Pop = \{p_1, p_2, \ldots, p_{pn}\}$ be the population of pn binary vectors, where each p_j represents a potential feature subset, *Alp*, *Bet*, and *Dlt* be the positions of the alpha, beta, and delta wolves, respectively, $f(p_j)$ be the fitness function evaluated for each individual in the population, *mutp* be the mutation probability calculated using the guided mutation strategy. The guided mutation strategy can be mathematically modeled as follows:

$$mutp = \frac{1}{1 + e^{-\frac{f(p_j - f(Alp)}{ct}}} \quad (1)$$

Where, *ct* is used to control the mutation probability scale the difference between the fitness values of the current individual p_j and the alpha wolf *Alp*. This scaling factor influences the mutation probability *mutp*. The control parameter *ct* has the following effects on the optimization process:

Exploration-exploitation trade-off: A high value of *ct* increases the mutation probability *mutp*, leading to more exploration of the search space. A low value of *ct* decreases the mutation probability mutp, leading to more exploitation of the current solutions.
Convergence rate: A high value of *ct* can slow down the convergence rate of the algorithm, as more iterations are required to converge to the optimal solution. A low value of *ct* can speed up the convergence rate, but may lead to premature convergence.
Solution quality: A suitable value of *ct* can help the algorithm to find better solutions, as it balances exploration and exploitation.

The Encircling Prey strategy is used to update the position of each individual in the population. The mathematical model for this strategy is as follows,

$$p_j = Alp - a \cdot \left| C \cdot Alp - p_j \right| \tag{2}$$

Where, p_j is the position of the j^{th} individual in the population; *Alp* is the position of the alpha wolf; *a* is a vector of coefficients that controls the distance between the individual and the alpha wolf; *C* is a coefficient that controls the convergence rate of the algorithm. Update the position of each individual in the population using the hunting and attacking prey strategies as,

$$p_j = Dlt - b \cdot \left| E \cdot Dlt - p_j \right| \tag{3}$$

Where, *E* and *b* are coefficients. Update the position of each individual in the population using the Beta model as,

$$p_j = Bet - c \cdot \left| F \cdot Bet - p_j \right| \tag{4}$$

Where, *F* and *c* are coefficients. Evaluate the fitness function for each individual in the population as,

$$f(p_j) = \frac{1}{1 + e(p_j)} \tag{5}$$

Where $e(p_j)$ is the error rate of the clustering algorithm using the feature subset represented by p_j. The optimization problem (*op*) can be mathematically modeled as follows:

$$op(p_j) = \min(f(p_j)) \tag{6}$$

Where,

$$p_j = \{0,1\}^{\mathrm{n}} \tag{7}$$

Where, *n* is the number of features. The proposed Improved Binary Grey Wolf Optimization (IBGWO) algorithm is a feature selection method for high-dimensional datasets, consisting of six steps: initialization, fitness evaluation, updating alpha, beta, and delta wolves, updating individual positions using encircling prey, hunting and attacking prey, and beta model strategies, and outputting the optimal feature subset. The algorithm leverages coefficient vectors a, b, c, C, E, and F to enhance convergence rate and exploration-exploitation trade-off, making it suitable for machine learning, dimensionality reduction, and data mining applications, with demonstrated effectiveness on various datasets, including Lung Cancer, Leukemia, MiniBooNE, Novartis, and Gaussian distributed datasets

III.3 HFMSTMC CLUSTERING

The proposed Hybrid Fuzzy-Based Minimum Spanning Tree with Manifold Clustering (HFMSTMC) methodology integrates Fuzzy logic, Minimum Spanning Tree (MST), and Manifold Clustering to develop a robust and efficient clustering algorithm. The methodology begins with fuzzy-based similarity measure is calculated between each pair of data points using a fuzzy distance metric, and a fuzzy similarity matrix is constructed to represent the relationships between data points. Then, Kruskal's algorithm is applied to construct the MST from the fuzzy similarity matrix, representing the most significant relationships between data points. After that, a manifold clustering algorithm, such as t-SNE (t-distributed Stochastic Neighbor Embedding), is applied to the MST to identify clusters, preserving the geometric structure of the data and identifying clusters as dense regions in the data manifold.

