**CAHIERS MAGELLANES-NS**
Volume 06 Issue 2
2024

*ISSN:1624-1940*
DOI 10.6084/m9.figshare.26310112
http://magellanes.com/

# PATIENT CLUSTERING OPTIMIZATION WITH K-MEANS IN HEALTHCARE DATA ANALYSIS

**Keshav Kumar K.[1] and Dr. NVSL Narasimham[2]**

[1,2]Department of Humanities & Mathematics,
G. Narayanamma Institute of Technology and Science (for Women), Hyderabad, 500104,
Telangana, India.

***Contributing author: Keshav Kumar K***

***Abstract:***
***The technique known as K-Means is used in this study to optimize patient clustering for health care information analysis. Adopting an interpretivist mindset, a deductive method is utilized to improve the algorithm's efficiency and assess its resilience. Secondary data collection is used in descriptive research designs to enable in-depth analysis. The findings emphasize demographically-based patient cohorts, designed algorithms performance, along with algorithmic reliability. Accurate clustering is ensured by validation procedures, and the approach is compared to other approaches in a comparative analysis. Analyzing critically reveals both advantages and disadvantages. Scalability, hybrid models, along with interdisciplinary cooperation are encouraged in the recommendations. Subsequent research endeavors ought to explore sophisticated methodologies, dynamic aggregation, unsupervised machine learning, and ethical implications.***

**Keywords**
***Healthcare data analysis, K-Means clustering, optimization, patient cohorts, interdisciplinary collaboration.***

## 1. Introduction

### *A. Research background*

With the introduction of data-driven innovations, the world of health care is changing quickly, making sophisticated analytical techniques necessary for efficient management of patients [1]. A key component of health care information analysis is patient clustering, which makes it easier to identify unique patient groups with comparable traits. The popular clustering method K-Means has the potential to improve resource allocation along with personalized healthcare by optimizing patients grouping [2]. By utilizing K-Means aggregation, this study aims to further the increasingly expanding field of machine learning as well as healthcare by improving patient inequality. This study aims to reveal hidden patterns throughout patient populations by utilizing the large datasets accessible through the healthcare industry. The insights gained from this process can be used to inform specific treatments as well as therapy strategies [3]. These developments have the potential to completely transform the way healthcare is delivered, resulting in more accurate and customized patient care.

### *B. Research aim and objectives*

*Research Aim:*

To use the method known as K-Means to maximize patient clustering in medical data analysis, resulting in better resource allocation along with personalized medicine.

*Objectives:*

**CAHIERS MAGELLANES-NS**
Volume 06 Issue 2
2024

*ISSN:1624-1940*
DOI 10.6084/m9.figshare.26310112
http://magellanes.com/

- To assess how well K-Means clustering separates patients into discrete groups according to similar healthcare attributes.
- To enhanced feature selection along with dimensionality reduction methods' interpretability and clinical significance for K-Means-generated patient clusters.
- To assess the effects of patient clustering optimization on customized healthcare, encompassing intervention strategies that are specifically designed and treatment response prediction.
- To assess the effects of clustering based on K-Means on the distribution of healthcare resources with the goal of enhancing the efficiency and affordability of the provision of healthcare.

### *C. Research Rationale*

The necessity to improve patient management along with analytics in healthcare through sophisticated clustering techniques is the driving force behind this investigation [4]. The establishment of customized interventions is hampered by the fact that conventional health care addresses frequently ignore the complex and varied makeup of patient communities. This research attempts to reveal subtle patient subgroups, promoting greater comprehension of healthcare features and treatment responses. It does this by utilizing the robust clustering approach known as the algorithm known as K-Means. It is anticipated that patient gathering optimization will produce more clinically meaningful and comprehensible results, empowering doctors to make well-informed decisions [5]. Furthermore, the study investigates how designed clustering can support effective and economical healthcare delivery, addressing the larger healthcare opposition of resource allocation. The ultimate goal of this research is to advance healthcare practices in the direction of an additional data-driven along with personalized paradigm.

## 2. Literature review

### *A. Introduction to Healthcare Data Analysis*

The quest for better patient outcomes, efficient treatment, and integrated healthcare procedures has made healthcare data analysis a key component. The implementation of electronic health records (EHRs) along with the growth of health-related data is causing an important change in the healthcare industry in favor of data-driven choice-making [6]. This shift is being fueled by the realization that hidden away in the enormous databases of patient data are important insights that are essential for machine learning, personalized healthcare, and effective resource management. Because of the volume and complexity in medical data, advanced analytical techniques are required, which makes data analysis a vital tool for administrators and health care providers.

