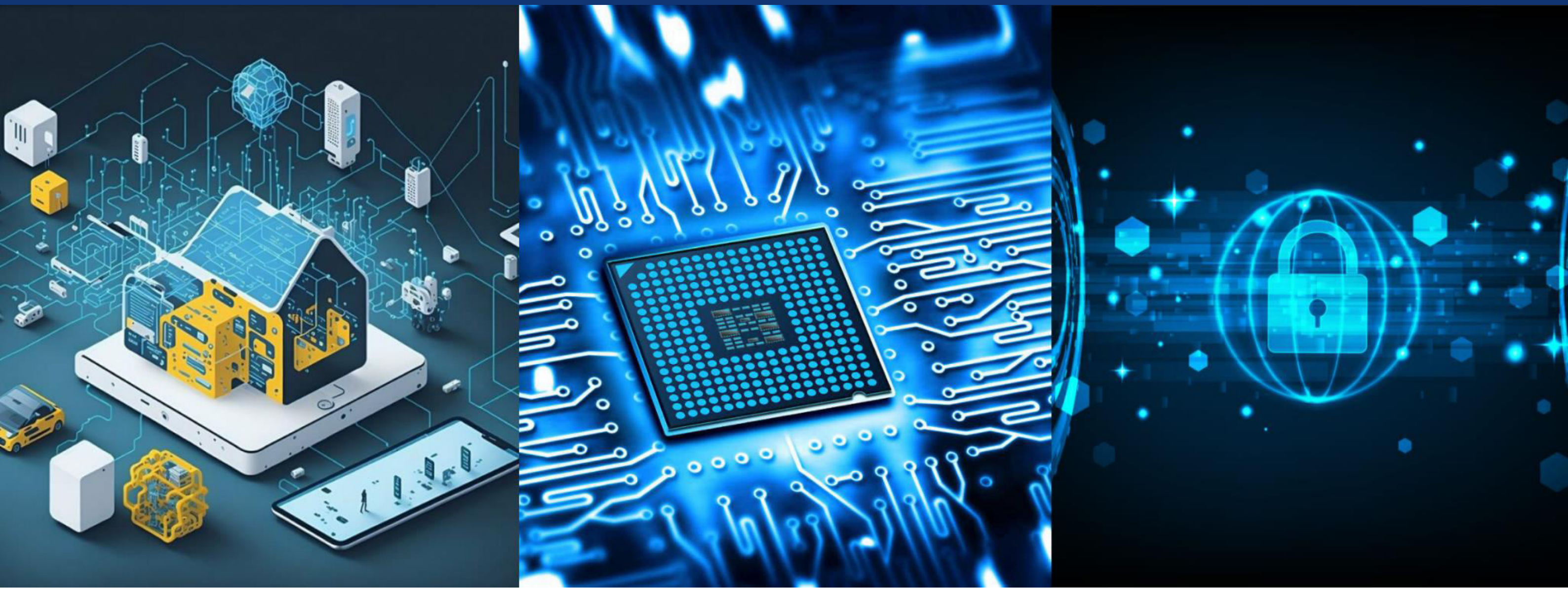
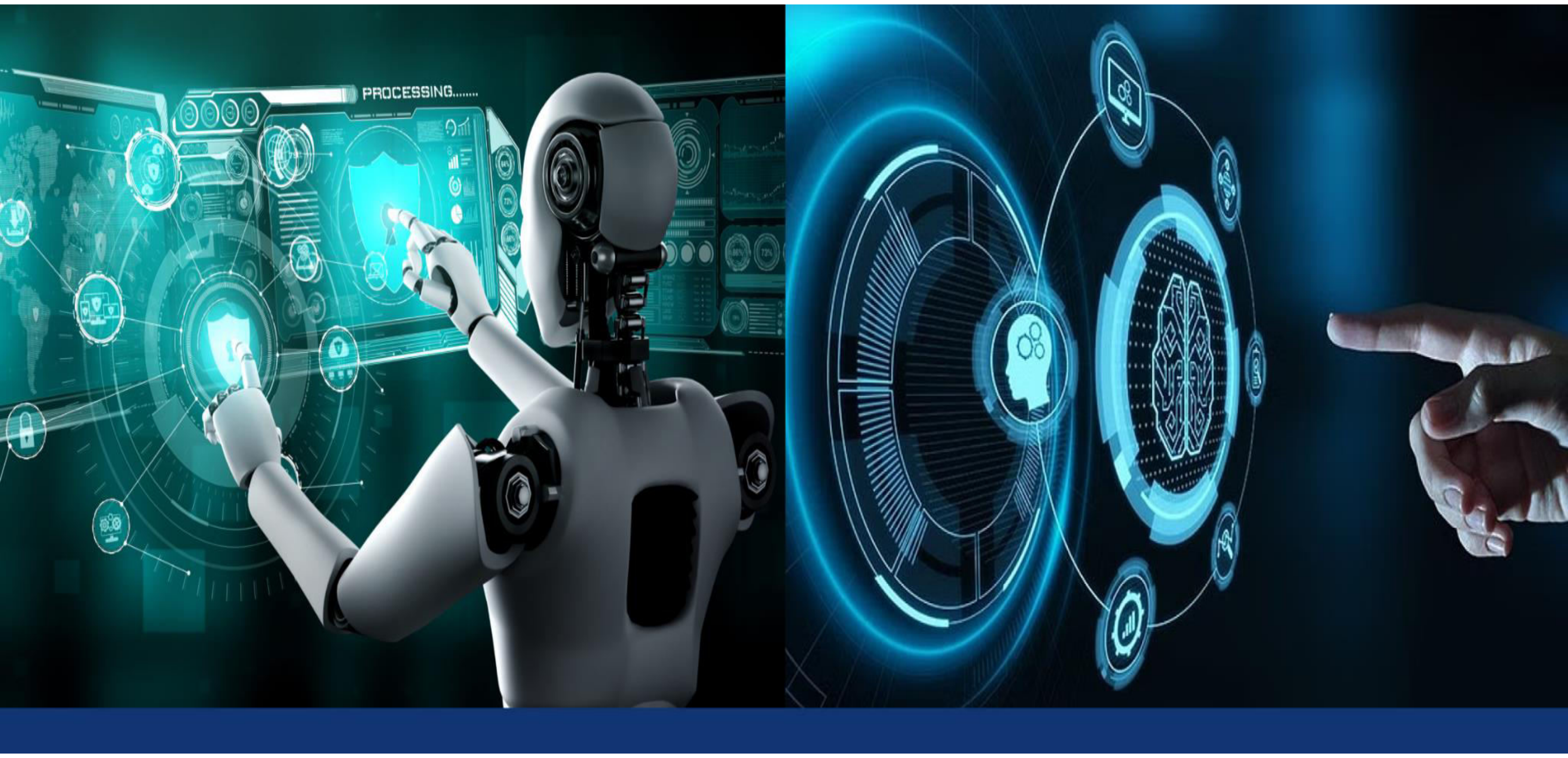
ISSN(O): 2320-9801
ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.771