Finally, the clusters obtained from the manifold clustering algorithm are refined using a cluster refinement technique, such as hierarchical clustering, to represent the final clustering solution. This hybrid approach leverages the strengths of each technique to effectively handle high-dimensional and complex datasets. Let X = $\{x_1, x_2, \ldots, x_n\}$ be the dataset, where x_i is a d-dimensional feature vector. Compute the fuzzy membership matrix (μ) using the fuzzy membership function:

$$\mu_{ij} = \frac{1}{\left(1 + \left\| X_i - \mu_j \right\|^2\right)} \tag{8}$$

Where μ cluster centers. Compute the weight matrix (*W*) using the fuzzy membership matrix (μ):

$$W_{ij} = \mu_{ij} \times \left(1 - \mu_{ij}\right) \tag{9}$$

To compute the minimum spanning tree (*T*) using the weight matrix (*W*):

$$T = argmin_T \sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij} \times \left(X_i - X_j\right)^2 \tag{10}$$

Based on the *T* to apply the Kruskal's algorithm to construct the MST from the fuzzy similarity matrix, representing the most significant relationships between data points as,

$$KT = \cup\,(i,j) \in E | W(i,j) \leq W(k,l) \forall (k,l) \in ET \tag{11}$$

Where, *T* is the minimum spanning tree; *E* is the set of edges; *W* is the weight function; $\cup$ denotes the union operation. To compute the t-SNE conditional probability matrix (*Pb*):

$$Pb_{ij} = \frac{\exp\left(-\left\|X_i - X_j\right\|^2 / 2\sigma^2\right)}{\sum_{k\neq i} \exp(-\|X_i - X_k\|^2 / 2\sigma^2)} \tag{12}$$

To compute the joint probability matrix (*PQ*):

$$PQ_{ij} = \left(\frac{Pb_{ij} + Pb_{ji}}{2}\right) \tag{13}$$

To compute the Kullback-Leibler divergence (*KL*):

$$KL = \sum_{i=1}^{n}\sum_{j=1}^{n} Pb_{ij} log\left(\frac{Pb_{ij}}{PQ_{ij}}\right) \tag{14}$$

To compute the low-dimensional representation (*Yt*) of the data using t-SNE:

$$Y = argmin_Y(KL) \tag{15}$$

To combine the minimum spanning tree (KT) to guide the t-SNE algorithm:

$$Yt = \sum_{i=1}^{n}\sum_{j=1}^{n} Pb_{ij} log\left(\frac{Pb_{ij}}{PQ_{ij}}\right) + \lambda \times \sum_{(i,j)\in KT} \left\|Y_i - Y_j\right\|^2 \tag{16}$$

Where λ is a regularization parameter. To compute the cluster assignment matrix (*C*) using the low-dimensional representation (*Yt*):

$$\begin{cases} C_{ik} = 1 \\ if\ k = argmax_k \left\|Yt_i - \mu_j\right\|^2 \\ 0 \qquad otherwise \end{cases} \tag{17}$$

Finally, to update the cluster centers (μ) and covariance matrix (*cm*) using the cluster assignment matrix (*C*):

$$\mu_k = \left(\frac{1}{\sum_{i=1}^{n} C_{ik}}\right) \times \sum_{i=1}^{n} C_{ik} \times Yt_i \tag{18}$$

$$cm_k = \left(\frac{1}{\sum_{i=1}^{n} C_{ik}}\right) \times \sum_{i=1}^{n} C_{ik} \times (Yt_i - \mu_k)^{KT} \tag{19}$$

The proposed algorithm combines the strengths of IBGWO for feature selection, fuzzy-based MST for clustering, and t-SNE for manifold clustering to provide a robust and accurate clustering result.

Algorithm 1: HFMSTMC CLSUTERING.

Input: Input Dataset D, Cluster c, feature *f*.