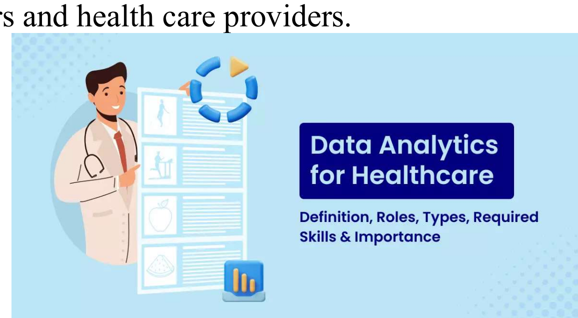


**Fig. 2.1**: Introduction to Healthcare Data Analysis

Healthcare data analysis has the potential to completely change how we provide medical care by revealing patterns in disease, forecasting patient outcomes, and streamlining treatment regimens. But this potential also presents a unique set of difficulties, such as the requirement for strong analytical

ISSN:1624-1940
DOI 10.6084/m9.figshare.26310112
http://magellanes.com/

techniques, privacy issues, and data integration [7]. Given this, it is critical for both scientists and clinicians to comprehend the state of medical data analysis. Doing so will pave the way for innovations that have the potential to improve patient care and the effectiveness of the medical industry as a whole.

### *B. Unsupervised Learning in Healthcare*

With no predetermined labels or direction, independent learning has become a disruptive force in healthcare, providing fresh insights through patient data. Unsupervised algorithms for learning are essential for revealing hidden relationships and frameworks in the field of healthcare, where data volumes and complexity are enormous [8]. Unsupervised learning enables the algorithm to find relationships throughout datasets on its own, making it easier to find complex connections that might be missed by human observers. Specifically, methods for clustering are notable for being important components of unsupervised learning in the medical field.


**Fig. 2.2:** Unsupervised Learning in Healthcare

These algorithms, such as the popular K means technique, are excellent at classifying patient data into discrete groups according to commonalities, which facilitates the discovery of cohorts that have similar features. This method works very well for research on epidemiology, treatment personalization, and illness stratification [9]. Unsupervised training in the healthcare industry does, however, come with certain difficulties, such as understanding along with information assurance for accuracy. Unsupervised learning has the potential to improve treatment plans, diagnosis, and treatment for patients in general as the healthcare sector continues to take advantage of its power.

### *C. K-Means Clustering: Concepts and Applications*

The ability of K-Means clustering, a basic unsupervised machine learning algorithm, to reveal underlying structures in intricate patient datasets is making it more and more popular in the healthcare industry. Basically, K-Means is trying to divide data points into different clusters, with the center of each cluster being a prototype of divulged features [10]. This idea makes it possible to identify patient groups that are similar, which makes it easier to implement focused interventions along with individualized treatment plans. The algorithm's efficiency comes from its iterative optimization procedure, which minimizes within-cluster variance by continuously improving cluster assignments.

**CAHIERS MAGELLANES-NS**
Volume 06 Issue 2
2024

*ISSN:1624-1940*
DOI 10.6084/m9.figshare.26310112
http://magellanes.com/

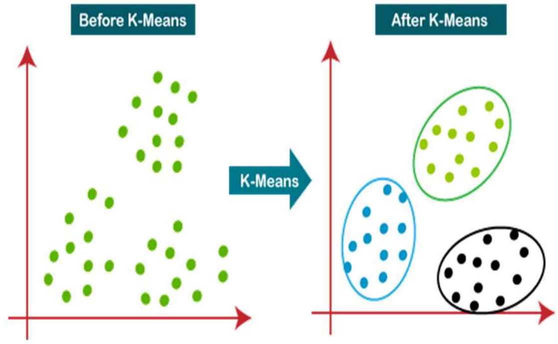


**Fig. 2.3:** K-Means Clustering

But there are some issues that need to be carefully considered, like the sensitivity to the starting conditions as well as the requirement to specify the total number of groupings, or K. K-Means is a valuable tool in healthcare settings as it helps with patient classification and the identification of specific populations that share comparable medical individuals or responses to therapy [11]. It also aids in the optimization of resource allocation since customized medical treatments can be created for every patient cluster. Even though K-Means has strong potential, researchers need to understand its subtleties and look into ways to improve it in order to best tailor it to the particular difficulties presented by healthcare data.

### *D. Existing Research in Patient Clustering Optimization*

The corpus of research on receptive clustering the improvement for health care data analysis that is currently available demonstrates a wide variety of approaches and uses. To improve healthcare decisions, a number of studies have used algorithms that use clustering, most notably K-Means, to identify patterns in patient datasets. These studies have shown how effective clustering is at identifying patient subgroups that share traits, which makes it possible to implement more specialized and individualized healthcare solutions [12]. Patient clustering has been used by researchers in a number of areas, such as the allocation of healthcare resources, response to therapy prediction, along with disease stratification. Research has investigated how clustering can enhance the precision of diagnoses, optimize treatment regimens, and customize interventions for particular patient profiles.

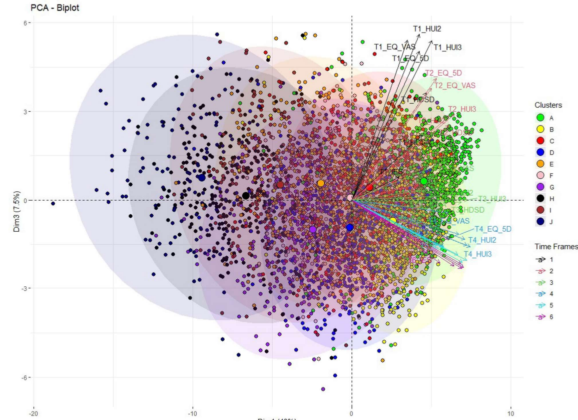


**Fig. 2.4:** Existing Research in Patient Clustering Optimization

Although K-Means is a popular option, some studies explore the incorporation of combination models or alternate methods for clustering to improve the preciseness of patient organization. Nevertheless, issues still exist, such as the requirement for strong validation procedures and the highly

**CAHIERS MAGELLANES-NS**
Volume 06 Issue 2
2024

*ISSN:1624-1940*
DOI 10.6084/m9.figshare.26310112
http://magellanes.com/

sensitive nature of clustering results to algorithm parameters required [13]. The strengths and weaknesses of the current research are carefully evaluated in this review, which lays the groundwork for future investigations and suggests directions for improving patient clustering methods in healthcare evaluation of data.