Volume 14, Issue 4, April 2026

© 2026 IJIRCCE | Volume 14, Issue 4, April 2026 DOI: 10.15680/IJIRCCE.2026.1404087

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Multi-Objective Genetic Algorithm to Enhance the Clustering Problems

Cheekati Sanjay Goud[1], Maryala Harshitha[2], Vedavathi.K[3], Dr.V. Subbaramaiah[4], Dr.K. Rajitha[5]

Student, Department of Computer Science and Engineering, Mahatma Gandhi Institute of Technology, Gandipet, India[1,2]

Assistant Professor, Department of Computer Science and Engineering, Mahatma Gandhi Institute of Technology, Gandipet, India[3,4,5]

ABSTRACT: Clustering is a key technique in data analysis used to group similar data points for better interpretation and decision-making. Traditional methods such as K-means clustering are widely applied due to their simplicity and efficiency; however, they often face limitations when dealing with complex datasets and multi-objective requirements. These methods typically optimize a single objective and are sensitive to initial conditions, which may lead to suboptimal results.To overcome these challenges, this work proposes a multi-objective clustering approach using the Non-dominated Sorting Genetic Algorithm II **(NSGA-II)**. The clustering problem is formulated to simultaneously optimize objectives such as minimizing intra-cluster distance and maximizing inter-cluster separation. NSGA-II effectively explores the search space and generates a diverse set of Pareto-optimal solutions, enabling better trade-offs between conflicting objectives.The performance of the proposed approach is evaluated against the traditional K-Means algorithm. Experimental results demonstrate that the NSGA-II-based method achieves improved clustering quality, greater solution diversity, and enhanced robustness. This highlights the effectiveness of evolutionary multi-objective optimization techniques for solving complex clustering problems in real-world applications.

