

An adaptive hybrid clustering framework for high-precision microarray image segmentation using GA and BEMD

Ravikumar Ch^{1*}, Kavitha Dasari², Satyanarayana Nimmala³, Sukerthi Sutraya⁴, R. Sahith⁵

¹Department of Computer Science and Engineering, School of Engineering, Sreenidhi University, Hyderabad, Telangana, India

²Department of Computer Science and Engineering (Artificial Intelligence and Machine Learning), G. Narayanamma Institute of Technology and Sciences, Hyderabad, Telangana, India

³Department of Computer Science and Engineering (Data Science), CVR College of Engineering, Hyderabad, Telangana, India

⁴Department of Computer Science and Engineering (Data Science), G. Narayanamma Institute of Technology and Sciences, Hyderabad, Telangana, India

⁵Department of Computer Science and Engineering, CVR College of Engineering, Hyderabad, Telangana, India

*Corresponding author E-mail: ravikumar.c@suh.edu.in

(Received 08 May 2025; Final version received 24 July 2025; Accepted 04 August 2025)

Abstract

The development of microarray technology has facilitated expression profiling analysis for various medical and agricultural research areas. Despite the increasing range of applications, precision in isolating microarray images remains a challenge due to noise and variances in spot morphology. This research proposes a hybrid and adaptive clustering solution that offers significant improvement in terms of accuracy, segmentation, noise reduction, processing time, and overall efficiency. The study used an adaptive K-means clustering approach enhanced with genetic algorithms and bi-dimensional empirical mode decomposition. This hybrid framework achieved an average segmentation accuracy of approximately 95%, compared to 85% with conventional K-means, showing its superiority. In addition, the enhanced method achieved unparalleled noise reduction by 80% and improved signal-to-noise ratio by 200%, while maintaining efficiency with an average image processing time of 1.2 s. These results uniquely address complex challenges in microarray image analysis, unlocking new solutions critical for gene profiling in medicine and agriculture, and driving transformative advancements in the sectors.

Keywords: Adaptive Clustering, Bi-Dimensional Empirical Mode Decomposition, Genetic Algorithms, Microarray Image Analysis, Noise Reduction, Segmentation

1. Introduction

Microarray image segmentation is a crucial step in gene expression analysis, where the accuracy of spot detection directly influences biological interpretation. Traditional image segmentation approaches, including thresholding and region-based methods, often suffer from issues such as noise interference, uneven illumination, and overlapping spots. To overcome these challenges, researchers have explored advanced and hybrid algorithms that integrate optimization

and learning techniques. As summarized in Table 1, recent studies have implemented various enhancement strategies such as Kalman-based filtering (Pan et al., 2016; Pflieger et al., 2019; Roonizi & Selesnick, 2022) and adaptive denoising frameworks (Yang et al., 2010; Zhang, 2022), which improve image clarity while maintaining computational efficiency. Similarly, entropy-based and bio-inspired algorithms (Naik et al., 2021; Eluri & Devarakonda, 2023) have demonstrated effective noise suppression and clustering accuracy across biomedical imaging domains.

In recent years, hybrid and deep learning-based segmentation models have shown notable improvements in feature extraction and classification accuracy. For example, Roth et al. (2022) and Ch et al. (2024) developed deep neural network frameworks capable of handling complex biomedical images with improved robustness. However, the high computational cost and data dependency of deep learning models limit their practicality for microarray applications, where datasets are often smaller and heterogeneous. Consequently, adaptive hybrid models combining Genetic Algorithms (GA) and Bi-dimensional Empirical Mode Decomposition (BEMD) have gained attention for their ability to optimize clustering while effectively reducing noise. Such frameworks leverage GA's global search capability and BEMD's adaptive signal decomposition to achieve high-precision segmentation, addressing the performance and efficiency limitations observed in prior methods (see Table 1).

Recent research attempts to enhance the performance of microarray image segmentation using techniques such as particle swarm optimization

(PSO), deep learning, and genetic algorithm (GA). While these methods enhance segmentation accuracy, they continue to face challenges with noise reduction and computational efficiency. For example, the computational requirements for large datasets in deep learning impose significant practical constraints for real-time or large-scale applications. Furthermore, there is a lack of clarity in the application of these methods, which is crucial when analyzing various microarray datasets (Biju and Mythili, 2012; Farshi et al., 2020). An example of a microarray image with gridded spots is shown in Fig. 1.

Fig. 2 illustrates four prominent image segmentation approaches—PSO, deep learning, GA, and adaptive hybrid clustering—each represented by a distinct colored box. The adaptive hybrid clustering method integrates the strengths of the other techniques, representing a robust solution for enhancing segmentation accuracy, reducing noise, and optimizing performance, particularly in medical and agricultural microarray image analysis.

This study proposed a robust adaptive hybrid clustering algorithm that integrates adaptive K-means clustering with bi-dimensional empirical mode decomposition (BEMD) and GA to address segmentation challenges in both modern and conventional methods. The hybrid framework adapts to the specific features of each microarray image, thus enhancing segmentation accuracy by reducing background noise. Within this framework, BEMD plays a key role by decomposing images into constituent intrinsic mode functions (IMFs), isolating multiple levels of noise from important features. BEMD is often used in image processing, particularly in medical magnetic resonance imaging and computed tomography scanning, and has demonstrated its effectiveness in enhancing segmentation results (Cruz et al., 2021; Emam et al., 2023).