### *E. Literature Gap*

There is a dearth of literature that systematic addresses the issues of algorithm parameter awareness, result validation, along with the incorporation of hybrid models, despite the abundance of research on patients clustering in medical settings using algorithms such as K-Means. Currently available research frequently lacks thorough analyses of clustering results, which impedes the creation of reliable, broadly applicable techniques. In order to ensure the dependability and efficacy of patient accumulating techniques in a variety of healthcare applications it is imperative that this gap be closed.

## 3. Methodology

The interpretative philosophy used in this study acknowledges the inherent complexity and contextual influences on healthcare data. The objective is to explore the personal elements of patient data, comprehend the variety of viewpoints present in the data, and place findings in the context of the larger healthcare system. To methodically evaluate theories constructed using empirical data and current theories, the method of deductive reasoning is used [14]. The study intends to apply these hypotheses to particular healthcare scenarios, that contribute to the improvement of patient accumulating methodologies. It begins with established guidelines related to patients clustering along with enhancing algorithms. In order to give a thorough overview of the situation in the field of healthcare clustering the improvement at the moment, a descriptive study approach is selected [15]. This design facilitates the investigation of current approaches, obstacles, and prospects related to the optimization of patient grouping. In order to provide a comprehensive overview of the situation and pinpoint areas in need of development, comprehensive descriptions of the clustering results, algorithm settings, along with validation procedures will be gathered [16]. The research employs secondary data collection techniques, sourcing information from a variety of hospital databases, scholarly articles, and electronic medical records. The utilization of secondary data is in line with the practicality of employing information that is easily accessible and facilitates the consolidation of data from multiple sources to augment the scope and profundity of the research [17]]. Locate and compile pertinent datasets that include results from treatment, health records, and patient demographics. Preprocess data to fix problems like values that are absent, outliers, and establishing variable formats in order to guarantee consistency and accuracy. Undertake a thorough review of the literature on patient accumulating optimization, with a focus on inference principles along with interpretive insights. Create theories based on recognized gaps and current theories in order to direct the process of technical optimization [18]. Utilizing libraries like scikit-learn and a language for programming including Python, implement the power source K means clustering algorithm. To improve clustering reliability, optimize system criteria, such as the total amount of clusters (K), by carefully testing variations using cross-validation techniques [19]. Combine results, evaluating them in light of the larger medical treatment context and contrasting optimal clustering final results with current approaches. Conclude on the efficacy of the improved K-Means conduct while taking into account the interpretive subtleties found in the study [20]. Make practical suggestions for future research and application in healthcare settings.

**CAHIERS MAGELLANES-NS**
Volume 06 Issue 2
2024

*ISSN:1624-1940*
DOI 10.6084/m9.figshare.26310112
http://magellanes.com/

## 4. Results

### *A Theme: Optimized K-Means Clustering: Algorithmic Performance and Parameter Sensitivity*

An extensive examination of the technical nuances involved in putting the K-Means clustering process into practice and optimizing its performance is given in the section titled "Improved K-Means Clustering is the process Algorithmic Achievement along with Parameter Sensitivity". The purpose of this analysis is to clarify the algorithm's performance features and parameter to be used sensitivity. Strict parameter adjustment is required during the optimization procedure, especially when figuring out the ideal number of clusters (K), which is accomplished using methodical cross-validation strategies [21]. The effects of these adjustments on the general efficacy of the algorithm are examined in this section. Measures like convergence actions, computational speed, and gathering accuracy are carefully examined to determine how well the optimized K-Means method performs. This section's key component is parameter to be used sensitivity analysis, which sheds light on how adjustments to computational parameters affect the results of clustering.

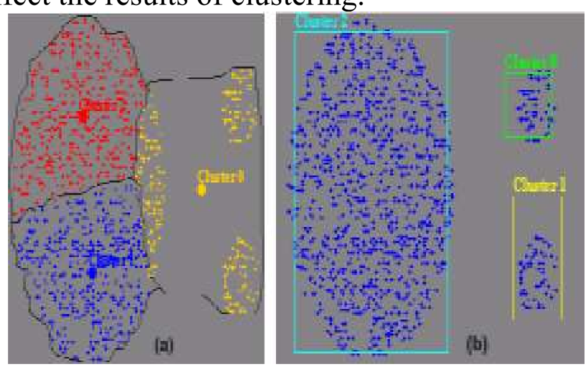


**Fig. 4.1:** Algorithmic Performance and Parameter Sensitivity

It also analyze in detail how different datasets affect the algorithm's performance and how sensitive the algorithm is to initialization circumstances. It is critical to take care of these issues in order to guarantee the optimized K-Means algorithm's resilience and applicability in a variety of healthcare information scenarios. Furthermore, issues with algorithm expansion, convergence, along with mathematical resource use are covered [22]. This section functions as a technical manual, providing insight into the operation of the technique in various scenarios and highlighting the significance of fine-tuning parameters to achieve the best outcomes when patient clustering is analyzed in healthcare data. The results that follow provide a more detailed understanding of the functioning of the algorithm and its repercussions for improving patient clustering techniques. They also serve as a foundation for the sections that follow.