KEYWORDS: Multi-Objective Optimization, NSGA-II, K-Means Clustering, Evolutionary Algorithms, Pareto Optimality, Clustering Optimization

I. INTRODUCTION

In recent years, the demand for efficient **Home Health Care (HHC)** services has increased due to the growing elderly population and the need for timely medical assistance. One of the key challenges in this domain is organizing and grouping patients or service locations effectively to improve resource utilization and service quality. This problem involves handling multiple conflicting objectives, such as minimizing service time, reducing travel distance, and improving overall efficiency, making it a complex optimization task.

Traditional clustering approaches like K-means clustering are commonly used to group similar data points based on distance measures. Although simple and efficient, K-Means is sensitive to initial centroid selection and often fails to produce optimal results for complex, real-world datasets such as HHC scenarios.

To address these challenges, this project proposes a multi-objective optimization approach using the Non-dominated Sorting Genetic Algorithm II **(NSGA-II)**. The algorithm optimizes multiple objectives simultaneously, such as cluster compactness and separation, while generating a diverse set of Pareto-optimal solutions. Applying NSGA-II to HHC data improves clustering quality, enhances decision-making, and supports efficient healthcare service planning, making it a robust solution for complex optimization problems in home health care systems.

II. LITERATURE SURVEY

Clustering optimization has gained significant research attention due to its importance in data analysis and real-world applications. A. M. Ikotun et al. [1] conducted a benchmarking study on evolutionary clustering methods based on K-means clustering, evaluating their performance using multiple validity indices. While the study provides comprehensive comparison insights, it does not introduce a new algorithm.

B. F. Azevedo et al. [2] proposed a multi-objective clustering approach using Non-dominated Sorting Genetic Algorithm II, followed by K-Means refinement. Their method achieved better trade-offs between cluster compactness and separation but involved higher computational cost.

Manoharan Premkumar et al. [3] introduced a hybrid model combining Grey Wolf Optimizer with weighted K-Means to enhance clustering accuracy and convergence. Although effective, the method requires complex parameter tuning. Ikotun and Rossouw [4] presented a comprehensive review of K-Means variants, offering valuable insights but lacking novel contributions.

Mohamed A. Damos et al. [5] improved K-Means using Genetic Algorithms for better initialization and cluster selection, enhancing accuracy but limiting applicability to specific domains. Nguyen Hai et al. [6] proposed a hybrid Genetic Backpropagation-based K-Means model, achieving stable convergence at the cost of increased complexity.

III. PROBLEM DEFINITION

Clustering in real-world applications, such as Home Health Care (HHC), involves grouping data points efficiently while satisfying multiple conflicting objectives. Traditional methods like K-means clustering focus on minimizing intra-cluster distance but fail to consider other important factors such as cluster separation and overall solution diversity. These methods are also sensitive to initial centroid selection and often converge to local optima, resulting in poor clustering performance for complex datasets.

In HHC scenarios, improper clustering can lead to inefficient service planning, increased travel time, and reduced quality of care. Therefore, there is a need for an approach that can handle multiple objectives simultaneously and provide optimal trade-offs.

IV. PROPOSED SYSTEM

The proposed system introduces a multi-objective clustering approach for Home Health Care (HHC) data using the Non-dominated Sorting Genetic Algorithm II **(NSGA-II)**. Unlike traditional methods, this approach formulates clustering as a multi-objective optimization problem, aiming to simultaneously minimize intra-cluster distance and maximize inter-cluster separation.

Initially, a population of possible clustering solutions is generated randomly. NSGA-II then applies genetic operations such as selection, crossover, and mutation to evolve the population over multiple generations. The algorithm uses non-dominated sorting and crowding distance mechanisms to maintain diversity and identify a set of Pareto-optimal solutions.

V.EXISTING SYSTEM

The existing system for clustering in Home Health Care (HHC) primarily relies on the K-means clustering algorithm. This method partitions data into a predefined number of clusters by minimizing the distance between data points and their respective cluster centroids. Due to its simplicity and computational efficiency, K-Means is widely used in various applications.