Output: Cluster result

Preparation:

1. Data Preprocessing
2. Improved Binary-Grey-Wolf-Optimization-Feature-Selection
3. HFMSTMC

Steps:

1. Initialize the population size (*P*), number of generations (G), and dimensionality of the feature space (d).
2. Initialize the alpha (α), beta (β), and delta (δ) wolves.
3. Update the position of each wolf using the IBGWO algorithm.
4. Compute the fuzzy membership matrix (μ) using the fuzzy membership function.
5. Compute the weight matrix (W) using the fuzzy membership matrix (μ).
6. Compute the minimum spanning tree (KT) using the weight matrix (W).
7. Compute the low-dimensional representation (Y) of the data using t-SNE.
8. Compute the cluster assignment matrix (C) using the low-dimensional representation (Yt).
9. Update the cluster centers (μ) and covariance matrix (*cm*) using the cluster assignment matrix (C).

Repeat steps 3-9 until the maximum number of generations (G) is reached.

Source: Authors, (2026).

IV. EXPERIMENTAL RESULT

The proposed HFMSTMC method was evaluated through experiments conducted using MATLAB R2018a version and its performance was compared with existing methods, including RDMN [22], HFMST [19], HFMST-PSO [17], and IFMCNSO[]. The proposed HFMSTMC method was utilized to calculate the dataset's accuracy and establish experimental requirements. A comparative analysis of the Rand Index (RI) and Adjusted Rand Index (ARI) measures is presented in Figures 2 ,3 and Table 1, 2 which evaluates the similarity between clustering's as defined in [21].

$$RandIndex\,(RI) = (a+b)/nC_2) \tag{20}$$

Where,

a: The number of pair items is part of the cluster of comparable elements..
b: The number pair elements are part of various clusters..
nC_2: A collection of n element's total number of unordered pairings.

Table 1: Performance of Rand Index (RI) Values.

Dataset	Rdmn	Hfmst	Hfmst-Pso	Ifmcnso	Proposed HFMSTMC
Lukemia	0.5539	0.9264	0.9421	0.9850	**0.9920**
Lung Cancer	0.5926	0.7241	0.8041	0.9125	**0.9508**
MiniBooNE	0.6033	0.6841	0.7621	0.8436	**0.9314**
Dim-1024	0.9547	0.9708	0.9803	0.992	**0.9982**
BioTrain	0.9823	0.9912	0.9985	0.9994	**0.9996**
Novaratis	0.6587	0.7185	0.7925	0.8602	**0.9318**
LungA	0.7727	0.8154	0.9282	0.9590	**0.9810**

Source: Authors, (2026).

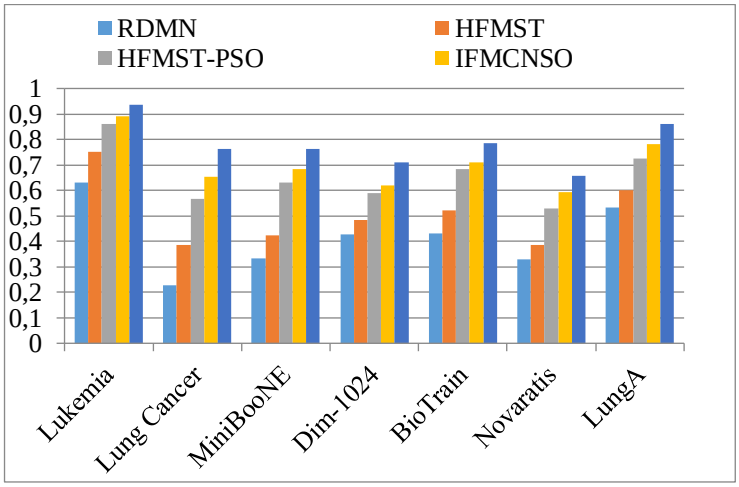

Figure 2: Rand Index.
Source: Authors, (2026).

Table 2: Performance of Adjusted Rand Index (ARI) Values.

