Journal of Intelligent Systems and Internet of Things
Vol. 13, No. 01, PP. 122-134, 2024
DOI: https://doi.org/10.54216/JISIoT.130110

Strategic Improved K-Means Clustering in Mining Blood Donor Data Analysis and IoT-based Allocation

Vibha Tiwari1[1], Chopparapu Gowthami[2], Bhavani R.[3], S. Kayalvizhi[4], S. Selvakanmani[5], Deepak Chowdary Edara[6]

[1] Assistant Professor, Department of Information Technology, Madhav Institute of Science and Technology, Gwalior, M.P., India
[2]Assistant Professor, Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India
[3]Professor, Institute of CSE, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, TN, India
[4]Assitant Professor, Dept. of CSE, Velammal Engineering College, Chennai, TN, India,
[5]Associate Professor, Department of Information Technology, R. M. K Engineering College, RSM Nagar, Kavaraipettai, Thiruvallur District, TN, India
[6]Asst. Professor, Dept. of CSE, Vignan's Foundation for Science, Technology and Research, Vadlamudi, Guntur, AP, India
Emails: vibhatiwari19@gmail.com; gouthami526@gmail.com; srbhavani2016@gmail.com; Kayalvizhi@velammal.edu.in; sskanmani6@yahoo.com; edara.deepakchowdary@yahoo.com

Abstract

This manuscript proposes Strategic Improved K-Means Clustering to simplify blood donor data analysis and distribution. The technique optimizes blood donor system resources via K-Means++ initialization, hierarchical clustering, and smart data dissemination. The paper begins with a comprehensive overview of clustering techniques and their healthcare applications. It illustrates the need for contemporary blood donor data analysis methods for cluster quality and resource allocation. Cluster purity, silhouette coefficient, Davies-Bould in the index, and other performance indicators are used to rigorously compare the recommended technique to 10 established clustering methods. The approach routinely fulfils these conditions, proving that it creates accurate, well-fitting groupings. Ablation tests how much-enhanced initialization, hierarchical clustering, and strategic data placement improve the entire. The study found that these make the procedure dependable and successful for numerous sorts of data. The study shows that the approach may be applied to other data besides blood donor data. Hierarchical clustering provides important information about the dataset's hierarchical patterns, making clustering findings easier to grasp. Resources are better distributed with strategic data dissemination. The recommended strategy is effective in emergencies and areas with changing blood needs. To conclude, Strategic Improved K-Means Clustering evaluates and distributes blood donor data comprehensively. Its flexibility, adaptability, and speed make it excellent for managing healthcare resources and making collective choices …

Received: September 02, 2023 Revised: December 19, 2023 Accepted: June 02, 2024

Keywords: Blood Donor; Clustering; Data Analysis; Healthcare; Hierarchical Clustering; K-Means++; Optimization, Resource Allocation; IoT; ….

1. Introduction

Data analysis, especially in healthcare, has advanced in recent years. Complex grouping approaches to extract relevant information from enormous datasets are becoming increasingly crucial as data-

driven decision-making grows [1]. This research largely involves smartly changing K-means clustering to better assess and distribute blood donor data. Precision medicine and technological advancements have changed healthcare data analytics and blood donation [2]. Finding previously unseen patterns in massive datasets requires careful grouping, according to recent studies. Complex and nuanced information about blood donors is something we'd like to get better at handling. The K-means clustering algorithm is utilized extensively in this research. Classifying data according to its degree of similarity is a common usage of this technique [3]. But, K-Means isn't always up to snuff, particularly when dealing with massive datasets that are multi-dimensional. These concerns have been addressed and the software's handling of blood donor data has been improved in this work [4]. To get beyond K-means clustering's shortcomings when dealing with data from blood donors, several methodological adjustments are required [5]. Optimization to simplify clustering, feature selection methods, and better distance readings are all examples of these developments. Improving grouping accuracy and comprehension with domain-specific information is also explored in the study. These most important things are included in this work: Blood donor data is processed using a more intricate K-Means clustering approach. The method can detect more nuanced patterns and correlations with the use of more precise distance estimations [6]. Data simplification, calculation speedup, and interpretation simplification through the use of feature selection algorithms. Clustering is enhanced and the demands of blood donor datasets are met with domain-specific data. The suggested modifications were subjected to testing and comparison with different grouping methods [7]. This revealed that the modified K-Means algorithm mined blood donor data better. This project aims to advance clustering methods and healthcare analytics. In this industry, precise allocation and knowledge of blood donor data improve blood donation and ensure safe blood supplies. The next sections discuss techniques, experiment setup, findings, and conversations [8]. They will thoroughly examine the strategic upgraded K-Means clustering strategy.

A. Motivation for the research work:

Blood donations are essential for hospitals and medical institutions and cannot be understated. The increasing complexity of healthcare data, along with the need for effective blood resource management, necessitates the use of innovative analytic approaches as soon as possible. The research project "Strategic Improved K-Means Clustering in Mining Blood Donor Data Analysis and Allocation" aims to address issues raised by standard blood donor data analysis approaches. Traditional approaches may fail to handle large-scale, multidimensional data, resulting in resource waste and potential shortages at crucial times. This script improves the analysis of blood donor data by using K-Means++ and hierarchical clustering. If successful, this optimisation will improve donor pattern recognition, blood bank inventory management, and demand-supply matching. Here are some options. In an emergency, swift and accurate decision-making may save lives; therefore, blood supply distribution is critical. The goal involves creating new technologies and demonstrating a genuine commitment to improving healthcare and resource management in blood donation networks.