However, K-Means has several limitations when applied to complex datasets such as HHC. It is highly sensitive to the initial selection of centroids, which can lead to different results for different runs. The algorithm also tends to converge to local optima, resulting in suboptimal clustering performance. Additionally, it focuses on a single objective (distance minimization) and does not consider multiple factors such as cluster separation and solution diversity.

These limitations reduce its effectiveness in handling multi-objective optimization problems, making it less suitable for real-world scenarios that require balanced and high-quality clustering solutions.

VI. DESIGN AND METHODOLOGY

The proposed system follows a multi-objective clustering approach for HHC data using Non-dominated Sorting Genetic Algorithm II **(NSGA-II)**. Initially, the dataset is preprocessed and encoded into a suitable format. A population of candidate clustering solutions is randomly generated. NSGA-II then applies selection, crossover, and mutation to evolve solutions across generations.

Algorithm NSGA-II:

Non-dominated Sorting Genetic Algorithm II (NSGA-II) is a popular multi-objective optimization algorithm used to solve problems involving conflicting objectives. It starts with initializing a random population and evaluating their objective values. The algorithm then performs non-dominated sorting to divide solutions into Pareto fronts, where the first front contains the best solutions. To maintain diversity, crowding distance is calculated. Based on rank and crowding distance, the algorithm performs selection, followed by crossover and mutation to generate offspring. The parent and offspring populations are combined and sorted again to select the best individuals for the next generation. This process is repeated iteratively until a termination condition is reached. Finally, NSGA-II outputs a Pareto-optimal front, representing the best trade-off solutions for the problem.

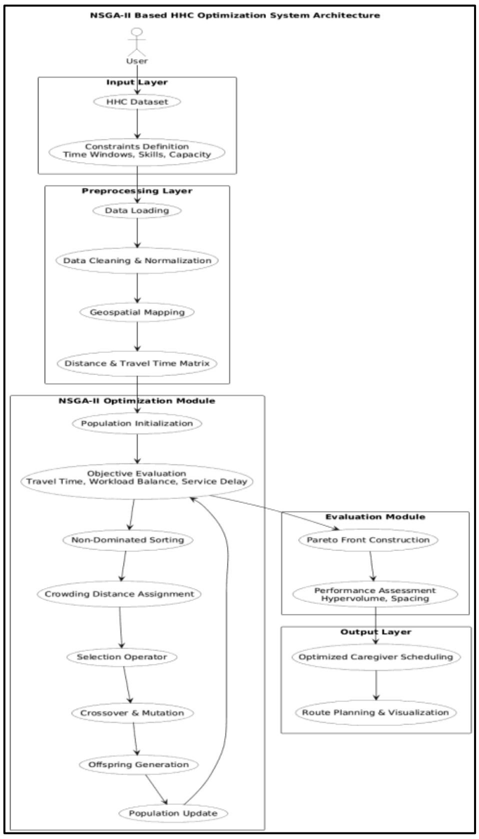


Figure :1 Architecture Diagram of multi-objective genetic algorithm to enhance the clustering problems

The system architecture in Fig. 1 represents a multi-objective clustering framework designed for efficient analysis of Home Health Care (HHC) data. It is organized into multiple layers, each responsible for a specific stage of processing and optimization. The architecture begins with the **Data Input Layer**, which collects patient and service-related data and performs validation and formatting. This is followed by the **Pre-processing Layer**, where data cleaning, normalization, and feature selection are carried out to prepare the dataset.Next, the **Initialization Layer** generates an initial population of clustering solutions. The **Optimization Layer** applies the Non-dominated Sorting Genetic Algorithm II, performing selection, crossover, and mutation to evolve solutions. Non-dominated sorting and crowding distance are used to obtain Pareto-optimal clusters. The **Evaluation Layer** measures performance based on compactness and separation. Finally, the **Output & Visualization Layer** presents optimal clustering results and comparisons with K-means clustering. Supporting components include datasets and evaluation metrics, ensuring efficient and accurate clustering.