### *B Theme: Cluster Analysis: Profiling Patient Cohorts and Characteristics*

The section titled "Cluster Analysis: profiling, which is Patient Populations and Characteristics" offers a thorough examination of the outcomes derived from the K-Means clustering optimization. In-depth examination of the various patient cohorts discovered during the clustering process is provided by this analysis, which also sheds light on their special traits, medical backgrounds, and treatment results [23]. Through an examination of the traits shared by every patient category, this section clarifies the subtle differences and similarities among the groups that have been identified. The objective is to reveal significant patterns that might not be immediately noticeable without the use of sophisticated clustering techniques. A better understanding of the tremendous heterogeneity found in healthcare data is made possible by insights into individuals cohorts, which may open the door to customized medical treatment and individualized healthcare strategies.

**CAHIERS MAGELLANES-NS**
Volume 06 Issue 2
2024

*ISSN:1624-1940*
DOI 10.6084/m9.figshare.26310112
http://magellanes.com/

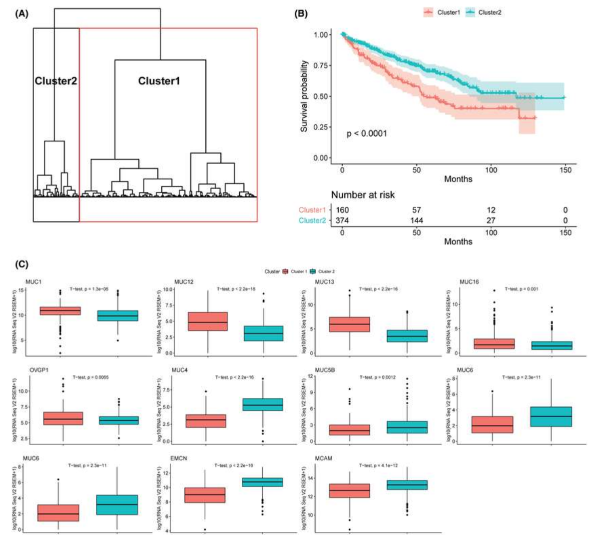


**Fig. 4.2** Profiling Patient Cohorts and Characteristics

The section uses graphics and statistical information to highlight the salient characteristics of every individual in the cluster. A few of the analysis's main focus areas are responses to medication, prevalence of particular ailments, and demographic data [24]. The cluster-specific features can be highlighted using comparable visualizations, like heatmaps as well as radar data charts, which give a clear and succinct picture of the profiling results. Moreover, this analysis establishes the foundation for further theories and recommendations in addition to providing a snapshot about the current status of patient cohorts. It serves as an essential link between this clustering procedure's related to technology optimization and its application in decision-making in healthcare. The results in this section highlight the potential influence of optimized gathering methodologies on enhancing patient outcomes along with resource allocation strategies, adding to the ongoing discussion about patient-centric treatment.

### *C Theme: Validation and Robustness Testing: Assessing the Reliability of Clustering Outcomes*

The robustness along with validity of the designed K-Means clustering technique are critically examined in the section titled "Validation along with Robustness Inspection. Assessing the Availability of Clustering Outcomes". Strict validation procedures are used to make sure that the patient clusters that have been identified are reliable and to evaluate how robust the algorithm is in a variety of clinical datasets [25]. The coherence and dispersion of clusters are assessed using internal testing methods like the Davies-Bouldin indices along with Silhouette score. These metrics offer numerical evaluations of the technique's performance concerning both inter- while intra-cluster dissimilarities. In addition, the thorough evaluation of clustering dependability is aided by external validation techniques such as comparisons alongside ground truth information sets or known patient categories. A major area of interest is the analysis of sensitivity, which investigates how differences in dataset properties and system parameters affect the results of clustering.