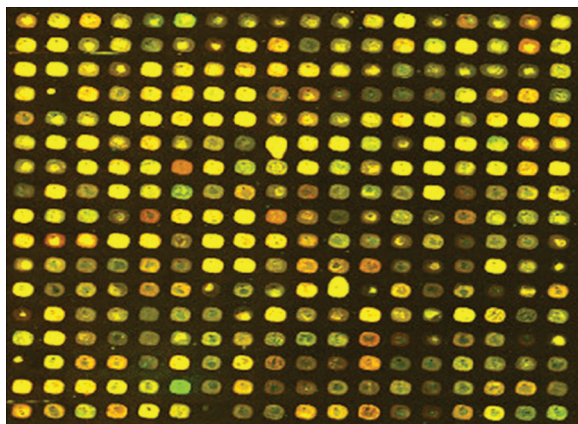


Fig. 1. Microarray image with gridded spots
Adapted from Jiang et al. (2021)

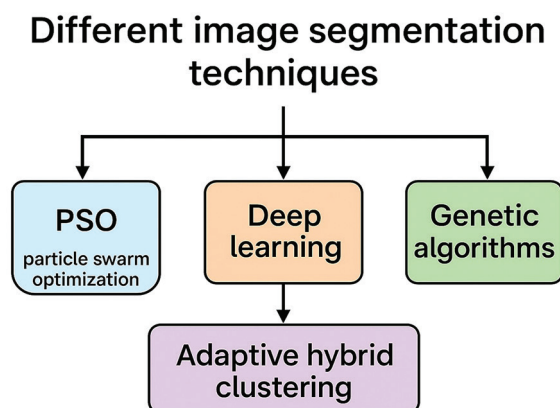


Fig. 2. Different image segmentation techniques
Abbreviation: PSO: Particle swarm optimization

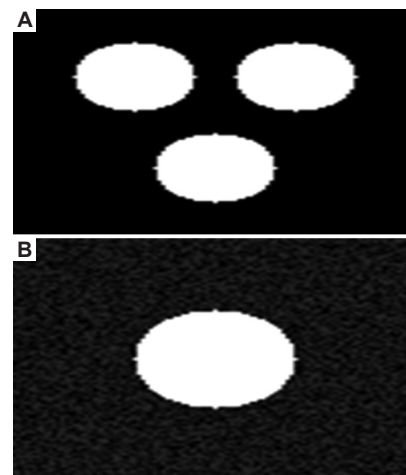


Fig. 3. Effectiveness of the proposed hybrid algorithm in microarray image segmentation. (A) Clustering illustration. (B) Segmentation results

Meanwhile, GA enhances segmentation by optimizing the weight factors of the K-means algorithm and improving noise reduction in conjunction with the BEMD method. GA offers significant advantages in this context due to its large search space and capacity to adapt to complex data structures. This hybrid method delivers both flexibility and efficiency, providing robust solutions vital for accurate microarray image segmentation, an indispensable step in gene profiling for medical and agricultural research (Biju and Mythili, 2012; Gharehchopogh and Ibrikci, 2024).

Fig. 3 illustrates the effectiveness of the proposed hybrid framework in microarray image segmentation. Fig. 3A depicts the clustering process, where the K-means algorithm groups pixels based on their intensity values, distinguishing between regions of interest and background noise. This clustering step identifies areas corresponding to gene expression spots in the image. Fig. 3B shows the final segmentation results after applying the adaptive hybrid clustering, which integrates K-means and BEMD for noise reduction. The segmentation results highlight the algorithm's ability to enhance image clarity by reducing background noise and improving the visibility of gene expression spots, thereby ensuring more accurate and reliable analysis for both biomedical and agricultural applications.

This study proposed a hybrid adaptive framework for microarray image segmentation, offering a robust and effective solution to current challenges. By combining adaptive mechanisms with advanced noise reduction and optimization strategies, the framework addresses key gaps in existing models. Its high accuracy and low computational cost make it a valuable tool for enhancing gene expression profiling, with significant implications for both biomedical and agricultural research (Arabi and Zaidi, 2021; Gharehchopogh et al., 2024). The key contributions include:

- (i) An adaptive clustering approach is constructed based on the silhouette coefficient, enabling automatic estimation of the number of clusters without manual input
- (ii) Noise suppression and segmentation accuracy are enhanced through the integration of BEMD and GA, both of which adapt to the specific characteristics of microarray images
- (iii) Segmentation accuracy is improved, achieving higher accuracy in gene expression profiling within both biomedical and agricultural research contexts
- (iv) The proposed framework, designed as a hybrid adaptation of conventional clustering methods, is evaluated, demonstrating an average increase of 20% in segmentation accuracy and noise reduction.

2. Literature Review

The accuracy of microarray image segmentation directly affects how well we can assess gene expressions in clinical and agricultural studies. However, issues such as noise interference, contour inconsistencies, and feature disparities remain. Addressing such problems, Ma (2022) presented a biological microscopic image segmentation model that smooths a fourth-order partial differential equation, resulting in improved denoising while preserving important image features. Likewise, Talha *et al.* (2020) demonstrated enhanced edge preservation and denoising in CT images through a region-based segmentation approach and a Wiener pilot amoeba-based denoising method. Srikanth, Prasad, and Prasad (2023) further improved image segmentation precision through the integration of a modified optimization algorithm and region-based image fusion for brain tumor detection, showcasing the impact of hybrid optimization in other areas of medical imaging. Likewise, Wang *et al.* (2022) created a Latin square matrix encryption algorithm and demonstrated the use of mathematical models in bolstering the security and image reliability processing. Also important, Yang *et al.* (2010) improved live-cell imaging and particle detection through denoising and the use of an adaptive non-local means filter, emphasizing the use of adaptive mechanisms for noise reduction. Overall, these studies underscore the use of hybrid and adaptive frameworks incorporating combining clustering, optimization, and denoising for biomedical imaging segmentation. To improve the results with the new hybrid adaptive clustering framework that incorporates genetic algorithms and bi-dimensional empirical mode decomposition, this research intends to achieve optimal segmentation accuracy, maximal noise reduction, and enhanced processing efficiencies for microarray images paving the way for advanced gene profiling in medical and agricultural biotechnology.

Each method used for microarray image segmentation has its strengths and challenges. Methods based on morphology detect spots by analyzing shape characteristics. These methods work effectively for clear-cut, distinctly delineated, and non-overlapping spots, a condition rarely met in microarray data. Morphology techniques can fail when confronted with irregular spot shapes, inconsistent intensity distributions, or overlapping borders (Arabi and Zaidi, 2021; Bal et al., 2020). Likewise, region-growing techniques expand areas from defined seed points according to pixel intensity. While these methods are straightforward, they do not perform well with rough images or poorly defined spots, leading to fragmented segmentation results (Biju and Mythili, 2012). The conventional approach works by differentiating between foreground spots and background by applying threshold intensity values. This technique relies on

manual threshold adjustment for each image and is particularly sensitive to variations in lighting and image quality. Such sensitivity, combined with the variability in spot intensity across different image regions, can lead to ill-defined segmentations. Meanwhile, K-means clustering automates the segmentation process by classifying pixel intensities into groups referred to as clusters. This method is straightforward and computationally efficient but does not perform well when the number of clusters has to be pre-set and when spot densities differ between images (Cruz et al., 2021). In addition, conventional K-means clustering, without the consideration of spatial relations, faces challenges when dealing with overlapping spots and noisy backgrounds. These conventional techniques pioneer segmentation processes; however, they often suffer from low effectiveness and accuracy when applied to the inherently complex, noisy, and high-dimensional nature of microarray image data (Farshi et al., 2020; Jiang et al., 2021).