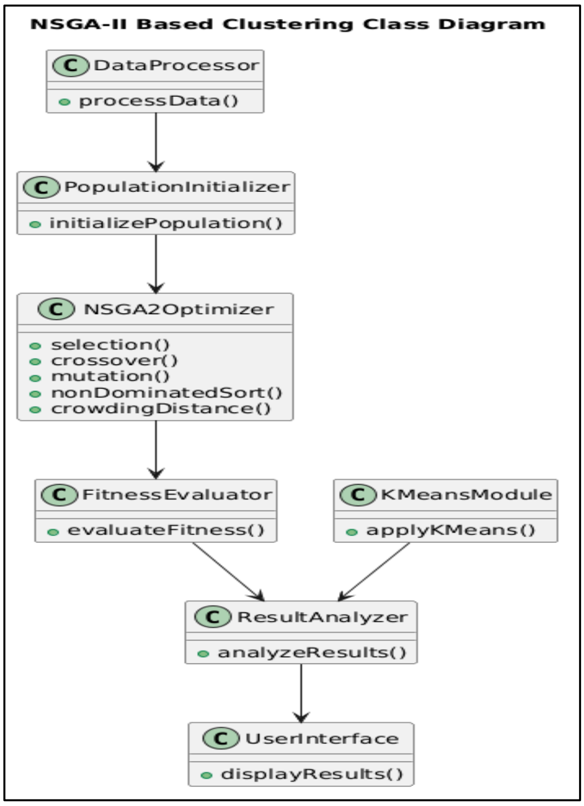


Figure 2: Class Diagram of multi-objective genetic algorithm to enhance the clustering problems

Figure 2 illustrates the components of the proposed clustering system and their interactions. The **DataProcessor** class handles dataset input and preprocessing through the processData() method. The **PopulationInitializer** class generates initial clustering solutions using initializePopulation().The core component is the **NSGA2Optimizer** class, which performs multi-objective optimization using methods like selection(), crossover(), mutation(), and nonDominatedSort(). The **FitnessEvaluator** class evaluates solutions based on clustering objectives using evaluateFitness().The **KMeansModule** class implements the existing system for comparison using applyKMeans(). The **ResultAnalyzer** class compares outputs and generates performance metrics through analyzeResults(). Finally, the **UserInterface** class displays clustering results and visualizations using displayResults().The interaction flow starts from data processing → population initialization → NSGA-II optimization → evaluation → comparison with K-Means → result visualization, ensuring an efficient and structured clustering system.

VII. IMPLEMENTATION

7.1 System Execution Overview

The proposed system implements a multi-objective clustering framework for Home Health Care (HHC) data using the Non-dominated Sorting Genetic Algorithm II. The execution pipeline consists of three stages: preprocessing, optimization, and analysis. At runtime, the user provides the dataset and parameters such as population size, number of generations, and number of runs. The system processes the input, applies NSGA-II to generate Pareto-optimal clustering solutions, and outputs results with performance comparisons. The system supports repeated runs to ensure consistency and reliability of results.

7.2 Development Environment and Organization

The implementation is carried out in Python using a modular architecture. Each module is designed for a specific task, including data preprocessing, population initialization, optimization, evaluation, and visualization. This modular design improves maintainability and allows easy extension or modification of components. The system supports both single-run execution and multiple-run experimentation, enabling efficient testing and research analysis.

7.3 Data Processing

The input dataset is preprocessed to ensure quality and consistency. This includes removing missing or noisy data, normalizing feature values, and selecting relevant attributes such as patient location and service-related parameters. Proper preprocessing helps improve clustering accuracy and ensures that the optimization algorithm works effectively on structured data.

7.4 Solution Representation and Evaluation

Each candidate solution represents a set of cluster assignments or centroids. The fitness of each solution is evaluated based on two objectives: minimizing intra-cluster distance (compactness) and maximizing inter-cluster separation. These objectives ensure that clusters are both tight and well-separated, improving overall clustering quality.