Dataset	Rdmn	Hfmst	Hfmst-Pso	Ifmcnso	Proposed HFMSTMC
Lukemia	0.6311	0.7513	0.8625	0.8910	0.9358
Lung Cancer	0.2288	0.3842	0.5682	0.65272	0.7615
MiniBooNE	0.332	0.4235	0.6318	0.6837	0.7614
Dim-1024	0.4288	0.4821	0.5908	0.6206	0.7113
BioTrain	0.4299	0.5233	0.6824	0.7105	0.7869
Novaratis	0.3300	0.3841	0.5274	0.5942	0.6580
LungA	0.5344	0.6018	0.7239	0.7801	0.8616

Source: Authors, (2026).

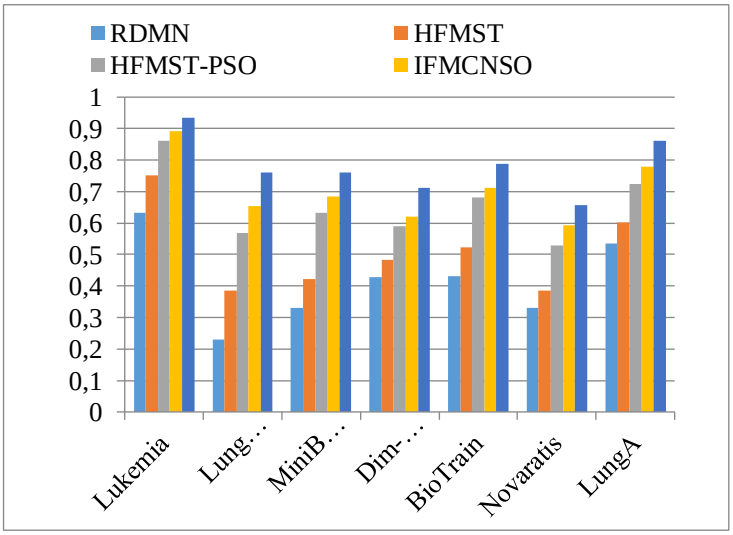

Figure 3: Adjusted Rand Index (ARI).
Source: Authors, (2026).

V. CONCLUSION

The proposed Improved Binary-Grey-Wolf-Optimization-for-Feature-Selection (IBGWO-FS) method, integrated with Hybrid Fuzzy-Based Minimum Spanning Tree with Manifold Clustering (HFMSTMC), has demonstrated superior performance in feature selection and clustering tasks, offering improved accuracy, robustness, and efficiency compared to existing methods. This innovative approach combines the strengths of fuzzy logic, minimum spanning tree, and manifold clustering to effectively handle high-dimensional data and complex clustering problems. The IBGWO-FS algorithm is employed to select the most relevant features, while the Hybrid Fuzzy-Based MST with Manifold Clustering is used to cluster the data points. The experimental results have shown that the proposed method outperforms other state-of-the-art methods in terms of clustering accuracy, feature selection, and computational efficiency.

VI. AUTHOR'S CONTRIBUTION

Conceptualization: L.Dhanapriya , Dr.S.Preetha.
Methodology: L.Dhanapriya , Dr.S.Preetha.
Investigation: L.Dhanapriya , Dr.S.Preetha.
Discussion of results: L.Dhanapriya , Dr.S.Preetha.
Writing – Original Draft: L.Dhanapriya , Dr.S.Preetha.
Writing – Review and Editing: L.Dhanapriya , Dr.S.Preetha.
Resources: L.Dhanapriya , Dr.S.Preetha.
Supervision: L.Dhanapriya , Dr.S.Preetha.
Approval of the final text: L.Dhanapriya , Dr.S.Preetha.

VII. REFERENCES

[1] H. A. Edelstein, Introduction to data mining and knowledge discovery (3rd ed), Potomac, MD: Two Crows Corp. 1999.

[2] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection", J. Mach. Learn. Res., vol. 3, pp. 1157-1182, Mar. 2003.

[3] P. Dhal and C. Azad, "A comprehensive survey on feature selection in the various fields of machine learning", Appl. Intell., vol. 52, pp. 4543-4581, 2022.

[4] M. Wang, H. Han, Z. Huang and J. Xie, "Unsupervised spectral feature selection algorithms for high dimensional data", Front. Comput. Sci., vol. 17, no. 5, 2023.