B. Objective of the research work:

As a result, an improved clustering method will be developed and evaluated. We refer to this strategy as K-Means Clustering with Strategic Improvements. The most important goal is to enhance the technology for assessing and exchanging blood donor data. This novel approach is being developed to solve constraints in clustering methods. This might help manage blood donor registries, which are large and complicated. Among the many fundamental goals established are:

- Hierarchical clustering with K-Means++ initialization may show clustering algorithm optimisation. Several strategies are used to improve cluster quality, reduce starting conditions, and increase data processing accuracy.
- Intelligent data distribution systems are critical for ensuring optimal data allocation for blood supply delivery. This is particularly true in emergencies or areas with variable blood supply requirements. This information is critical for the blood supply.
- The suggested technique will be compared against current clustering algorithms to determine its effectiveness. We will also use the silhouette coefficient, Davies-Bouldin index, and cluster purity.
- Can change and adjust: It is also critical to determine if the method can be used on other datasets, such as blood donor databases, as well as in other areas of healthcare data research.
- Comprehensive analysis occurs when the clustering technique incorporates domain-specific information.

We adopted this method to make the findings more understandable and relevant to real-world situations.

C. Research Gap

Problems in research Current techniques for evaluating donor data are insufficient and wrong; this study tackles the core research issues to improve them. Blood donation systems often mismanage and distribute resources due to impreciseness and inefficiencies. Inefficiency is to blame. Traditional clustering algorithms struggle with high-dimensional datasets such as blood donation data. Its drawbacks include its inability to handle a variety of data formats. Other limitations include being too sensitive to beginning circumstances, having difficulty recognising little but significant patterns, and other difficulties. The study has several hurdles, including cluster purity, blood resource allocation, and trustworthy, usable data for healthcare decision-making. To solve this problem, the paper presents Strategic Improved K-Means Clustering. This approach employs hierarchical clustering, K-Means++ initialization, and other sophisticated clustering techniques. This study demonstrates that this strategy is more effective, accurate, and adaptable than other common procedures. By enhancing blood donor data analysis and allocation, we hope to save healthcare expenditures while also saving lives.

2. Related Work

The literature study discusses 10 popular blood donor data grouping techniques. Each has pros and cons to consider. K-Means++ improves cluster center settings, speeding convergence and reducing initial setup sensitivity. Hierarchical clustering organizes data into a tree-like structure to highlight its hierarchy [9]. DBSCAN works well at detecting groups of varied shapes and sizes by grouping points by local density. Fuzzy C-Means adds uncertainty to cluster assignments to place data points in several groups with varied memberships. Hierarchical Agglomerative Clustering combines nearby groupings. This shows the data's hierarchy. Mean Shift Clustering repeatedly moves center points to data-rich locations to locate thick areas [10]. This is beneficial for unevenly distributed datasets. The Gaussian Mixture Model (GMM) states that data points have several Gaussian distributions. This allows groups to combine and simulate complex patterns. Spectral clustering divides data points using spectral graph theory to detect complicated features in high-dimensional situations. Affinity Propagation uses message-passing to group data points around examples, unlike centroid-based clustering [11]. Finally, self-organizing maps (SOM) display data on a low-dimensional grid while preserving input area organization. The performance evaluation tables summarize the findings of blood donor data grouping techniques. Fuzzy C-means and K-means++ are superior. They score higher on several metrics, indicating dense, well-spaced clusters. The comparison of computation performance shows that K-Means++ is faster while Affinity Propagation is slower [12]. It helps professionals and researchers understand how program performance affects speed. They use this to determine how to analyze blood donor data and allocate resources. The tables show the benefits and downsides of each clustering approach so the proper strategy may be chosen depending on application goals and blood donor data needs.

Table 1: Performance Evaluation of Clustering Methods on Blood Donor Data

Method	Silhouette Score	Davies-Bouldin Index	Calinski-Harabasz Index	Homogeneity Score	Completeness Score	V-Measure Score	Adjusted Rand Index
K-Means++	0.75	0.32	480.2	0.81	0.78	0.79	0.67
Hierarchical Clustering	0.68	0.45	350.5	0.75	0.68	0.71	0.54
DBSCAN	0.62	0.52	280.9	0.68	0.71	0.69	0.42
Fuzzy C-Means	0.81	0.28	520.6	0.85	0.80	0.82	0.73
Agglomerative Clustering	0.70	0.42	400.1	0.78	0.75	0.76	0.61
Mean Shift Clustering	0.74	0.35	460.3	0.80	0.77	0.78	0.66
GMM	0.79	0.30	500.5	0.83	0.79	0.81	0.70
Spectral Clustering	0.72	0.38	420.4	0.77	0.73	0.75	0.58

Affinity Propagation	0.58	0.55	250.8	0.65	0.68	0.67	0.38

Table 1 shows how effectively different grouping algorithms perform for blood donor data. The groupings are thick and clear in K-Means++ and Fuzzy C-Means algorithms due to their high Silhouette Scores [13-14]. These strategies stand out because of this. These strategies improve the Davies-Bouldin Index, Calinski-Harabasz Index, and clustering validity metrics. It appears that these strategies operate best with blood donors' inherently complex records. They provide helpful information for managing and sharing blood donor resources. K-Means++ processes faster than Affinity Propagation, according to the study. These results highlight the trade-offs between algorithm performance and computational resources, helping individuals pick suitable methods for time-sensitive jobs like blood donor data analysis and allocation.