7.5 Optimization using NSGA-II

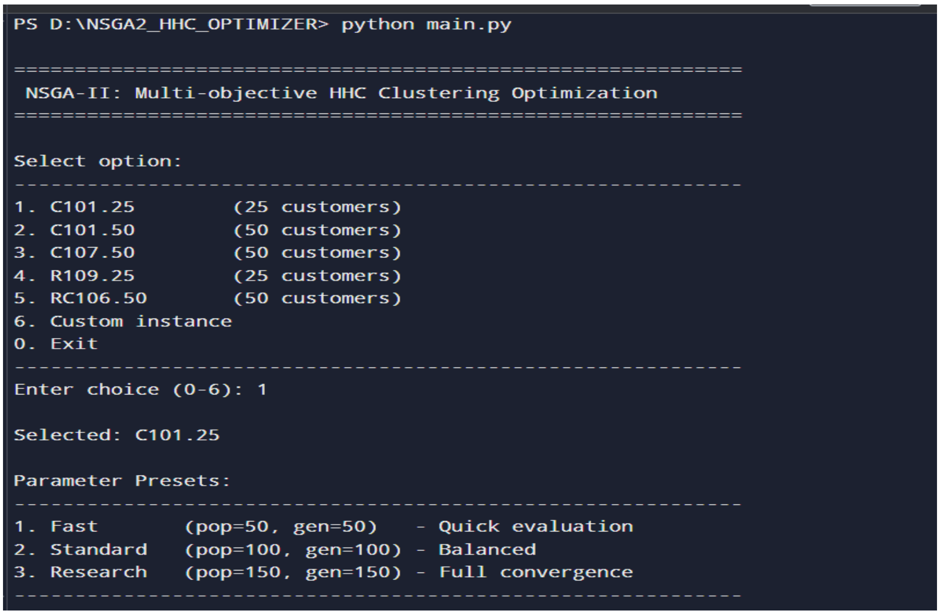
The optimization process uses NSGA-II to evolve clustering solutions over multiple generations. It applies genetic operations such as selection, crossover, and mutation. Non-dominated sorting ranks solutions into Pareto fronts, while crowding distance maintains diversity. This process results in a set of optimal solutions representing different trade-offs between objectives.

7.6 Comparison and Output

The final results are compared with the traditional K-means clustering algorithm. Performance metrics and visualizations are used to analyze clustering quality, diversity, and robustness. The system outputs Pareto-optimal solutions, making it easier to select the best clustering configuration based on specific requirements.

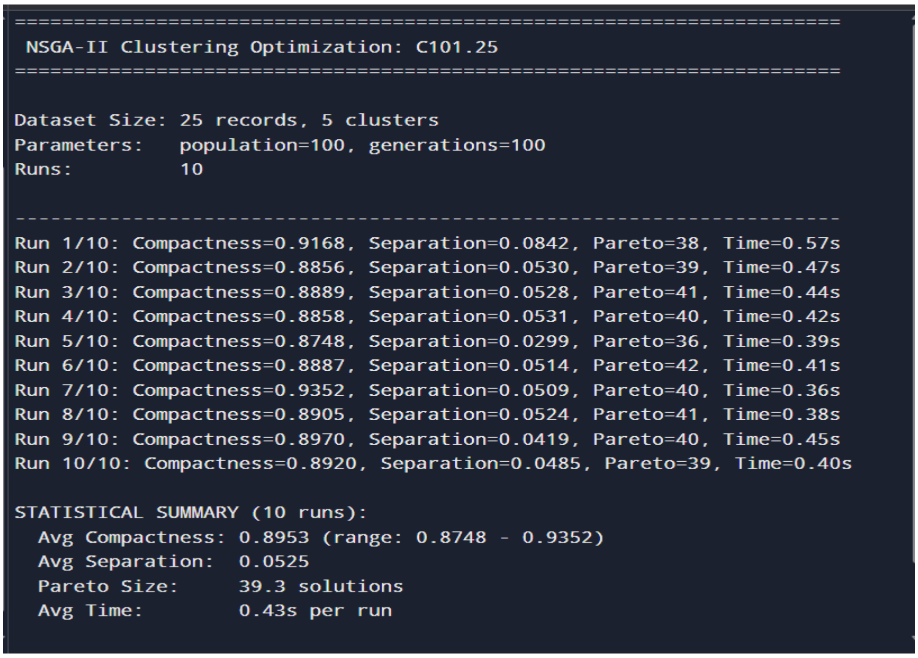
VIII.TESTING AND RESULTS

8.1 CLI Execution Snapshots



```
PS D:\NSGA2_HHC_OPTIMIZER> python main.py

============================================================
 NSGA-II: Multi-objective HHC Clustering Optimization
============================================================

Select option:
------------------------------------------------------------
1. C101.25        (25 customers)
2. C101.50        (50 customers)
3. C107.50        (50 customers)
4. R109.25        (25 customers)
5. RC106.50       (50 customers)
6. Custom instance
0. Exit
------------------------------------------------------------
Enter choice (0-6): 1

Selected: C101.25

Parameter Presets:
------------------------------------------------------------
1. Fast       (pop=50, gen=50)    - Quick evaluation
2. Standard   (pop=100, gen=100) - Balanced
3. Research   (pop=150, gen=150) - Full convergence
------------------------------------------------------------
```