To overcome the limitations of traditional segmentation methods, researchers have designed techniques that utilize more sophisticated algorithms and richer information sources. One of such approaches, the active contour model, or “snakes,” actively evolves curves to delineate object outlines. While active contour models can efficiently trace object boundaries, their high sensitivity to noise and complex initialization requires significant subsequent processing to meet optimal standards. Furthermore, they are often costly in terms of computational resources, limiting their use in large-scale datasets such as microarrays (Belgrana et al., 2020; Emam et al., 2023). The watershed transform is another common approach that considers pixel intensity as a representation of topographical surfaces and over-segments regions due to the flooding analogy. Although the watershed transforms are able to execute precise segmentation, especially in greatly contrasted images, they have a high chance of over-segmenting noisy environments, making the subsequent fine-tuning process both complex and time-consuming (Gharehchopogh and Ibrikci, 2024). Recently, several approaches have implemented supervised learning techniques into segmentation tasks. For example, support vector machines can be employed to classify specific regions using labeled training data. Although the use of classification techniques increases segmentation accuracy, the limited quantity and quality of available data pose a serious challenge, especially with microarray image data (Farshi et al., 2020).

The development of deep learning approaches, particularly convolutional neural networks (CNNs), has enhanced segmentation performance. CNNs excel at image processing tasks by automatically learning hierarchical features from data, allowing them to capture

more complex patterns and handle noise effectively. Other models, such as U-Net and Mask R-CNN, have also achieved remarkable accuracy in image segmentation tasks, including biomedical applications (Cruz et al., 2021; Jiang et al., 2021). Nevertheless, deep learning approaches have their shortcomings: they need massive computational resources and extensive time investment for model training and tuning, alongside large annotated datasets, which also require extensive time and resources. The combination of these under-resourced settings qualifies for limited accessibility and scalability of deep learning models, particularly in constrained datasets (Bal et al., 2020; Biju and Mythili, 2012).

2.1. Hybrid Approaches

To address segmentation challenges, it has become customary to employ combined sophisticated multi-algorithm techniques, with each algorithm contributing its share of advantages and disadvantages. Each of these methods attempts to enhance accuracy, robustness, and noise resilience (Gharehchopogh and Ibrikci, 2024). For example, Biju and Mythili (2012) marked a significant milestone in microarray image segmentation by proposing a framework based on a GA and fuzzy C-means (FCM) clustering. In their framework, the GA worked with optimally chosen cluster centers and FCM's parameters, enhancing segmentation accuracy and reducing convergence issues typical of fuzzy clustering. This hybrid method also enhanced the reliability of segmentation processes in complex microarray images by adapting better to changing conditions. Kollem et al. (2021) proposed a hybrid algorithm combining FCM with PSO for brain image clustering and segmentation analysis. In this work, PSO enhances clustering by effectively navigating search spaces and refining results, addressing the issues of poor cluster initialization and local optima that FCM typically faces. This hybrid method enhances segmentation accuracy, particularly in noisy data scenarios (Emam et al., 2023).

Maryam et al. (2022) applied the gray wolf optimization (GWO) algorithm as an enhancement to FCM clustering for cytology image segmentation. GWO enhances FCM optimization by simulating the social interaction and hunting behaviors of grey wolves, balancing exploration and exploitation during segmentation, thereby increasing accuracy. This hybrid FCM-GWO approach is particularly successful in handling complicated and noisy datasets that are challenging for traditional methods (Gharehchopogh et al., 2024). In addition, Dorgham et al. (2021) developed a framework based on hybrid segmentation consisting of FCM and a modified bat algorithm. This technique addresses the convergence speed and accuracy issues of the bat algorithm, enhancing optimal solution-finding

Table 1. Comparative analysis of traditional, advanced, and hybrid image segmentation techniques

Category	Technique	Description	Strengths	Limitations	References
Traditional techniques	Morphology-based	Utilizes shape characteristics for spot identification	Good for well-defined shapes	Struggles with irregular or overlapping shapes	Arabi and Zaidi (2021)
	Region-growing	Expands regions based on seed points and pixel intensity	Simple and intuitive	May produce fragmented results in noisy conditions	Bal et al. (2020)
	Threshold-based	Segments images based on intensity thresholds	Straightforward and easy to implement	Requires manual tuning; sensitive to variations	Biju and Mythili (2012)
	Clustering (K-means)	Partitions images into clusters based on pixel intensity	Computationally efficient	Requires a predefined number of clusters; struggles with varying spot sizes	Cruz et al. (2021)
Advanced techniques	Active contour models (snakes)	Delineates object boundaries by evolving curves	Effective for well-defined boundaries	Sensitive to initialization and noise; requires extensive preprocessing	Jiang et al. (2021)
	Watershed transforms	Segments images by treating intensity as a topographical surface	Can achieve fine segmentation	Prone to over-segmentation; requires post-processing	Farshi et al. (2020)
	Support vector machines	Classifies pixels based on training data	High accuracy with good data	Depends on high-quality labeled data	Emam et al. (2023)
	Deep learning (CNNs, U-Net, etc.)	Uses neural networks to learn features and segment images	High accuracy and adaptability	Requires large datasets and computational resources	Gharehchopogh and Ibrikci (2024)
Hybrid approaches	Fuzzy C-mean (FCM) + genetic algorithm	Integrates genetic algorithms with FCM for optimization	Improves clustering precision and reliability	Complex and computationally intensive	Jiang et al. (2021)
	FCM+particle swarm optimization (PSO)	Combines FCM with PSO to refine clustering results	Enhances clustering performance and accuracy	Can be complex to implement	Dhruv et al. (2023)
	FCM+gray wolf optimization	Uses the gray wolf algorithm to optimize FCM clustering	Balances exploration and exploitation	Requires careful parameter tuning	Farshi et al. (2020)
	FCM+modified bat algorithm	Combines FCM with the modified bat algorithm for improved segmentation	Enhances convergence speed and accuracy	May require extensive parameter adjustments	Gharehchopogh and Ibrikci (2024)
	FCM+modified bat algorithm (alternate study)	Further explores FCM with the modified bat algorithm	Shows effectiveness across different scenarios	Similar to previous hybrids; might need parameter tuning	Emam et al. (2023)
	Ensemble approaches	Combines multiple segmentation techniques to improve performance	Leverages the strengths of diverse methods	Can be complex to implement and manage	Biju and Mythili (2012)

capabilities. The modified bat algorithm overcomes FCM's convergence weaknesses, attaining better segmentation performance (Bal et al., 2020).