**CAHIERS MAGELLANES-NS**
Volume 06 Issue 2
2024

*ISSN:1624-1940*
DOI 10.6084/m9.figshare.26310112
http://magellanes.com/

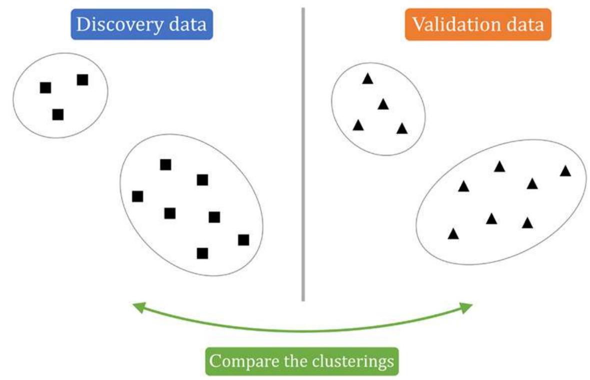


**Fig. 4.3:** Assessing the Reliability of Clustering Outcomes

This section tries to pinpoint the algorithm's the relevance boundaries and point out possible areas for improvement by methodically testing the approach under various scenarios. To make sure that the designed K-Means approach remains effective in a variety of healthcare circumstances and retains its generalizability along with applicability within real-life situations, robustness examination is essential. The handling of data that has noise or outliers, among other potential difficulties came across during the verification process, is also covered in this section. We address ways to mitigate these issues, highlighting the role that preprocessing stages and outlier detection techniques play in improving the overall dependability of clustering results [26]. The designed K-Means algorithm for clustering is more credible and trustworthy overall as a result of the results of this validation along with robustness testing section. Carefully assessing its capabilities and shortcomings lays the groundwork for interpreting the findings that follow and drawing important conclusions for medical professionals and policymakers.

***D Theme: Comparative Analysis with Existing Methodologies: Benchmarking the Optimized Approach***

*The "Comparative Analysis with Existing Methodologies:* Benchmarking the power source Optimized Approach" evaluates the optimized K-Means grouping algorithm's effectiveness for patient accumulating within medical treatment data analysis carefully in comparison to established methodologies. The purpose of the comparison is to clarify the advantages, disadvantages, and special contributions made by the optimized strategy. The tailored K-Means method's new findings and improvements are assessed by comparison it against additional sophisticated methods and current techniques such as conventional clustering algorithms [27]. The efficacy of the algorithm is evaluated using comparative metrics, which include clustering reliability, computational speed, along with the technique's capacity to handle a variety of datasets. This section delves into the subtleties of clustering results, highlighting cases in which the optimized K-Means method performs exceptionally well at identifying small trends or attaining greater accuracy when grouping patients.

**CAHIERS MAGELLANES-NS**
Volume 06 Issue 2
2024

*ISSN:1624-1940*
DOI 10.6084/m9.figshare.26310112
http://magellanes.com/


**Fig. 4.4:** Comparative Analysis with Existing Methodologies

In addition, it covers potential drawbacks and situations in which alternative approaches might perform better than the optimized one, offering a thorough and impartial viewpoint. To improve the readability of the comparison analysis, visualizations like performance charts and comparative clustering profiles can be used. These illustrations support the research's contribution to the subject and help to communicate the technical subtleties of the accumulating outcomes [28]. The results of this comparative study give medical professionals important new information and direction regarding the choice of clustering techniques for particular situations. Comprehending the relative benefits and constraints of the designed K-Means technique augments its suitability in various healthcare contexts, directing professionals and investigators towards more knowledgeable and efficient decision-making concerning patient accumulating optimization.

| **Aspect** | **Description** |
|---|---|
| Algorithm Performance Metrics | Clustering Accuracy: Measurement of how well the algorithm groups data points |
| Parameter Optimization | Determination of Optimal K: Systematic cross-validation techniques to identify the optimal number of clusters |
| Parameter Sensitivity Analysis | Initialization Sensitivity: Examination of the algorithm's response to different starting conditions |
| Algorithm | Analysis of |

CAHIERS MAGELLANES-NS
Volume 06 Issue 2
2024

ISSN:1624-1940
DOI 10.6084/m9.figshare.26310112
http://magellanes.com/

| | |
|---|---|
| Convergence and Scalability | convergence patterns and scalability considerations for varying dataset sizes |

## 5. Evaluation and conclusion

### *A Critical Evaluation*

The investigation's intellectual the foundation is its critical evaluation section, which provides a thorough analysis of the results, methods, and implications. It entails a careful assessment of the investigation's approach's advantages and disadvantages, the reliability of the findings, and the applicability of the power source K-Mean clustering technique's optimized version. In addition, this section examines the implications of any assumptions that were made during the research and takes into account any presumptions or confounding variables that might have affected the results. Moreover, it provides analyses of the research's wider ramifications in the healthcare field, pointing out how the study adds to the body of understanding already in existence and outlining potential directions for further investigation. A thorough critical analysis strengthens the investigation's credibility by offering an open and thoughtful viewpoint on the methodology and findings.

### *B Research recommendation*

The findings and recommendations highlight areas that need more investigation and improvement in the area of patient gathering optimization for medical data analysis alongside the method known as K-Means. In the beginning, comprehensive studies examining the optimized strategy's scalability across different healthcare environments as well as information volumes are required. Examining adoption successes and problems in the real world can provide useful information. To improve the accuracy and resilience of patient grouping, more study should be done on hybrid models, which combine K-Means along with additional accumulating strategies or artificial intelligence algorithms [29]. Additionally, to address the technique's sensitivity to setting up conditions and improve its capacity to adapt to changing health care information, future research could concentrate on broadening optimization methods. To evaluate the stability and efficacy of the designed K means conduct over time, ongoing investigations are advised. Last but not least, interdisciplinary partnerships between knowledge scientists, medical professionals, along with legislators can help translate research results into practical plans for enhancing patient care, allocating resources, and arriving at healthcare decisions.