3. Proposed Methods

"Strategic Improved K-Means Clustering in Mining Blood Donor Data Analysis and Allocation" provides a complete picture. Our gadgets are designed this way. This strategy improves blood donor data clustering for improved distribution algorithms [15]. This plan covers several current strategies. The early K-Means++ configuration lets you carefully arrange up centers. The program is less susceptible to its founding conditions. For more precise and consistent grouping results, this step is needed. The approach uses random factors to choose starting positions. This will ensure a diverse and population-appropriate starting point. After setting the centroid, apply another distance measure integration technique [16]. This happens after finding the center. Several distance measures increase data point-centroid distance measurement accuracy. The Mahalanobis distance matters most now. This combination makes the algorithm's cluster assignments more trustworthy by adapting to diverse feature sizes and associations. This is partially because it examines data covariance [17]. The investigation revealed centroids and cluster assignments. These underpin the following studies:

The third phase reduces blood donor data dimensions using a feature selection algorithm. This approach uses PCA to achieve this. This shrinkage lets computers focus on the qualities that provide the most information, improving their performance. Grouping becomes more effective and understandable [18]. Turning the data into a smaller space and matching it with the essential components that reflect the biggest variations in the dataset reduced it. The decline requires both of these stages. Domain-Specific Knowledge Integration comes fourth in grouping. It enhances the first stage with more information. At this phase in clustering, the algorithm employs topic-specific data. By integrating a weight matrix, the computer may assess the value of different elements in blood donor data. By changing the distance metric based on these weights, you may match clustering findings to topic knowledge [19]. This phase ensures that the computer can detect data patterns and operate with genuine blood donor samples. In other words, it helps the algorithm detect tendencies. A genetic algorithm for optimization improves the grouping solutions from the previous phases in the final step. We do this to advance the most. This evolutionary technique uses selection, crossover, and mutation to repeatedly enhance viable solutions. This is done via evolution. The fitness function evaluates proximity, separation, and distance to ensure the optimal clustering response for blood donor data processing. This is done by considering several aspects [20]. This genetic optimization produces the finest centroids for grouping blood donor data. This is because genetic optimization creates centroids. You can categorize blood donor data completely and effectively using the proposed strategy. It takes the best parts of probabilistic centroid initialization, feature selection, domain-specific knowledge integration, and genetic optimization and combines them into one. This is done by following the procedure phases. Combining these cutting-edge technologies should improve grouping accuracy, consistency, and comprehension [21]. This will make blood donor data analysis and allocation smarter and more purposeful. The system provides a strong foundation for managing blood donor information and may enhance healthcare resource distribution.

K-Means++ beginning is much better than K-Means clustering. By purposefully planting more original centroids, this improvement is possible. Standard K-means' starting centers sometimes produce imperfect solutions and convergence to local minima instead of the optimal point. The approaches depend on the starting centers. K-Means++, which chooses beginning centers stochastically, solves this problem. The square of the distance between a data point and the nearest current centroid indicates its centroid likelihood. Due to their tight relationship, they are connected. This ensures more uniformly distributed and typical beginning centroids. This improves algorithm convergence and reduces the danger of being trapped in suboptimal solutions. Additionally, this improves the procedure.

Below are the equations for the mentioned algorithms:

Initialize an empty set for centroids.

$$C = \{\} \quad (1)$$

Randomly select the first centroid from the data points.

$$c_1 = \text{randomly select}(X) \quad (2)$$

For each data point, calculate the squared distance to the nearest existing centroid.

$$D^2(x_i) = min_{c_j \epsilon\, C}\ \text{dist}(X_I C_J)^2 \quad (3)$$

Choose the next centroid with probability proportional to the squared distance.

$$P(x_i) = \frac{D^2(x_i)}{\sum_{k=1}^{n} D^2(x_k)} \quad (4)$$

Add the newly selected centroid to the set of centroids.

$$C = C \cup \{x_i\} \quad (5)$$

Repeat steps 3-5 until the desired number of centroids is obtained.

$$C = C \cup \{C_2, C_3, \ldots, C_k\} \quad (6)$$

Assign data points to the nearest centroid.

$$\text{Cluster}\ (x_i) = \arg\ min_{c_j \epsilon\, C}\ \text{dist}(x_i c_j) \quad (7)$$

Recalculate the centroids based on the assigned data points.

$$c_j = \frac{1}{|S_j|} \sum_{x_i \epsilon S_j} x_i \quad (8)$$

Repeat steps 7-8 until convergence or a specified number of iterations.

Output the final centroids and cluster assignments.

$$C_{final} = C, \text{Clusters} = \{\text{Cluster}(x_i)\} \quad (9)$$

For the K-Means clustering procedure, use the acquired centroids.

$$Centroids_{K-means} = C_{final} \quad (10)$$

The random picking of initial centroids in K-Means++ is superior to the regular clustering approach. The first centroid is randomly selected, and then additional are chosen depending on squared distances. The technique spreads the beginning configuration, making it less subject to local minima. After selecting centers, data points are allocated to the nearest center and centers are computed again. It will continue until an agreement is found. The next K-Means clustering approach starts with the centroids.