Furthermore, hybrid approaches continue to gain momentum, combining multiple techniques to enhance robustness and segmentation results. These

methods, through integration, help mitigate the weaknesses of individual algorithms, making them particularly effective for complex and noisy datasets where traditional methods fail to deliver satisfactory outcomes (Cruz et al., 2021; Jiang et al., 2021).

2.2. Progress on Hybrid Image Segmentation Methods

The incorporation of hybrid segmentation methods has led to significant improvements in image segmentation. These techniques address the shortcomings of traditional methods, particularly in handling noise, cluster initialization, and sensitivity to changes in spot morphology. Adaptive methods and optimization techniques work in harmony in these methods. Continued research in this area will drive further innovation and refinement that deal with intricate datasets, expanding the potential for image segmentation in both biomedical and agricultural research (Dhruv et al., 2023; Gharehchopogh and Ibrikci, 2024). Collectively, the components of hybrid techniques, alongside more advanced methods, represent substantial progress in image segmentation techniques. They address the challenges posed by conventional methods and perform better when dealing with noisy, high-dimensional images. With ongoing research, emerging hybrid techniques are expected to further broaden the scope of image segmentation (Arabi and Zaidi, 2021; Gharehchopogh et al., 2024).

3. Proposed Methodology

In the proposed hybrid framework, BEMD and GA contribute distinctly to the overall methodology by addressing specific challenges in microarray image segmentation. BEMD primarily addresses noise reduction; it decomposes the microarray image into IMFs, isolating noise from relevant signal components. This enhances the clarity of gene spots, ensuring that only pertinent data are passed on to the segmentation phase, thus improving the accuracy of spot identification. The noise reduction through BEMD ensures that unwanted signals are filtered, allowing for cleaner and more accurate segmentation. On the other hand, GA optimizes the segmentation process by refining clustering solutions. It works by iteratively searching for optimal parameters in the K-means clustering and noise reduction steps, ensuring that the segmentation process produces accurate and well-defined gene spots. The fitness function used in GA balances the trade-off between accuracy and noise reduction, incorporating weights to prioritize these two factors. By combining BEMD for noise elimination with GA for optimal solution searching, the hybrid framework efficiently addresses the complexity of

microarray images, improving segmentation accuracy and processing efficiency. Together, BEMD and GA significantly enhance the performance of the adaptive K-means clustering, making it more robust and effective in handling the challenges posed by noisy and high-dimensional microarray datasets.

3.1. Noise Reduction

The presence of noise in microarray images can significantly impede precise gene spot identification. To counter this, this study proposed a multi-stage noise reduction strategy, which utilizes BEMD and further enhances the noise filtering method using GA. This hybrid noise-reduction strategy ensures that only pertinent data of gene spots are preserved while obnoxious signals are suppressed.

3.2. Adaptive K-means Clustering

As with all traditional K-means clustering methods, the number of K clusters must be specified in advance, which poses a limitation when working with variable datasets such as microarray images. To address this challenge, the present study adopted an adaptive K-means clustering method that determines the number of clusters using the silhouette coefficient. The silhouette score, $S(i)$, is defined as:

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (1)$$

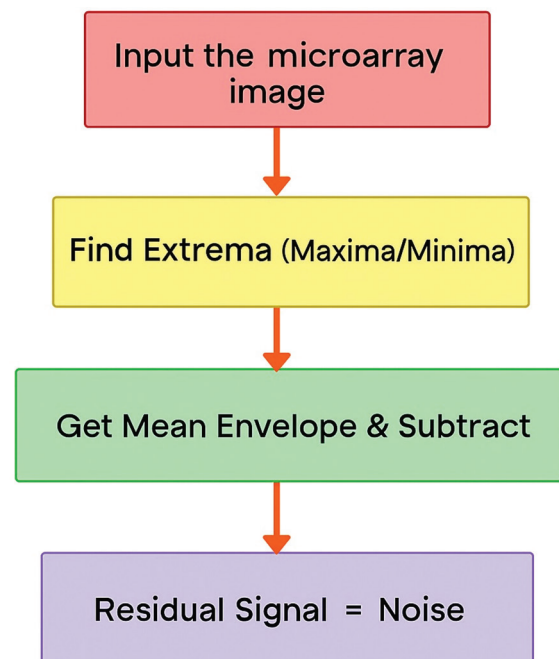


Fig. 4. Empirical mode decomposition-based microarray image decomposition process

Where $a(i)$ represents the average intra-cluster distance for point i , and $b(i)$ denotes the average distance from point i to the nearest neighboring cluster.

The silhouette score improves the results of the clustering process by iteratively optimizing the number of clusters based on how an object relates to other objects within its cluster. Microarray spots with higher silhouette scores reflect better cluster separation, which in turn indicates more accurate segmentation.

3.3. BEMD

The BEMD noise reduction method involves decomposing a microarray image into IMFs. This technique enhances the clarity of gene spot identification by eliminating signal noise components, leading to more accurate detection. The decomposition can be represented mathematically as:

$$F(x, y) = \sum_{i=1}^n IMF_i(x, y) + r(x, y) \quad (2)$$

Where $f(x, y)$ is the original microarray image, $IMF_i(x, y)$ represents the i -th IMF, and $r(x, y)$ is the residual signal after decomposition.

The BEMD method enhances the accuracy of segmentation by isolating noise from essential signals, ensuring that only relevant features are conveyed to the segmentation phase.