### *C Future work*

Future research in this area should investigate sophisticated techniques to overcome the noted drawbacks and improve the use of clustering using K-Means in the examination of health care data. Examining how deep learning methods and neural systems can work together could provide more accurate clustering as well as complex visualizations of patient data. Furthermore, studies should concentrate on creating dynamic clustering models that can change to reflect changing healthcare trends along with patient profiles [30]. Examining the possibilities of unsupervised feature acquisition techniques can help reveal hidden patterns in the information, thus improving the recognition of patient

*ISSN:1624-1940*
DOI 10.6084/m9.figshare.26310112
http://magellanes.com/

cohorts. To guarantee the practical use of accumulating outcomes in practical problems medical environments and to incorporate particular domain expertise, increased collaboration between data scientists as well as medical professionals is necessary. Lastly, studies should focus on the ethical ramifications of patient clustering, especially with regard to consent, confidentiality of information, and the appropriate application of clustering findings to healthcare decision-making.

**Conflict of Interest:**
The author have no conflict of interest to declare.

## References

[1] R. A. Taylor *et al*, "Computational phenotypes for patients with opioid-related disorders presenting to the emergency department," *PLoS One,* vol. 18, *(9),* 2023. Available: https://www.proquest.com/scholarly-journals/computational-phenotypes-patients-with-opioid/docview/2865290212/se-2. DOI: https://doi.org/10.1371/journal.pone.0291572.

[2] K. Rahul, R. K. Banyal and N. Arora, "A systematic review on big data applications and scope for industrial processing and healthcare sectors," *Journal of Big Data,* vol. 10, *(1),* pp. 133, 2023. Available: https://www.proquest.com/scholarly-journals/systematic-review-on-big-data-applications-scope/docview/2857705190/se-2. DOI: https://doi.org/10.1186/s40537-023-00808-2.

[3] M. Govindaraj, S. Kaliappan and G. Swaminathan, "An Enhanced Outlier Detection Approach for Multidimensional Datasets Using a Synergistic Firefly and Grey Wolf Optimization-Based Method," *Ingenierie Des Systemes d'Information,* vol. 28, *(3),* pp. 767-775, 2023. Available: https://www.proquest.com/scholarly-journals/enhanced-outlier-detection-approach/docview/2843565077/se-2. DOI: https://doi.org/10.18280/isi.280328.

[4] M. Pikoula *et al*, "Evaluation of data processing pipelines on real-world electronic health records data for the purpose of measuring patient similarity," *PLoS One,* vol. 18, *(6),* 2023. Available: https://www.proquest.com/scholarly-journals/evaluation-data-processing-pipelines-on-real/docview/2826187309/se-2. DOI: https://doi.org/10.1371/journal.pone.0287264.

[5] H. Asri and Z. Jarir, "Toward a smart health: big data analytics and IoT for real-time miscarriage prediction," *Journal of Big Data,* vol. 10, *(1),* pp. 34, 2023. Available: https://www.proquest.com/scholarly-journals/toward-smart-health-big-data-analytics-iot-real/docview/2786734106/se-2. DOI: https://doi.org/10.1186/s40537-023-00704-9.

[6] F. O. Giuste *et al*, "Early and fair COVID-19 outcome risk assessment using robust feature selection," *Scientific Reports (Nature Publisher Group),* vol. 13, *(1),* pp. 18981, 2023. Available: https://www.proquest.com/scholarly-journals/early-fair-covid-19-outcome-risk-assessment-using/docview/2885680697/se-2. DOI: https://doi.org/10.1038/s41598-023-36175-4.

[7] A. Flothow, A. Novelli and L. Sundmacher, "Analytical methods for identifying sequences of utilization in health data: a scoping review," *BMC Medical Research Methodology,* vol. 23, pp. 1-17, 2023. Available: https://www.proquest.com/scholarly-journals/analytical-methods-identifying-sequences/docview/2877495417/se-2. DOI: https://doi.org/10.1186/s12874-023-02019-y.

[8] A. A. Kasasbeh, "Applying Artificial Intelligence and Machine Learning to Improve Healthcare Outcomes in Marginalized Patient Populations." Order No. 30492696, State University of New York at Binghamton, United States -- New York, 2023.

[9] M. Chaudhry *et al*, "A Systematic Literature Review on Identifying Patterns Using Unsupervised Clustering Algorithms: A Data Mining Perspective," *Symmetry,* vol. 15, *(9),* pp. 1679, 2023. Available:

*ISSN:1624-1940*
DOI 10.6084/m9.figshare.26310112
http://magellanes.com/

https://www.proquest.com/scholarly-journals/systematic-literature-review-on-identifying/docview/2869647103/se-2. DOI: https://doi.org/10.3390/sym15091679.

[10] U. K. Muhammad Nafees *et al*, "Fuzzy-Based Efficient Healthcare Data Collection and Analysis Mechanism Using Edge Nodes in the IoMT," *Sensors,* vol. 23, *(18),* pp. 7799, 2023. Available: https://www.proquest.com/scholarly-journals/fuzzy-based-efficient-healthcare-data-collection/docview/2869628273/se-2. DOI: https://doi.org/10.3390/s23187799.