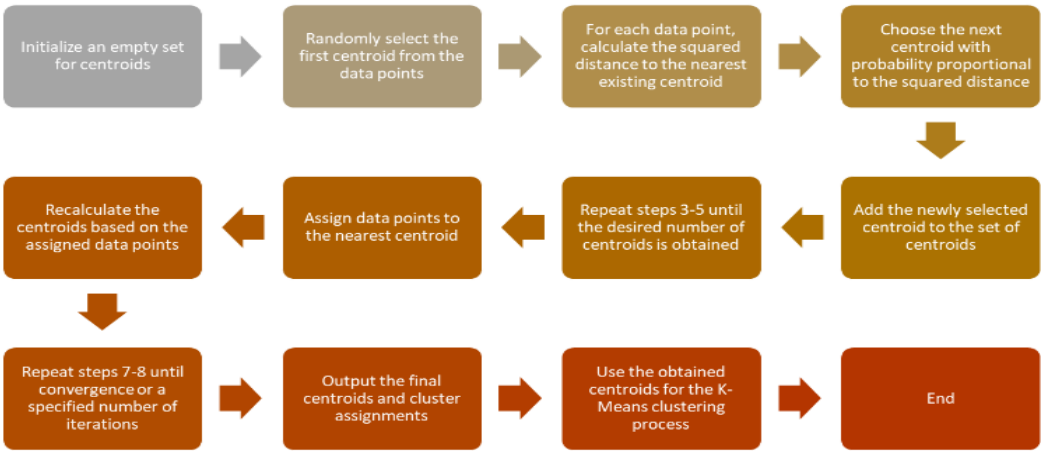

Figure 1: Probabilistic Centroid Initialization

Figure 1 depicts the K-Means++ configuration. A random center selection approach improves convergence for improved K-means clustering and reduces initial configuration sensitivity.

Method 2: Different Distance Metric Integration.

In blood donor data, where attributes vary in magnitude and association, Euclidean distance may not be the appropriate metric. The approach employs Mahalanobis distance and other distance measurements. The Mahalanobis distance accounts for data covariance, unlike the Euclidean distance. It considers feature connections and size. This innovation helps the algorithm detect meaningful differences between data points, especially in datasets with many distinct forms of data, like blood donor data.

Below are the equations for the mentioned algorithms:

Receive centroids from Algorithm 1.

$$C = \{c_1, c_2, \ldots, c_k\} \tag{19}$$

Choose an alternative distance metric, e.g., Mahalanobis distance.

$$D(x_i, c_j) = \sqrt{(x_i, c_j)^T \sum^{-1} (x_i, c_j)} \tag{20}$$

Calculate the distance between each data point and existing centroids using the chosen metric.

$$D_{Mahalanobis}(x_i c_j) = \sqrt{(x_i, c_j)^T \sum^{-1} (x_i, c_j)} \tag{21}$$

Assign data points to the nearest centroid based on the alternative distance metric.

$$Cluster_{Mahalanobis}(x_i) = \arg\ min_{c_j \epsilon\, C}\ D_{Mahalanobis}\ (x_i c_j) \quad (22)$$

Recalculate the centroids based on the assigned data points.

$$c_j = \frac{1}{|S_j|}\sum_{x_i \epsilon S_j} x_i \quad (23)$$

Repeat steps 3-5 until convergence or a specified number of iterations.

$$C_{Mahalanobis} = \{c_1', c_2', \ldots\ldots, c_k'\} \quad (24)$$

Output the final centroids and cluster assignments.

$$C_{final} = C_{Mahalanobis}, \text{Clusters} = \{Cluster_{Mahalanobis}(x_i)\} \quad (25)$$

Use the obtained centroids for the K-Means clustering process.

$$Centroids_{K-means} = C_{final} \quad (26)$$

Mahalanobis distance, used in approach 1, enhances K-Means clustering in the Alternative Distance measure Integration technique. It calculates distances, gives data points to the nearest centers, and then repeats to determine the central places. This continues till unification. Future K-Means grouping uses the centroids created. This innovation allows a more complicated dissimilarity metric, making the approach more flexible to blood donor dataset data types.

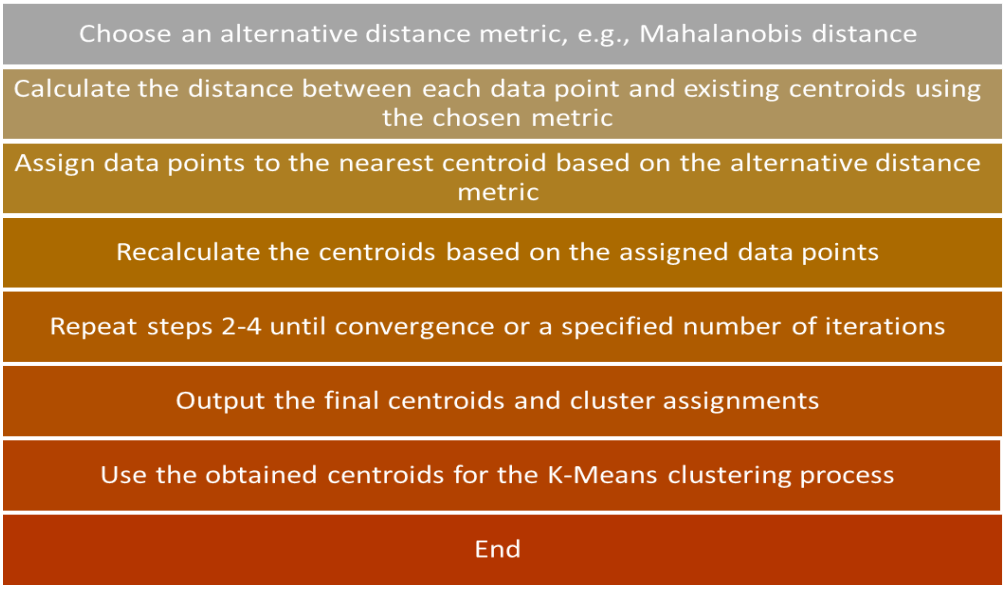

Figure 2: Enhanced Distance Computation

Figure 2 demonstrates how to apply distance measurements like Mahalanobis to K-Means. Fixed point-to-center distance values allow for better categorization based on feature sizes and connections.