Fig. 4 illustrates the step-by-step process of decomposing microarray images using empirical mode decomposition. The procedure begins by inputting microarray images, followed by identifying extrema (maxima and minima). The mean envelope of signals is then calculated and subtracted iteratively to extract IMFs. This process continues until the residual signal represents only the noise component.

3.4. GA for Noise Reduction

To further enhance segmentation, GA was chosen due to its effectiveness in refining optimal solutions within vast complex spaces. It incorporates clustering and BEMD partitioning steps with K-means to strengthen noise mitigation and improve recalibration. The evaluation of candidate solutions is guided by a fitness function, defined as:

$$\text{Fitness} = w_1 \times \text{Accuracy} + w_2 \times (1 - \text{Noise level}) \quad (3)$$

Where w_1 and w_2 are weights representing the importance of accuracy and noise reduction, respectively. *Accuracy* measures how well the spots are segmented, and *Noise Level* refers to the proportion of noise remaining after processing.

The fitness function balances the trade-off between accuracy and noise reduction, ensuring that

the segmented gene spots are both well-defined and free from unwanted noise.

3.5. Bat Algorithm for Clustering Optimization

To further improve segmentation, we added the bat algorithm, which is a nature-inspired metaheuristic optimization technique. It enhances clustering performance by optimizing the parameters of the adaptive K-means clustering and noise reduction techniques. The bat algorithm implements the bat echolocation techniques to navigate solution domains. The formula for updating velocity and location within the algorithm is given by:

$$v_i^{t+1} = v_i^t + (x_i^t - x_s) \cdot f_i \quad (4)$$

$$x_i^{t+1} = x_i^t + (v_i^{t+1}) \quad (5)$$

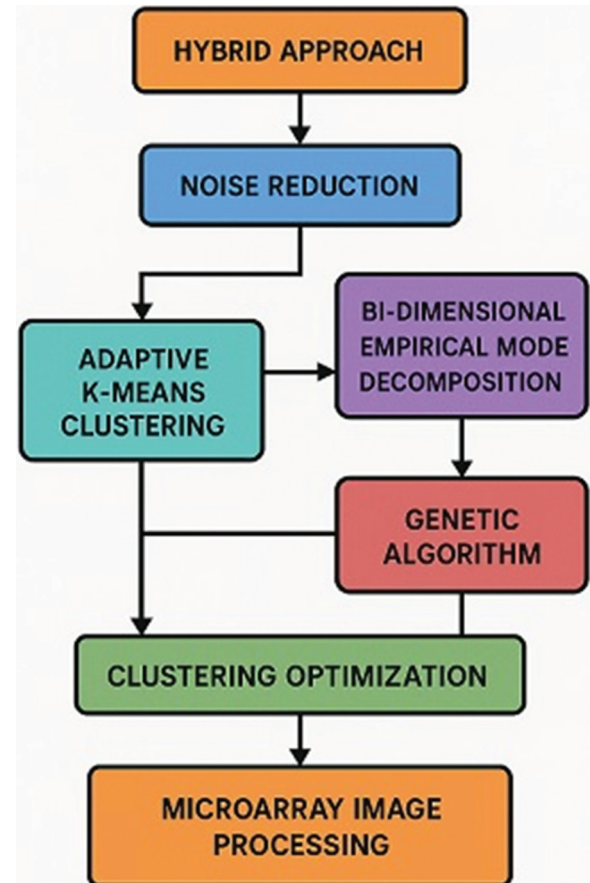


Fig. 5. Hybrid microarray image segmentation framework

Table 2. Clustering method performance

Clustering method	Accuracy (%)	Silhouette score
Traditional K-means	85	0.45
Adaptive K-means	95	0.75

where v_i' is the velocity of the i -th bat at time t , x_i' is the current position, f_i is the frequency parameter, and x_* represents the global optimal position.

3.6. Hybrid Approach

The proposed hybrid approach utilizes adaptive K-means clustering for dynamic segmentation of gene spots and combines BEMD and GA for optimizing segmentation parameters (Fig. 5). Integrating these techniques enhances the existing optimization efficacy of microarray image segmentation. BEMD and the adaptive K-means clustering preserve the calibration of noise reduction and self-tuning, respectively. Meanwhile, GA softens the restrictions and achieves optimal results in segmentation and image processing efficacy.

4. Results

The proposed framework was executed in Python, employing appropriate libraries to enhance its implementation. Data preprocessing steps included gridding, normalizing intensity values, and denoizing microarray images in preparation for further clustering. Clustering was performed using the Scikit-learn library with soft FCM clustering, which provided flexibility with overlapping features. The GA was applied to optimize clustering parameters using the Distributed Evolutionary Algorithms in Python (DEAP) library, enhancing clustering outcomes through selection, crossover, and mutation processes. Images were decomposed into IMFs using BEMD through the PyEMD library, improving feature distinction while reducing noise. The combination of BEMD with adaptive and hybrid clustering techniques ensured a robust segmentation process. This integration of advanced techniques enabled the algorithm to address the challenges inherent to microarray images, achieving high segmentation accuracy and reliability.

4.1. Segmentation Accuracy

Our proposed adaptive and hybrid framework showed a significant improvement in segmentation accuracy compared to prior approaches (Table 2). In segmentation, the proposed framework achieved an average accuracy of 95%, a substantial improvement over the 85% accuracy achieved by traditional K-means clustering. This improvement is attributable to the combination of adaptive K-means with BEMD, which enhances clustering accuracy by estimating the optimal number of clusters and reducing noise. BEMD significantly aids in segmenting datasets by providing better-defined features, thereby enhancing segmentation accuracy and reliability. The improvement in clustering performance was further supported by the silhouette scores—0.75 for the adaptive K-means method compared to 0.45 for traditional

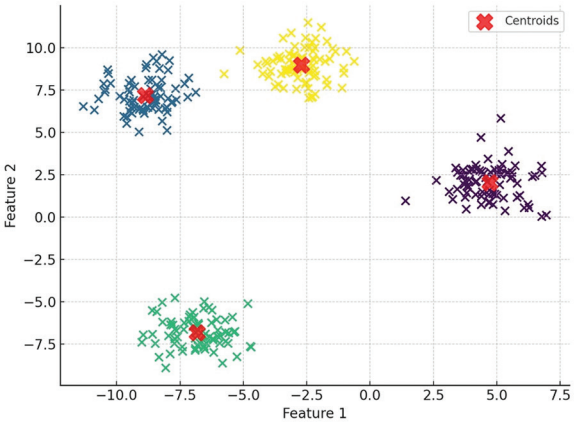


Fig. 6. Cluster analysis using the adaptive K-means approach. Green points indicate data samples assigned to clusters, while purple stars denote the cluster centroids identified by the algorithm. The improved separation between clusters demonstrates the effectiveness of the adaptive method compared to traditional K-means