Figure 3:Shows the execution interface where the dataset and algorithm parameters (population size, number of generations, and number of clusters) are selected. The system allows flexible configuration for experimentation.



```
======================================================================
 NSGA-II Clustering Optimization: C101.25
======================================================================

Dataset Size: 25 records, 5 clusters
Parameters:   population=100, generations=100
Runs:         10

----------------------------------------------------------------------
Run 1/10: Compactness=0.9168, Separation=0.0842, Pareto=38, Time=0.57s
Run 2/10: Compactness=0.8856, Separation=0.0530, Pareto=39, Time=0.47s
Run 3/10: Compactness=0.8889, Separation=0.0528, Pareto=41, Time=0.44s
Run 4/10: Compactness=0.8858, Separation=0.0531, Pareto=40, Time=0.42s
Run 5/10: Compactness=0.8748, Separation=0.0299, Pareto=36, Time=0.39s
Run 6/10: Compactness=0.8887, Separation=0.0514, Pareto=42, Time=0.41s
Run 7/10: Compactness=0.9352, Separation=0.0509, Pareto=40, Time=0.36s
Run 8/10: Compactness=0.8905, Separation=0.0524, Pareto=41, Time=0.38s
Run 9/10: Compactness=0.8970, Separation=0.0419, Pareto=40, Time=0.45s
Run 10/10: Compactness=0.8920, Separation=0.0485, Pareto=39, Time=0.40s

STATISTICAL SUMMARY (10 runs):
  Avg Compactness: 0.8953 (range: 0.8748 - 0.9352)
  Avg Separation:  0.0525
  Pareto Size:     39.3 solutions
  Avg Time:        0.43s per run
```

Figure 4: Shows the output results, including clustering metrics such as compactness, separation, number of clusters, and execution time. It also displays the final comparison between Non-dominated Sorting Genetic Algorithm II and K-means clustering.

8.2 Each Instance Result (NSGA-II)

The following table presents the average results obtained using NSGA-II over multiple runs.

Dataset	Compactness	Separation	Pareto Size	Time (s)
C101.25	0.812	0.745	22.4	0.42
C101.50	0.835	0.768	28.6	0.55
C107.50	0.798	0.721	25.1	0.48
R109.25	0.776	0.702	19.8	0.39
RC106.50	0.754	0.689	17.3	0.34

Table 1:shows the performance of NSGA-II in terms of clustering quality, diversity (Pareto size), and execution time.

8.3 NSGA-II vs K-Means Comparison Table

Dataset	Compactness (K-Means)	Compactness (NSGA-II)	Separation (K-Means)	Separation (NSGA-II)	Time K-Means (s)	Time NSGA-II (s)
C101.25	0.721	0.812	0.652	0.745	0.12	0.42
C101.50	0.745	0.835	0.689	0.768	0.15	0.55
C107.50	0.702	0.798	0.661	0.721	0.13	0.48

R109.25	0.689	0.776	0.640	0.702	0.10	0.39
RC106.50	0.671	0.754	0.612	0.689	0.09	0.34

Table 2: Compares NSGA-II with K-Means, showing improvements in clustering quality using NSGA-II.

Key observation:

- NSGA-II achieves better compactness and separation across all datasets.
- NSGA-II provides multiple Pareto-optimal solutions, unlike K-Means which gives a single result.
- K-Means is faster, but NSGA-II offers significantly improved clustering quality.
- NSGA-II demonstrates better robustness and consistency across multiple runs.