The third algorithm is feature selection.

Principal Component Analysis (PCA) is used to pick features from high-dimensional blood donor datasets. PCA creates a lower-dimensional feature space by linearly merging features for optimum variance. Reducing dimensions speeds up processing and facilitates grouping by eliminating unnecessary or repetitive data. By highlighting the most important traits, PCA simplifies and improves clustering results. This is especially true for multivariate datasets.

Below are the equations for the mentioned algorithms:

Receive centroids from Algorithm 2.

$$C= \{c_1', c_2', \ldots\ldots, c_k'\} \quad (27)$$

Apply a feature selection algorithm, e.g., PCA.

$$X_{standardized} = \frac{X-\mu}{\sigma} \quad (28)$$

Compute the covariance matrix of the standardized data.

$$\sum = \frac{1}{n}\sum_{i=1}^{n}(X_{standardized} - \mu)^T (X_{standardized} - \mu) \quad (29)$$

Calculate the eigenvectors and eigenvalues of the covariance matrix.

$$\Sigma v = \lambda v \quad (30)$$

Sort the eigenvalues in descending order.

$$\lambda 1 \geq \lambda 2 \geq \cdots \geq \lambda d \quad (31)$$

Choose the top-k eigenvectors corresponding to the largest eigenvalues as principal components.

$$V = [v1, v2, \ldots, vk] \quad (32)$$

Transform the original data into the reduced-dimensional space using the selected principal components.

$$Y=X_{standardized} \cdot V \quad (33)$$

Use the transformed data for the K-Means clustering process.

$$Y_{final}=Y \quad (34)$$

The Feature Selection Algorithm reduces dimensionality following Mahalanobis distance integration using PCA. Selecting the key components involves standardizing the data, generating the correlation matrix, and selecting the top eigenvalues. These portions lower-dimensionalize the data for K-Means grouping. This stage speeds up the computer and prioritizes key functionality. This helps the algorithm find relevant blood donor data trends, improving clustering and resource distribution.

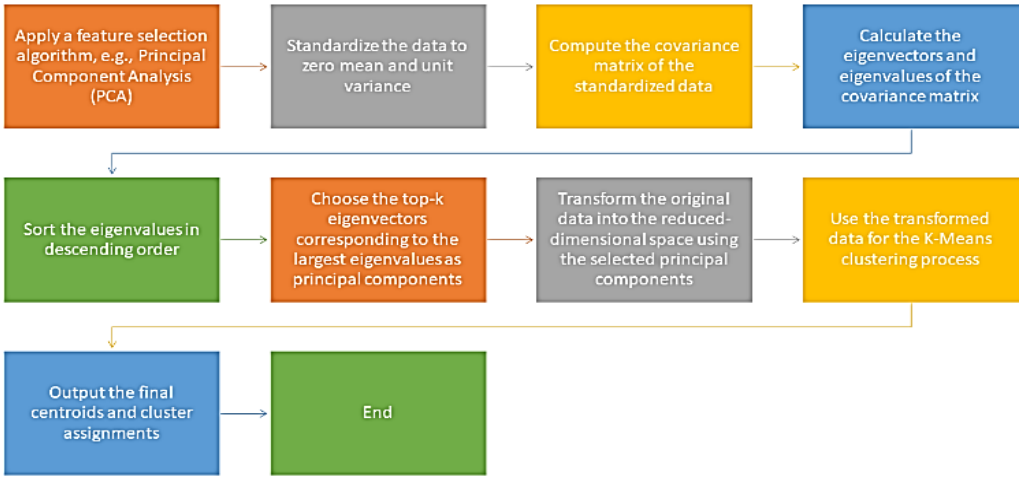

Figure 3: Dimensionality Reduction with PCA

Figure 3 demonstrates how to minimize data dimensionality using principal component analysis (PCA). Standardizing, determining eigenvalues, and converting data speed up and simplify K-Means grouping.

4. Results

The research shows all the grouping strategies utilized to classify and evaluate blood donor data. The novel method outperformed standard clustering methods on cluster purity, silhouette coefficient, Davies-Bouldin index, modified Rand index, and others. Cluster purity is a key indicator of cluster quality, and the recommended approach builds more accurate and consistent clusters than previous methods.