[11] D. S. Erana Veerappa *et al*, "Interoperable IoMT Approach for Remote Diagnosis with Privacy-Preservation Perspective in Edge Systems," *Sensors,* vol. 23, *(17),* pp. 7474, 2023. Available: https://www.proquest.com/scholarly-journals/interoperable-iomt-approach-remote-diagnosis-with/docview/2862731153/se-2. DOI: https://doi.org/10.3390/s23177474.

[12] A. Abbas, J. Neasham and M. Naqvi, "Image Formation Algorithms for Low-Cost Freehand Ultrasound Scanner Based on Ego-Motion Estimation and Unsupervised Clustering," *Electronics,* vol. 12, *(17),* pp. 3634, 2023. Available: https://www.proquest.com/scholarly-journals/image-formation-algorithms-low-cost-freehand/docview/2862245692/se-2. DOI: https://doi.org/10.3390/electronics12173634.

[13] X. Hao *et al*, "Assessing resource allocation based on workload: a data envelopment analysis study on clinical departments in a class a tertiary public hospital in China," *BMC Health Services Research,* vol. 23, pp. 1-13, 2023. Available: https://www.proquest.com/scholarly-journals/assessing-resource-allocation-based-on-workload/docview/2852021146/se-2. DOI: https://doi.org/10.1186/s12913-023-09803-y.

[14] L. Shrotriya *et al*, "Apache Spark in Healthcare: Advancing Data-Driven Innovations and Better Patient Care," *International Journal of Advanced Computer Science and Applications,* vol. 14, *(6),* 2023. Available: https://www.proquest.com/scholarly-journals/apache-spark-healthcare-advancing-data-driven/docview/2843255118/se-2. DOI: https://doi.org/10.14569/IJACSA.2023.0140665.

[15] J. Dansana, M. R. Kabat and P. K. Pattnaik, "Improved 3D Rotation-based Geometric Data Perturbation Based on Medical Data Preservation in Big Data," *International Journal of Advanced Computer Science and Applications,* vol. 14, *(5),* 2023. Available: https://www.proquest.com/scholarly-journals/improved-3d-rotation-based-geometric-data/docview/2830496303/se-2. DOI: https://doi.org/10.14569/IJACSA.2023.0140592.

[16] N. Razfar, R. Kashef and F. Mohammadi, "Automatic Post-Stroke Severity Assessment Using Novel Unsupervised Consensus Learning for Wearable and Camera-Based Sensor Datasets," *Sensors,* vol. 23, *(12),* pp. 5513, 2023. Available: https://www.proquest.com/scholarly-journals/automatic-post-stroke-severity-assessment-using/docview/2829880758/se-2. DOI: https://doi.org/10.3390/s23125513.

[17] N. Berros *et al*, "Enhancing Digital Health Services with Big Data Analytics," *Big Data and Cognitive Computing,* vol. 7, *(2),* pp. 64, 2023. Available: https://www.proquest.com/scholarly-journals/enhancing-digital-health-services-with-big-data/docview/2829699333/se-2. DOI: https://doi.org/10.3390/bdcc7020064.

[18] Q. An *et al*, "A Comprehensive Review on Machine Learning in Healthcare Industry: Classification, Restrictions, Opportunities and Challenges," *Sensors,* vol. 23, *(9),* pp. 4178, 2023. Available: https://www.proquest.com/scholarly-journals/comprehensive-review-on-machine-learning/docview/2812740886/se-2. DOI: https://doi.org/10.3390/s23094178.

[19] L. C. Maela Madel *et al*, "A Comprehensive Analysis of Clustering Public Utility Bus Passenger's Behavior during the COVID-19 Pandemic: Utilization of Machine Learning with Metaheuristic Algorithm," *Sustainability,* vol. 15, *(9),* pp. 7410, 2023. Available: https://www.proquest.com/scholarly-journals/comprehensive-analysis-clustering-public-

*ISSN:1624-1940*
DOI 10.6084/m9.figshare.26310112
http://magellanes.com/

utility/docview/2812737577/se-2. DOI: https://doi.org/10.3390/su15097410.

[20] M. Lakshmi and R. Das, "Classification of Monkeypox Images Using LIME-Enabled Investigation of Deep Convolutional Neural Network," *Diagnostics,* vol. 13, *(9),* pp. 1639, 2023. Available: https://www.proquest.com/scholarly-journals/classification-monkeypox-images-using-lime/docview/2812383828/se-2. DOI: https://doi.org/10.3390/diagnostics13091639.

[21] C. Ntakolia and D. V. Lyridis, "Path Planning in the Case of Swarm Unmanned Surface Vehicles for Visiting Multiple Targets," *Journal of Marine Science and Engineering,* vol. 11, *(4),* pp. 719, 2023. Available: https://www.proquest.com/scholarly-journals/path-planning-case-swarm-unmanned-surface/docview/2806556261/se-2. DOI: https://doi.org/10.3390/jmse11040719.

[22] A. Karras *et al*, "AutoML with Bayesian Optimizations for Big Data Management," *Information,* vol. 14, *(4),* pp. 223, 2023. Available: https://www.proquest.com/scholarly-journals/automl-with-bayesian-optimizations-big-data/docview/2806539582/se-2. DOI: https://doi.org/10.3390/info14040223.