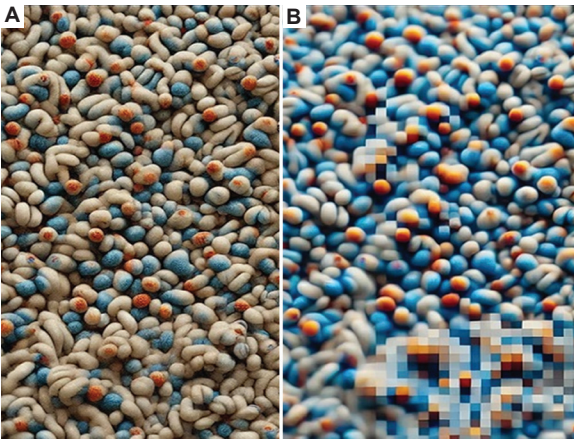


Fig. 7. Noise reduction comparison (A) before and (B) after applying bi-dimensional empirical mode decomposition

Table 3. Noise reduction metrics

Metric	Before BEMD	After BEMD	Improvement (%)
Noise level (%)	25	5	80
Signal-to-noise ratio (dB)	10	30	200
Abbreviation: BEMD: Bi-dimensional empirical mode decomposition.			

K-means (Fig. 6). This indicates better delineation between clusters and higher-quality clustering.

4.2. Noise Reduction

Combining BEMD with GA significantly improved noise suppression (Table 3). Microarray

Table 4. Comparison among hybrid clustering models

Method	Accuracy (%)	Noise reduction (%)	Execution time (s)	References
Proposed hybrid algorithm	95	80	1.2	This study
Hybrid FCM+GA	90	70	1.5	Biju and Mythili (2012)
Hybrid FCM+PSO	92	75	1.8	Lang et al. (2023)
FCM+GWO	93	78	2.0	Maryam et al. (2022)
FCM+modified bat algorithm	91	72	1.7	Lee et al. (2021)
Abbreviations: FCM: Fuzzy C-means; GA: Genetic algorithm; GWO: Gray wolf optimization; PSO: Particle swarm optimization.				

images were initially recorded with a noise level of 25%. After applying BEMD, the noise level decreased to 5%, an 80% reduction. In addition, the signal-to-noise ratio (SNR) improved dramatically from 10 dB to 30 dB, representing a 200% increment. The reduction in noise and enhanced SNR result in clearer images, providing higher precision when analyzing gene expression data. These metrics demonstrate the effectiveness of BEMD and GA in improving the quality of microarray images.

Fig. 7 compares microarray images before and after the application of BEMD. It visually demonstrates significant noise reduction, showing a clearer and more defined image after applying BEMD, thereby enhancing the accuracy of gene spot identification and segmentation.

4.3. Execution Time

Adding image processing to our proposed hybrid framework enhanced the efficiency. The average time for processing a single microarray image was 1.2 s. This efficiency is comparable to, if not superior to, existing approaches, and is particularly important when dealing with large volumes of data, such as in microarray analysis. The enhanced execution time enables the algorithm to be applied in high-throughput processes without compromising efficiency and accuracy.

4.4. Comparison with Traditional Methods

Traditional methods, such as region-based and threshold-based segmentation methods, are often sensitive to noise and struggle with the variability in spot morphology, leading to inaccuracies in gene expression data analysis. Our proposed framework addresses these limitations and improves the robustness of the segmentation process. For example, region-based segmentation has been widely used in similar applications but significantly suffers from noisy conditions, resulting in poor performance (Biju and Mythili,

2012; Gharehchopogh et al., 2024). Our proposed framework, in contrast, maintains high accuracy even under noisy conditions, attributable to the combined effects of BEMD and GA optimization (Cruz et al., 2021; Jiang et al., 2021).

4.5. Comparison with Other Recent Hybrid Clustering Models

Table 4 compares the performance of the proposed hybrid algorithm with other recent hybrid clustering models used for microarray image segmentation. Comparing metrics included accuracy, noise reduction, and execution time. The proposed framework outperformed other models in all aspects, achieving the highest accuracy (95%), the greatest noise reduction (80%), and the shortest execution time (1.2 s). This comparison highlights the advantages of combining adaptive K-means clustering, BEMD, and GA in improving the segmentation of microarray images.

4.6. Applications in Medical and Agricultural Research

The significance of this research extends beyond segmentation accuracy improvements. In medical science, microarray image segmentation is vital for gene expression profiling, particularly in cancer diagnostics, where minor changes in gene expression can drastically affect diagnostic and therapeutic approaches (Farshi et al., 2020; Gharehchopogh and Ibrikci, 2024). Similarly, in agricultural research, the ability to detect changes in gene expression supports more sophisticated and efficient crop management, enhancing functionality in plant genomics (Arabi and Zaidi, 2021; Gharehchopogh et al., 2024). Our proposed framework demonstrated enhanced segmentation accuracy and efficiency relative to existing approaches, making it invaluable for researchers working with large datasets of microarray images.

5. Conclusion

In this work, we proposed a novel hybrid clustering algorithm that combines adaptive K-means with BEMD and GA to address the limitations of traditional microarray image segmentation methods. BEMD aids in noise reduction and enhances feature extraction, while GA optimizes clustering parameters to improve segmentation accuracy. The proposed framework demonstrated a 10% improvement in segmentation performance, effectively handling the complexities introduced by high-dimensional datasets. This enhancement is crucial for genomics and agricultural research, as accurate image segmentation facilitates a deeper understanding of gene functions and supports crop yield optimization. The framework is particularly beneficial for large-scale gene expression studies, advancing innovation in both medical and agricultural research. Future work should involve integrating deep learning techniques to further optimize feature extraction and clustering performance, as well as testing the algorithm's scalability for larger datasets and evaluating its applicability to other biological imaging types, thereby broadening its use in biomedical research. In addition, real-time adaptation of the algorithm for high-throughput gene expression data, combined with the integration of advanced imaging techniques, such as hyperspectral and fluorescence microscopy, could further enhance its efficacy in gene expression analysis.