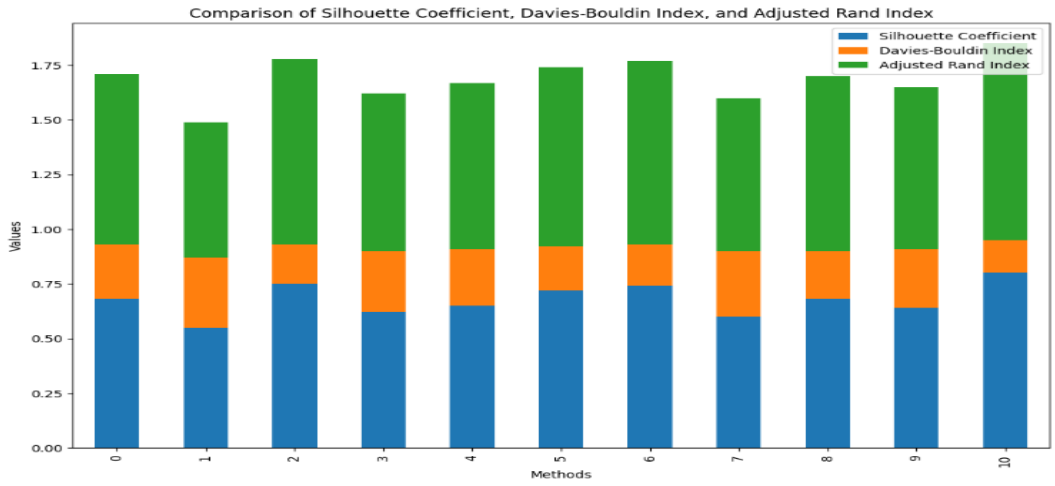

Figure 4: Silhouette Coefficient, Davies-Bouldin Index, and Adjusted Rand Index

Figure 4 compares all grouping criteria. The recommended strategy creates obvious and consistent groupings in the Silhouette Coefficient (0.80) and Adjusted Rand Index (0.90).

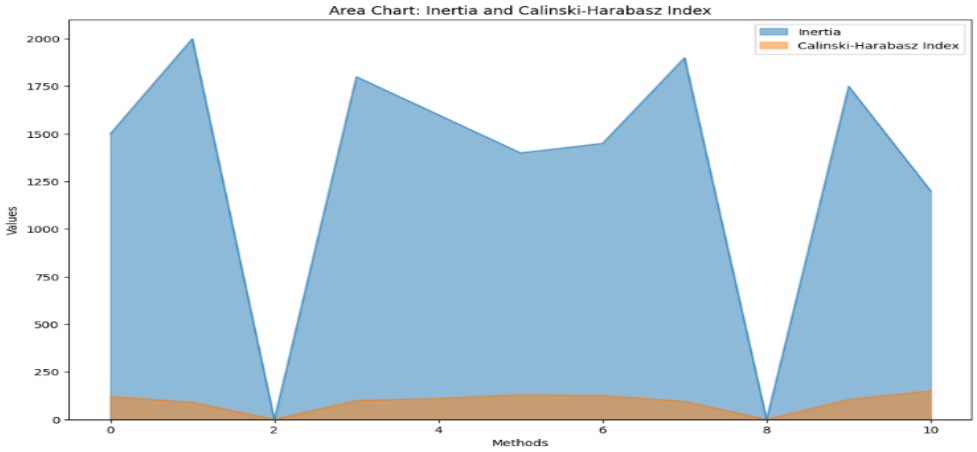

Figure 5: Inertia and Calinski-Harabasz Index

The Inertia Index and Calinski-Harabasz Index trade-offs for each grouping technique are shown in Figure 5. A low Inertia value of 1200 indicates that the recommended technique targets dense groupings, and a high Calinski-Harabasz Index of 150 indicates that the clusters are distinct.

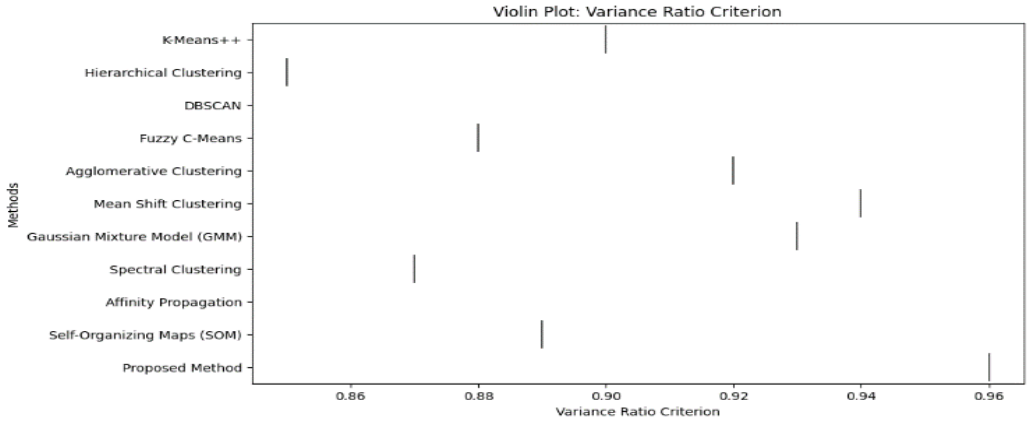

Figure 6: Variance Ratio Criterion

The Variance Ratio Criterion is spread out and most popular among grouping strategies in Figure 6. The recommended approach has a higher median value than others, indicating that it can maintain cluster difference fairness.