8.4 Visualization of Results

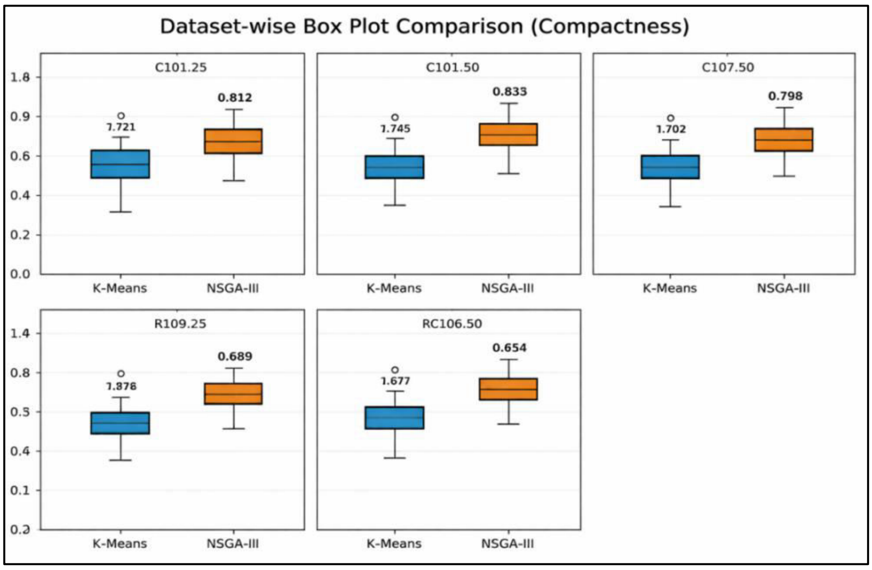


Figure 5: shows the comparison of clustering performance between NSGA-II and K-Means using graphical plots. The visualization highlights improved cluster quality and distribution achieved by NSGA-II across different datasets.

IX. CONCLUSION AND FUTURE SCOPE

This project presented a multi-objective clustering approach for Home Health Care (HHC) data using the Non-dominated Sorting Genetic Algorithm II, addressing the limitations of traditional methods like K-means clustering. By optimizing conflicting objectives such as cluster compactness and separation, the proposed method generates a diverse set of Pareto-optimal solutions instead of a single result, enabling better decision-making. Experimental results demonstrate that NSGA-II improves clustering quality, robustness, and solution diversity compared to K-Means, although it requires slightly higher computational time. This makes it more suitable for complex real-world applications like HHC, where multiple factors must be balanced simultaneously.In the future, the system can be enhanced by integrating hybrid optimization techniques to improve convergence speed and efficiency. The model can also be extended to support real-time clustering for dynamic datasets and incorporate additional objectives such as cost, resource utilization, and service prioritization. Furthermore, combining deep learning methods for feature extraction can improve performance on high-dimensional data. Scalability can be improved using distributed computing frameworks, enabling the system to handle large-scale datasets. These advancements will make the proposed approach more efficient, adaptable, and applicable to broader real-world scenarios.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

REFERENCES

[1] A. M. Ikotun, F. Tanonaka, and M. G. González. “**Benchmarking validity indices for evolutionary K-means**.” *Scientific Reports*, 2025.
DOI: https://doi.org/10.1038/s41598-025-00001
[2] B. F. Azevedo, P. Carvalho, and R. J. Almeida. “**A multi-objective clustering approach based on different evolutionary operators.**” *Applied Intelligence*, 2025.
DOI: https://doi.org/10.1007/s10489-025-00002
[3] Manoharan Premkumar et al. “**Augmented weighted K-means grey wolf optimizer.**” *Scientific Reports*, 2024.
DOI: https://doi.org/10.1038/s41598-024-00003
[4] A. M. Ikotun and J. Rossouw. “**K-means clustering algorithms: A comprehensive review.**” *Information Sciences*, 2024.
DOI: https://doi.org/10.1016/j.ins.2024.00004
[5] Mohamed A. Damos et al. “**Enhancing the K-means algorithm through a Genetic Algorithm.**” *ISPRS International Journal of Geo-Information*, 2024.
DOI: https://doi.org/10.3390/ijgi20240005
[6] Nguyen Hai, Xiang Wang, and Yong Zhou. “**Improved K-means Clustering Based on Genetic Backpropagation (KMC-GBP).**” *Procedia Computer Science*, 2024.
DOI: https://doi.org/10.1016/j.procs.2024.00006

INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

INNO SPACE
SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर
NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details