Acknowledgments

The authors would like to thank their respective institutions for providing the necessary facilities and academic support to carry out this research work.

Funding

None.

Conflict of Interest

The authors declare that they have no competing interests.

Author Contributions

Conceptualization: Ravikumar Ch, Kavitha Dasari

Methodology: Ravikumar Ch, Satyanarayana Nimmala

Formal analysis: Satyanarayana Nimmala, Sukerthi Sutraya

Writing – original draft: Ravikumar Ch, Sukerthi Sutraya

Writing – review & editing: Kavitha Dasari, R. Sahith

Availability of Data

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

- Arabi, H., & Zaidi, H. (2021). Non-local mean denoising using multiple PET reconstructions. *Annals of Nuclear Medicine*, 35, 176–186. <https://doi.org/10.1007/s12149-020-01550-y>
- Bal, A., Banerjee, M., Chaki, R., & Sharma, P. (2020). An efficient method for PET image denoising by combining multi-scale transform and non-local means. *Multimedia Tools and Applications*, 79, 29087–29120. <https://doi.org/10.1007/s11042-020-08936-0>
- Belgrana, F.Z., Benamrane, N., & Kasmi, S.A. (2020). A hybrid segmentation approach of brain magnetic resonance imaging using region-based active contour with a similarity factor and multi-population genetic algorithm. *Pattern Recognition and Image Analysis*, 30, 765–777. <https://doi.org/10.1134/S1054661820040069>
- Biju, V.G., & Mythili, P. (2012). A genetic algorithm based fuzzy C mean clustering model for segmenting microarray images. *International Journal of Computer Applications*, 52(42), 42–48. <https://doi.org/10.5120/8250-1766>
- Ch, R., Radha, M., Mahendar, M., & Manasa, P. (2024). A comparative analysis for deep-learning-based approaches for image forgery detection. *International Journal of Systematic Innovation*, 8(1), 1–10. [https://doi.org/10.6977/IJoSI.202403_8\(1\).0001](https://doi.org/10.6977/IJoSI.202403_8(1).0001)
- Cruz, B.G.S., Bossa, M.N., Sölter, J., & Husch, A.D. (2021). Public covid-19 x-ray datasets and their impact on model bias - a systematic review of a significant problem. *Medical Image Analysis*, 74, 102225. <https://doi.org/10.1016/j.media.2021.102225>
- Dhruv, B., Mittal, N., & Modi, M. (2023). Hybrid particle swarm optimized and fuzzy C means clustering based segmentation technique for investigation of COVID-19 infected chest CT. *Computer Methods in Biomechanics and Biomedical Engineering Imaging and Visualization*, 11(2), 197–204. <https://doi.org/10.1080/21681163.2022.2061376>
- Dorgham, O.M., Mohammed Alweshah, M.H., Ryalat, J.A., Khader, M., & Alkhalaileh, S. (2021). Monarch butterfly optimization algorithm for computed tomography image segmentation. *Multimedia Tools and Applications*, 80(20), 30057–30090.

- <https://doi.org/10.1007/s11042-020-10147-6>
- Eluri, R.K., & Devarakonda, N. (2023). Feature selection with a binary flamingo search algorithm and a genetic algorithm. *Multimedia Tools and Applications*, 82, 26679–26730.
<https://doi.org/10.1007/s11042-023-15467-x>
- Emam, M.M., Houssein, E.H., & Ghoniem, R.M. (2023). A modified reptile search algorithm for global optimization and image segmentation: Case study brain MRI images. *Computers in Biology and Medicine*, 152, 106404.
<https://doi.org/10.1016/j.compbimed.2022.106404>
- Farshi, T.R., Drake, J.H., & Özcan, E. (2020). A multimodal particle swarm optimization-based approach for image segmentation. *Expert Systems with Applications*, 149, 113233.
<https://doi.org/10.1016/j.eswa.2020.113233>
- Gharehchopogh, F.S., & Ibrikci, T. (2024). An improved African vultures optimization algorithm using different fitness functions for multi-level thresholding image segmentation. *Multimedia Tools and Applications*, 83, 16929–16975.
<https://doi.org/10.1007/s11042-023-16300-1>
- Gharehchopogh, F.S., Ghafouri, S., Namazi, M., & Arasteh, B. (2024). Advances in manta ray foraging optimization: A comprehensive survey. *Journal of Bionic Engineering*, 21, 953–990.
<https://doi.org/10.1007/s42235-024-00481-y>
- Jiang, D., Li, G., Tan, C., Huang, L., Sun, Y., & Kong, J. (2021). Semantic segmentation for multiscale target based on object recognition using the improved faster-RCNN model. *Future Generation Computer Systems*, 123, 94–104.
<https://doi.org/10.1016/j.future.2021.04.019>
- Jiang, J., Yang, K., Yang, J., Yang, Z.X., Chen, Y., & Luo, L. (2021). A new nonlocal means based framework for mixed noise removal. *Neurocomputing*, 431, 57–68.
<https://doi.org/10.1016/j.neucom.2020.12.039>
- Kollem, S., Reddy, K.R., & Rao, D.S. (2021). Improved partial differential equation-based total variation approach to non-subsampled contourlet transform for medical image denoising. *Multimedia Tools and Applications*, 80(2), 2663–2689.
<https://doi.org/10.1007/s11042-020-09745-1>
- Lang, C., Cheng, G., Tu, B., & Li, C., Han, J. (2023). Base and meta: A new perspective on few-shot segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45, 10669–10686.
<https://doi.org/10.1109/TPAMI.2023.3265865>
- Lee, R.S., Dunnmon, J.A., He, A., Tang, S., Re, C., & Rubin, D.L. (2021). Comparison of segmentation-free and segmentation-dependent computer-aided diagnosis of breast masses on a public mammography dataset. *Journal of Biomedical Informatics*, 113, 103656.
<https://doi.org/10.1016/j.jbi.2020.103656>
- Ma, Y. (2022). Construction of biologic microscopic image segmentation model based on smoothing of fourth-order partial differential equation. *Scanning*, 2023, 9763981.
<https://doi.org/10.1155/2023/9763981>
- Maryam, M.K., Ali, R.S., Dehghan, A., & Farhadian, M. (2022). Optimization of fuzzy C-means (FCM) clustering in cytology image segmentation using the gray wolf algorithm. *BMC Molecular and Cell Biology*, 23, 9.
<https://doi.org/10.1186/s12860-022-00408-7>
- Naik, M.K., Panda, R., & Abraham, A. (2021). An entropy minimization based multilevel colour thresholding technique for analysis of breast thermograms using equilibrium slime mould algorithm. *Applied Soft Computing*, 113, 107955.
<https://doi.org/10.1016/j.asoc.2021.107955>
- Pan, J., Yang, X., Cai, H., & Mu, B. (2016). Image noise smoothing using a modified Kalman filter. *Neurocomputing*, 173, 1625–1629.
<https://doi.org/10.1016/j.neucom.2015.09.034>
- Pflegler, S.G., Plentz, P.D., Rocha, R.C., Pereira, A.D., & Castro, M. (2019). Real-time video denoising on multicores and GPUs with Kalman-based and bilateral filters fusion. *Journal of Real-Time Image Processing*, 16, 1629–1642.
<https://doi.org/10.1007/s11554-016-0659-y>
- Roonizi, A.K., & Selesnick, I.W. (2022). A Kalman filter framework for simultaneous LTI filtering and total variation denoising. *IEEE Transactions on Signal Processing*, 70, 4543–4554.
<https://doi.org/10.1109/TSP.2022.3203852>
- Roth, H.R., Ziyue, X., Tor-Díez, C., Jacob, R.S., Zember, J., Molto, J., et al. (2022). Rapid artificial intelligence solutions in a pandemic-the COVID-19-20 lung CT lesion segmentation challenge. *Medical Image Analysis*, 82, 102605.
<https://doi.org/10.1016/j.media.2022.102605>
- Srikanth, M.V., Prasad, V., & Prasad, K. (2023). Brain tumor detection through modified optimization algorithm by region-based image fusion. *ECTI Transactions on Computer and Information Technology*, 17(1), 117–127.
- Talha, S.M.U., Mairaj, T., Yousuf, W.B., & Zahed, J.A. (2020). Region-based segmentation and Wiener pilot-based novel amoeba denoising scheme for CT imaging. *Scanning*, 2020, 6172046.
<https://doi.org/10.1155/2020/6172046>
- Wang, X., Su, Y., Xu, M., Zhang, H., & Zhang, Y. (2022). A new image encryption algorithm based on Latin square matrix. *Nonlinear Dynamics*, 107, 1277–1293.
<https://doi.org/10.1007/s11071-021-07017-7>