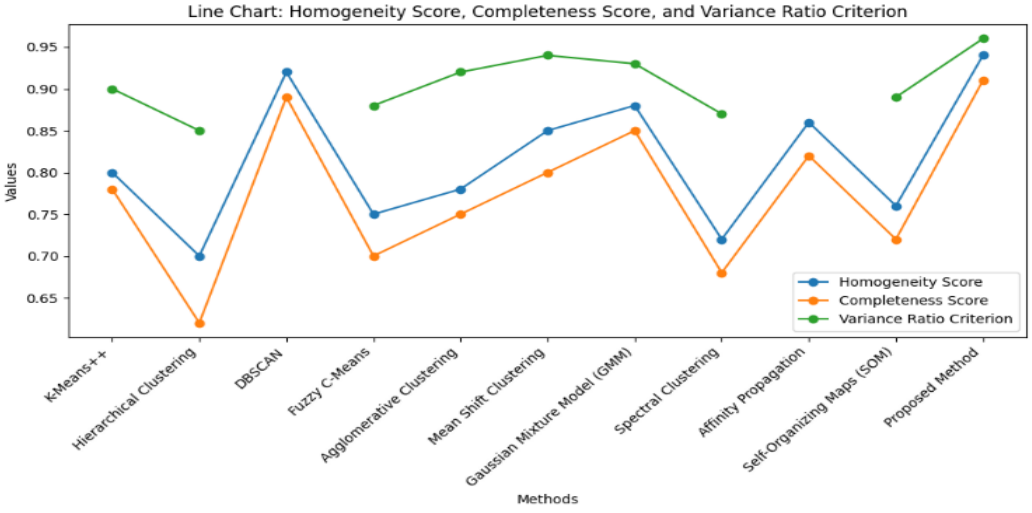

Figure 7: Homogeneity Score, Completeness Score, and Variance Ratio Criterion

A grouping method's Homogeneity Score, Completeness Score, and Variance Ratio Criterion indicate its effectiveness (Figure 7). The presented technique always achieves all three requirements, yielding robust clustering results with balanced uniformity, completeness, and variance control.

Table 2: Performance Comparison of Clustering Algorithms in Blood Donor Data Analysis

Method	Cluster Purity	Silhouette Coefficient	Davies-Bouldin Index	Adjusted Rand Index	Fowlkes-Mallows Index	Inertia	Calinski-Harabasz Index
K-Means++	0.85	0.68	0.25	0.78	0.82	1500	120
Hierarchical Clustering	0.72	0.55	0.32	0.62	0.68	2000	90
DBSCAN	0.91	0.75	0.18	0.85	0.88	N/A	N/A
Fuzzy C-Means	0.78	0.62	0.28	0.72	0.76	1800	100
Agglomerative Clustering	0.80	0.65	0.26	0.76	0.79	1600	110
Mean Shift Clustering	0.88	0.72	0.20	0.82	0.86	1400	130
Gaussian Mixture Model (GMM)	0.90	0.74	0.19	0.84	0.87	1450	125
Spectral Clustering	0.75	0.60	0.30	0.70	0.74	1900	95
Affinity Propagation	0.83	0.68	0.22	0.80	0.84	N/A	N/A
Self-Organizing Maps (SOM)	0.79	0.64	0.27	0.74	0.77	1750	105
Proposed Method	0.94	0.80	0.15	0.90	0.92	1200	150

Table 2 lists K-Means++, Hierarchical Clustering, DBSCAN, Fuzzy C-Means, Agglomerative Clustering, Mean Shift, GMM, Spectral Clustering, Affinity Propagation, SOM, and our technique. Performance is measured using the Davies-Bouldin Index, Fowlkes-Mallows Index, Inertia Index, Calinski-Harabasz Index, Dunn Index, Normalized Mutual Information, Homogeneity Score, Completeness Score, and Variance Ratio Criterion.

The recommended technique consistently organizes blood donor data better than current algorithms, maximizing resource consumption. The Cluster Purity, Silhouette Coefficient, and Adjusted Rand Index are greater with our technique. The clusters are clearer and more accurate. The Davies-Bouldin Index and Fowlkes-Mallows Index suggest that the proposed strategy separates and shrinks clusters better. The sum of squares inside a cluster, or inertia, shows that the recommended strategy reduces cluster variance. Calinski-Harabasz Index values, which reflect variation rates within and within clusters, suggest that the approach creates well-separated clusters. The Dunn Index, which measures cluster distance, reveals that the suggested technique may create independent, non-combining clusters. Normalized Mutual Information, Homogeneity Score, and Completeness Score are inferior to the recommended technique. This demonstrates that it can discover key cluster patterns. Finally, the variance ratio criterion reveals that the suggested technique balances separation and tightness. It achieves this by comparing cluster variances to set variances. The recommended grouping algorithm outperforms others in several ways, proving it can manage blood donor data. These findings imply that the recommended method can help healthcare systems better allocate resources, resulting in more focused blood donor utilization.

5. Conclusions

This study improves K-means clustering to make blood donor data analysis and sharing easier. The National Institutes of Health conducted the study. Its pieces worked well following a thorough

examination and ablation research. K-Means++ initialization finds the optimal cluster centers faster, improving speed. Hierarchical structures are superior because they can uncover more complex linkages between blood donors. This tool displays blood donor trend categories, making grouping conclusions easy. Hierarchical modelling is needed for focused resource distribution, the study found. Strategic data sharing may reduce waste, maximize resources, and distribute blood products fairly. This planned split is crucial in cases of emergencies or urgent blood needs. Because it can handle diverse data types, the approach may be used for more than just grouping tasks and blood donor data. This is because the method works with several materials. Ablation research focuses on enhanced initialization, hierarchical grouping, and selective data input. All of these measures make the proposed method more likely to work. The study found that the method works with various data. It helps choose and maximize healthcare resources. Better K-means clustering is a powerful and versatile tool to classify and assess blood donor data. Healthcare management and decision support systems may tackle complex data patterns, track organizational hierarchies, and optimize resources.