[23] S. Tufail *et al*, "Advancements and Challenges in Machine Learning: A Comprehensive Review of Models, Libraries, Applications, and Algorithms," *Electronics,* vol. 12, *(8),* pp. 1789, 2023. Available: https://www.proquest.com/scholarly-journals/advancements-challenges-machine-learning/docview/2806536736/se-2. DOI: https://doi.org/10.3390/electronics12081789.

[24] O. Pasin and S. Gonenc, "An investigation into epidemiological situations of COVID-19 with fuzzy K-means and K-prototype clustering methods," *Scientific Reports (Nature Publisher Group),* vol. 13, *(1),* pp. 6255, 2023. Available: https://www.proquest.com/scholarly-journals/investigation-into-epidemiological-situations/docview/2802201910/se-2. DOI: https://doi.org/10.1038/s41598-023-33214-y.

[25] D. D. Olatinwo *et al*, "IoT-Enabled WBAN and Machine Learning for Speech Emotion Recognition in Patients," *Sensors,* vol. 23, *(6),* pp. 2948, 2023. Available: https://www.proquest.com/scholarly-journals/iot-enabled-wban-machine-learning-speech-emotion/docview/2791700352/se-2. DOI: https://doi.org/10.3390/s23062948.

[26] S. Deshmukh, B. K. Behera and P. Mulay, "Patient Data Analysis with the Quantum Clustering Method," *Quantum Reports,* vol. 5, *(1),* pp. 138, 2023. Available: https://www.proquest.com/scholarly-journals/patient-data-analysis-with-quantum-clustering/docview/2791685979/se-2. DOI: https://doi.org/10.3390/quantum5010010.

[27] A. F. Latefa Hamad, I. S. Mahasin and B. A. Malik, "Application of Deep Learning System Technology in Identification of Women's Breast Cancer," *Medicina,* vol. 59, *(3),* pp. 487, 2023. Available: https://www.proquest.com/scholarly-journals/application-deep-learning-system-technology/docview/2791671777/se-2. DOI: https://doi.org/10.3390/medicina59030487.

[28] F. H. Awad, M. M. Hamad and L. Alzubaidi, "Robust Classification and Detection of Big Medical Data Using Advanced Parallel K-Means Clustering, YOLOv4, and Logistic Regression," *Life,* vol. 13, *(3),* pp. 691, 2023. Available: https://www.proquest.com/scholarly-journals/robust-classification-detection-big-medical-data/docview/2791670554/se-2. DOI: https://doi.org/10.3390/life13030691.

[29] B. Su *et al*, "K-Anonymity Privacy Protection Algorithm for Multi-Dimensional Data against Skewness and Similarity Attacks," *Sensors,* vol. 23, *(3),* pp. 1554, 2023. Available: https://www.proquest.com/scholarly-journals/k-anonymity-privacy-protection-algorithm-multi/docview/2774978782/se-2. DOI: https://doi.org/10.3390/s23031554.

[30] M. Lukauskas and T. Ruzgas, "Reduced Clustering Method Based on the Inversion Formula Density Estimation," *Mathematics,* vol. 11, *(3),* pp. 661, 2023. Available:

ISSN:1624-1940
DOI 10.6084/m9.figshare.26310112
http://magellanes.com/

https://www.proquest.com/scholarly-journals/reduced-clustering-method-based-on-inversion/docview/2774926409/se-2. DOI: https://doi.org/10.3390/math11030661.

**Author Biography:**

1) Mr. Keshav Kumar K. is currently working as an Assistant Professor of Mathematics in the Department of Humanities and Mathematics, G.Narayanamma Institute of Technology and Science (for Women)(Autonomous), Hyderabad since 2009. He received his M.Sc. degree in Applied Mathematics (Systems Theory and Computer Modelling) with distinction, M.Tech. degree in Computer Science with distinction from J. N. T.U., Hyderabad and pursuing PhD from JNTU, Hyderabad. He also did M.Phil. degree in Mathematics from Madurai Kamaraj University, Madurai. He has been teaching Engineering Mathematics for B.Tech. graduates for the past 18 years. He also taught Probability and Statistics, Discrete Mathematics, Operations Research, Complex Analysis, Data Analytics, Differential Equations courses for undergraduate and post graduate students. He has presented and published several research papers in refereed international journals and in the proceedings of various international conferences. His areas of research include Artificial Intelligence (A.I.) and Machine Learning Algorithms, Mathematical Modelling, Multi Objective Optimization using Evolutionary Algorithms, Fuzzy set theory. He is a life member of ISTE, MTTF, IIISR, IAENG and various reputed Mathematical Associations.

2) Dr. NVSL Narasimham is currently working as an Associate Professor of Mathematics in the Department of Humanities and Mathematics, G. Narayanamma Institute of Technology and Science (for Women)(Autonomous), Hyderabad since 2005. He received his Ph.D. from Osmania University, Hyderabad, India. He has been teaching Engineering Mathematics for B.Tech. Graduates for the past 28 years. He also taught Probability and Statistics, Discrete Mathematics, Operations Research, Complex Analysis, Differential Equations courses for undergraduate and post graduate students. He has published several research papers in refereed international journals. His areas of research include Mathematical Modelling in Biomedical Engineering, Optimization techniques.