[5] M. Nazir, A. Majid-Mirza, S. Ali-Khan, PSO-GA Based Optimized Feature Selection Using Facial and Clothing Information for Gender Classification, Journal of Applied Research and Technology, Volume 12, Issue 1, 2014.

[6] Aparna U.R. and S. Paul, "Feature selection and extraction in data mining," 2016 Online International Conference on Green Engineering and Technologies (IC-GET), Coimbatore, India, 2016.

[7] ALASALI, Tasnim & Ortakci, Yasin. "Clustering Techniques in Data Mining: A Survey of Methods, Challenges, and Applications", Computer Science. 9. 2024.

[8] Jahirabadkar, Sunita & Kulkarni, Parag, "Clustering for High Dimensional Data: Density based Subspace Clustering Algorithms", International Journal of Computer Applications. 63. 29-35, 2013.

[9] J. Prinzbach, T. Lauer and N. Kiefer, "Accelerating Density-Based Subspace Clustering in High-Dimensional Data," 2021 International Conference on Data Mining Workshops (ICDMW), Auckland, New Zealand, 2021.

[10] V. K. Sharma and A. Bala, "Clustering for high dimensional data," 2014 First International Conference on Networks & Soft Computing (ICNSC2014), Guntur, India, 2014.

[11] Anand Rajaraman, Jure Leskovec, and Jeffrey D. Ullman, "Clustering, " in Mining of Massive Datasets, 4th ed., 2013, ch. 7, pp 239-278.

[12] F. Li, M. Zhou, S. Li and T. Yang, "A New Density Peak Clustering Algorithm Based on Cluster Fusion Strategy," in IEEE Access, vol. 10, pp. 98034-98047, 2022.

[13] Z. Yu, D. Wang, X. -B. Meng and C. L. P. Chen, "Clustering Ensemble Based on Hybrid Multiview Clustering," in IEEE Transactions on Cybernetics, vol. 52, no. 7, pp. 6518-6530, July 2022

[14] F. Li, J. Wang, Y. Qian, G. Liu and K. Wang, "Fuzzy Ensemble Clustering Based on Self-Coassociation and Prototype Propagation," in IEEE Transactions on Fuzzy Systems, vol. 31, no. 10, pp. 3610-3623, Oct. 2023.

[15] Xiaobo Li, Qiyong Fu, Qi Li, Weiping Ding, Feilong Lin, Zhonglong Zheng, "Multi-objective binary grey wolf optimization for feature selection based on guided mutation strategy", Applied Soft Computing, Volume 145, 2023.

[16] L. Sun, "Multi-Path Minimum Spanning Tree and Superpixel Based Cost Aggregation for Stereo Matching," in IEEE Access, vol. 11, pp. 121096-121108, 2023.

[17] L. Dhanapriya and S. Preetha, "HFMST-PSO: An Efficient Hybrid Fuzzy Based Minimum Spanning Tree (HFMST) With Particle Swarm Optimization Clustering Algorithm", Metszet journal, 2023.

[18] L. Mei, L. Zhang, Y. Zeng, T. Yan, P. Jiang and S. Li, "AFSC: An Improved Spectral Clustering Based on Adaptive Neighbor and Fuzzy Affiliation," in IEEE Access, vol. 12, pp. 133426-133440, 2024.

[19] L. Dhanapriya and S. Preetha, "HFMST: An Efficient Adaptive Fuzzy Linkage Feature Selection with Hybrid Fuzzy Based Minimum Spanning Tree (HFMST) Clustering Algorithm", Journal of Harbin Engineering University, Vol 44, No. 7, July 2023.

[20] F. Pasi et al., "Clustering datasets," 2015. [Online]. Available: http://cs.uef.fi/sipu/datasets/.

[21] Markelle Kelly, Rachel Longjohn, Kolby Nottingham, The UCI Machine Learning Repository, https://archive.ics.uci.edu

[22] Gaurav Mishra and Sraban Kumar Mohanty, "RDMN: A Relative Density Measure Based on MST Neighborhood for Clustering Multi-Scale Datasets", IEEE Transactions On Knowledge And Data Engineering, VOL. 34, NO. 1, January 2022