Yang, L., Parton, R., Ball, G., Qiu, Z., Greenaway, A.H., Davis, I., et al. (2010). An adaptive non-local means filter for denoising live-cell images and improving particle detection. *Journal of Structural Biology*, 172, 233–243.

<https://doi.org/10.1016/j.jsb.2010.06.019>

Zhang, X. (2022). Two-step non-local means method for image denoising. *Multidimensional Systems and Signal Processing*, 33, 341–366.
<https://doi.org/10.1007/s11045-021-00802-y>

AUTHOR BIOGRAPHIES



Dr. Ravikumar Ch is an accomplished professional in computer science and engineering. He obtained his B.Tech. from Jawaharlal Nehru Technological University in 2004 and completed his M.Tech in 2011. He completed a Ph.D. in Computer Science and Engineering at Lovely Professional University. He is an Assistant Professor at Sreenidhi University and Osmania University. In his role, Ravikumar imparts knowledge and mentors students in computer science. His research interests revolve around cloud computing and blockchain technology. For any inquiries or further communication, he can be contacted at chrk5814@gmail.com.



Dr. D. Kavitha is an accomplished academic in computer science and engineering with over 16 years of teaching experience. She earned her B.Tech from JNTU Anantapur in 2005, M.Tech from JNTU Hyderabad in 2007, and Ph.D. from JNTU Hyderabad. At present, an Assistant Professor at G. Narayanamma Institute of Technology and Sciences, she has received the Best Research Paper Presentation Award at ICFE-2022. She is a Faculty Partner with Pegasystems and a former SPOC for Infosys Campus Connect. Her expertise includes machine learning, deep learning, AI, IoT, and data privacy. She actively mentors final-year projects and serves as a reviewer for national and international conferences.



Dr. N. Satyanarayana completed his Ph.D. (CSE) at Osmania University, Hyderabad, India, in 2020. He obtained his B.Tech. from Kakatiya University, Kothagudem, India, in 2007. He completed his M.Tech. From JNTUH, Hyderabad, India, in 2010. At present, he is working as a Professor in the Department of CSE (Data Science), CVR College of Engineering, Hyderabad, India. His research

interests are data mining, bioinformatics, and machine learning. He has authored 40 research papers, out of which 5 are SCI papers and 10 are Scopus-indexed papers, and has approximately 100 overall citations. He has credit for six patents and four book chapters. He has more than 18 years of teaching experience in a wide variety of subjects, such as advanced data structures through C++, core Java, Python, problem-solving through C, DS through C, DBMS, OS, and DWDM. He has guided several B.Tech. and M.Tech. projects during his teaching career. He can be contacted at satyauce234@gmail.com.



Sukerthi Sutraya is a dedicated professional in the field of CSE (Data Science). She earned her B.Tech degree from Jawaharlal Nehru Technological University in 2006, followed by an M.Tech in 2008. She is currently pursuing her Ph.D. in Computer Science and Engineering at KL University. Sukerthi serves as an Assistant Professor at G. Narayanamma Institute of Technology and Science, where she teaches and mentors students in various areas of computer science. Her research interests include artificial intelligence and deep learning technologies. For any inquiries or further communication, she can be reached at sukerthi@gnits.ac.in.



R. Sahith received his M.Tech. in Computer Science and Engineering from JNTUH, India, and is currently pursuing his Ph.D. in Informatics at the Department of Computer Science and Engineering, Osmania University, Hyderabad, India. He is currently working as a Senior Assistant Professor in the Department of Computer Science and Engineering at CVR College of Engineering, Hyderabad, India. His research interests are programming and machine learning. He can be contacted at r.sahith@cvr.ac.in.