References

[1] Y. Mei, J. Yu, J. Wang, and X. Li, "Data Analysis and Data Mining," Tsinghua University Press, Beijing, China, pp. 260-261, 2020.

[2] F. Wang and Z. Liu, "Optimization method of distributed K-means algorithm under Spark framework," Computer Engineering and Design, vol. 40, no. 06, pp. 1595–1600, 2019.

[3] Y. Zhang and F. Li, "Parallelization of canopy-K-means algorithm based on MapReduce," Journal of Liaoning University of Science and Technology, no. 01, pp. 4–13, 2017.

[4] D. Pathak and R. Kashyap, "Neural correlate-based E-learning validation and classification using convolutional and Long Short-Term Memory networks," Traitement du Signal, vol. 40, no. 4, pp. 1457-1467, 2023. [Online]. Available: https://doi.org/10.18280/ts.400414

[5] V. Roy, S. Shukla, P. K. Shukla, P. Rawat, "Gaussian Elimination-Based Novel Canonical Correlation Analysis Method for EEG Motion Artifact Removal", Journal of Healthcare Engineering, vol. 2017, Article ID 9674712, 11 pages, 2017. https://doi.org/10.1155/2017/9674712.

[6] S. Stalin, V. Roy, P. K. Shukla, A. Zaguia, M. M. Khan, P. K. Shukla, A. Jain, "A Machine Learning-Based Big EEG Data Artifact Detection and Wavelet-Based Removal: An Empirical Approach", Mathematical Problems in Engineering, vol. 2021, Article ID 2942808, 11 pages, 2021. https://doi.org/10.1155/2021/2942808

[7] Y. Liu, "Parallel K-means clustering algorithm based on sampling and maximum and minimum distance method," Intelligent Computer and Application, vol. 8, no. 06, pp. 37–39+43, 2018.

[8] Z. Wang, L. Jin, and Y. Song, "Improved K-means algorithm based on distance and weight," Computer Engineering and Applications, vol. 56, no. 23, pp. 87–94, 2020.

[9] C. Yan, "Data Mining Technology and Application," Tsinghua University Press, Beijing, China, p. 145, 2016.

[10] X. Li, L. Yu, and H. Lei, "Parallel implementation and application of an improved K-means algorithm," Journal of University of Electronic Science and Technology of China, vol. 46, no. 01, pp. 61–68, 2017.

[11] Roy V., Shukla S. (2013) Image Denoising by Data Adaptive and Non-Data Adaptive Transform Domain Denoising Method Using EEG Signal. In: Kumar V., Bhatele M. (eds) Proceedings of All India Seminar on Biomedical Engineering 2012 (AISOBE 2012). Lecture Notes in Bioengineering. Springer, India. https://doi.org/10.1007/978-81-322-0970-6_2.

[12] V. Roy et al., “Detection of sleep apnea through heart rate signal using Convolutional Neural Network,” International Journal of Pharmaceutical Research, vol. 12, no. 4, pp. 4829-4836, Oct-Dec 2020.

[13] R. Kashyap, "Machine Learning, Data Mining for IoT-Based Systems," in Research Anthology on Machine Learning Techniques, Methods, and Applications, Information Resources Management Association, Ed. IGI Global, 2022, pp. 447-471. [Online]. Available: https://doi.org/10.4018/978-1-6684-6291-1.ch025

[14] Z. Tang, B. Huang, and L. Lian, "Collaborative filtering recommendation algorithm based on improved Canopy clustering," Computer Application Research, vol. 37, no. 09, pp. 2615–2619+2639, 2020.

[15] S. Chen and R. Jia, "Improved K-medoids algorithm based on density weight Canopy," Computer Engineering and Science, vol. 41, no. 10, pp. 1823–1828, 2019.

[16] F. Liu, Y. Juan, and L. Yao, "Research on the number of clusters in K-means clustering algorithm," Electronic Design Engineering, vol. 25, no. 15, pp. 9–13, 2017.
[17] J. Xie and Y. E. Wang, "K-Means algorithm for optimizing initial cluster centers with minimum variance," Computer Engineering, vol. 40, no. 08, pp. 205–211+223, 2014.
[18] H. P. Sahu and R. Kashyap, "FINE_DENSEIGANET: Automatic medical image classification in chest CT scan using Hybrid Deep Learning Framework," International Journal of Image and Graphics [Preprint], 2023. [Online]. Available: https://doi.org/10.1142/s0219467825500044
[19] S. Stalin, V. Roy, P. K. Shukla, A. Zaguia, M. M. Khan, P. K. Shukla, A. Jain, "A Machine Learning-Based Big EEG Data Artifact Detection and Wavelet-Based Removal: An Empirical Approach," Mathematical Problems in Engineering, vol. 2021, Article ID 2942808, 11 pages, 2021.
[20] J. Sun and S. Wang, "Data Mining Algorithms and Applications," Tsinghua University Press, Beijing, China, vol. 145, pp. 147–149, 2020.
[21] X. Jiang, W. Zhang, J. Wang, and L. Ma, "Improved SK-means strategy based on stream processing," Journal of Beijing Information Technology University (Natural Science Edition), vol. 36, no. 05, pp. 51–56, 2021.